KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ESAT
AFDELING PSI
Kardinaal Mercierlaan 94 — 3001 Heverlee, Belgium

SELF-CALIBRATION AND METRIC 3D RECONSTRUCTION FROM UNCALIBRATED IMAGE SEQUENCES

Promotor:
Prof. L. VAN GOOL

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de Toegepaste Wetenschappen

door

Marc POLLEFEYS

Mei 1999

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ESAT
AFDELING PSI
Kardinaal Mercierlaan 94 — 3001 Heverlee, Belgium

SELF-CALIBRATION AND METRIC 3D RECONSTRUCTION FROM UNCALIBRATED IMAGE SEQUENCES

Jury:
Voorzitter: Prof. E. Aernhoudt
Prof. L. Van Gool, promotor
Prof. P. Wambacq
Prof. A. Zisserman (Oxford Univ.)
Prof. Y. Willems
Prof. H. Maître (ENST, Paris)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de Toegepaste Wetenschappen

door

Marc POLLEFEYS

U.D.C. 681.3*I48

Mei 1999

©Katholieke Universiteit Leuven - Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

Wettelijk depot D/1999/7515/23

ISBN 90-5682-193-8

Acknowledgements

At this point I would like to express my gratitude towards my advisor, Prof. Luc Van Gool, who gave me the opportunity to work in his research group. He provided an exciting working environment with many opportunities to develop new ideas, work on promising applications and meet interesting people.

I would also like to thank Andrew Zisserman and Patrick Wambacq for accepting to be in my reading committee. I am especially indebted to Andrew Zisserman for the fruitful discussions we had and for the interesting visits to his group in Oxford. I am also grateful to the other members of the jury, Henri Maître and Yves Willems, who accepted this task with enthusiasm.

Of course, I am also indebted to many colleagues. I would like to especially acknowledge Reinhard Koch and Maarten Vergauwen for the work we did together. Marc Proesmans, Tinne Tuytelaars, Joris Vanden Wyngaerd and Theo Moons also deserve a special mention for contributing to some results presented in this work. Besides this I would also like to thank all my colleagues who turned these years at ESAT into a pleasant time.

The financial support of the IWT is also gratefully acknowledged.

Last but not least, I would like to thank my parents, my family and my friends for their patience and support. This was very important to me.

Abstract

This thesis discusses the possibility to obtain three dimensional reconstructions of scenes from image sequences. Traditional approaches are based on a preliminary calibration of the camera setup. This, however, is not always possible or practical. The goal of this work was to investigate how this calibration constraint could be relaxed.

The approach was twofold. First, the problem of self-calibration was studied. This is an approach which retrieves the calibration from the image sequence only. Several new methods were proposed. These methods were validated on both real and synthetic data. The first method is a stratified approach which assumes constant calibration parameters during the acquisition of the images. A second method is more pragmatic and allows some parameters to vary. This approach makes it possible to deal with the zoom and focus available on most cameras.

The other important part of this work consisted of developing an automatic system for 3D acquisition from image sequences. This was achieved by combining, adapting and integrating several state-of-the-art algorithms with the newly developed self-calibration algorithms. The resulting system offers an unprecedented flexibility for the acquisition of realistic three-dimensional models. The visual quality of the models is very high. The metric qualities were verified through several validation experiments. This system was succesfully applied to a number of applications.

Notations

To enhance the readability the notations used throughout the text are summarized here.

For matrices bold face fonts are used (i.e. $\mathbf{A}$). 4-vectors are represented by $\mathtt{A}$ and 3-vectors by $\mathtt{a}$. Scalar values will be represented as a.

Unless stated differently the indices i, j and k are used for views, while l and m are used for indexing points, lines or planes. The notation $\mathbf{A}_{ij}$ indicates the entity $\mathbf{A}$ which relates view i to view j (or going from view i to view j). The indices i,j and k will also be used to indicate the entries of vectors, matrices and tensors. The subscripts P, A, M and E will refer to projective, affine, metric and Euclidean entities respectively

$\mathbf{P}$	camera projection matrix (3×4 matrix)
$\mathtt{M}$	world point (4-vector)
$\mathtt{\Pi}$	world plane (4-vector)
$\mathtt{m}$	image point (3-vector)
$\mathtt{l}$	image line (3-vector)
$\mathbf{H}_{ij}^{\mathtt{\Pi}}$	homography for plane $\mathtt{\Pi}$ from view i to view j (3×3 matrix)
$\mathbf{H}_{\mathtt{\Pi}i}$	homography from plane $\mathtt{\Pi}$ to image i (3×3 matrix)
$\mathbf{F}$	fundamental matrix (3×3 rank 2 matrix)
$\mathtt{e}_{ij}$	epipole (projection of projection center of viewpoint i into image j)
$\mathbf{T}$	trifocal tensor ($3 \times 3 \times 3$ tensor)
$\mathbf{K}$	calibration matrix (3×3 upper triangular matrix)
$\mathbf{R}$	rotation matrix
$\mathtt{\Pi}_\infty$	plane at infinity (canonical representation: $W = 0$)
Ω	absolute conic (canonical representation: $X^2 + Y^2 + Z^2 = 0$ and $W = 0$)
Ω^*	absolute dual quadric (4×4 rank 3 matrix)
ω_∞	absolute conic embedded in the plane at infinity (3×3 matrix)
ω_∞^*	dual absolute conic embedded in the plane at infinity (3×3 matrix)
ω	image of the absolute conic (3×3 matrices)
ω^*	dual image of the absolute conic (3×3 matrices)

$\sim$	equivalence up to scale ($A \sim B$
$\|\mathbf{A}\|_F$	indicates the Frobenius norm of
$\mathbf{F}(\mathbf{A})$	indicates the matrix $\mathbf{A}$ scaled to h (i.e. $\frac{\mathbf{A}}{\|\mathbf{A}\|_F}$)
$\mathbf{A}^\top$	is the transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	is the inverse of $\mathbf{A}$ (i.e. $\mathbf{A}\mathbf{A}^{-1} =$
$\mathbf{A}^\dagger$	is the Moore-Penrose pseudo inve

Contents

1 Introduction **1**
- 1.1 Scope of the work 1
- 1.2 3D models from images 2
- 1.3 Main contributions 6
- 1.4 Outline of the thesis 7

2 Projective geometry **9**
- 2.1 Introduction 9
- 2.2 Projective geometry 9
  - 2.2.1 The projective plane 10
  - 2.2.2 Projective 3-space 11
  - 2.2.3 Transformations 11
  - 2.2.4 Conics and quadrics 11
- 2.3 The stratification of 3D geometry 14
  - 2.3.1 Projective stratum 14
  - 2.3.2 Affine stratum 15
  - 2.3.3 Metric stratum 18
  - 2.3.4 Euclidean stratum 22
  - 2.3.5 Overview of the different strata 22
- 2.4 Conclusion 22

3 Camera model and multiple view geometry **25**
- 3.1 Introduction 25
- 3.2 The camera model 25
  - 3.2.1 A simple model 25
  - 3.2.2 Intrinsic calibration 27
  - 3.2.3 Camera motion 28
  - 3.2.4 The projection matrix 29
  - 3.2.5 Deviations from the camera model 31
- 3.3 Multi view geometry 33
  - 3.3.1 Two view geometry 33
  - 3.3.2 Three view geometry 36
  - 3.3.3 Multi view geometry 38

3.4 Conclusion 39

4 Self-calibration **41**
4.1 Introduction 41
4.2 Projective ambiguity 41
4.3 Calibration 43
  4.3.1 Scene knowledge 43
  4.3.2 Camera knowledge 44
4.4 Self-calibration 45
  4.4.1 General motions 45
  4.4.2 Restricted motions 52
  4.4.3 Critical motion sequences 55
4.5 Conclusion 60

5 Stratified self-calibration **61**
5.1 Introduction 61
5.2 The modulus constraint 62
5.3 Self-calibration from constant intrinsic parameters 63
  5.3.1 Finding the plane at infinity 63
  5.3.2 Finding the absolute conic and refining the calibration results . 64
  5.3.3 Stratified self-calibration algorithm 64
  5.3.4 Some simulations 65
  5.3.5 Experiments 69
5.4 Special cases 75
  5.4.1 Two images and two vanishing points 75
  5.4.2 Varying focal length 79
  5.4.3 Stereo rig with varying focal length 82
5.5 Conclusion 84

6 Flexible self-calibration **87**
6.1 Introduction 87
6.2 Some theory 88
  6.2.1 A counting argument 88
  6.2.2 A geometric interpretation of self-calibration constraints . . . 88
  6.2.3 Minimal constraints for self-calibration 90
6.3 Self-calibration of a camera with varying intrinsic parameters 92
  6.3.1 Non-linear approach 92
  6.3.2 Linear approach 94
  6.3.3 Maximum Likelihood approach 96
6.4 Critical motion sequences 97
  6.4.1 Critical motion sequences for varying focal length 98
  6.4.2 Detecting critical motion sequences 101
6.5 Constraint selection 102
6.6 Experiments 104
  6.6.1 Simulations 104

6.6.2 Castle sequence 106
6.6.3 Pillar sequence 108
6.7 Conclusions 113

7 Metric 3D Reconstruction 117
7.1 Introduction 117
7.2 Overview of the method 118
7.3 Projective reconstruction 119
7.3.1 Relating the images 119
7.3.2 Initial reconstruction 122
7.3.3 Adding a view 124
7.3.4 Relating to other views 126
7.4 Upgrading the reconstruction to metric 127
7.5 Dense depth estimation 128
7.5.1 Rectification 130
7.5.2 Dense stereo matching 130
7.5.3 Multi view matching 131
7.6 Building the model 132
7.7 Some implementation details 133
7.8 Some possible improvements 135
7.8.1 Interest point matching 135
7.8.2 Projective reconstruction 136
7.8.3 Surviving planar regions 136
7.8.4 Generalized rectification 138
7.9 Conclusion 141

8 Results and applications 143
8.1 Introduction 143
8.2 Acquisition of 3D models from photographs 143
8.3 Acquisition of 3D models from preexisting image sequences 148
8.4 Acquisition of plenoptic models 150
8.5 Virtualizing archaeological sites 153
8.5.1 Virtualizing scenes 154
8.5.2 Reconstructing an overview model 155
8.5.3 Reconstructions at different scales 158
8.5.4 Combination with other models 159
8.6 More applications in archaeology 159
8.6.1 3D stratigraphy 159
8.6.2 Generating and testing building hypotheses 160
8.7 Applications in other areas 160
8.7.1 Architecture and conservation 160
8.7.2 Other applications 161
8.8 Conclusion 162

9 Conclusion 163

9.1 Summary 163

9.2 Discussion and further research 164

A The modulus constraint 183

A.1 Derivation of the modulus constraint 183

A.2 Expressions for l_3, l_2, l_1 and l_0 184

A.3 Expressions for l_3, l_2, l_1 and l_0 for a varying focal length 185

B Self-calibration from rectangular pixels 187

C Planar sections of imaginary cones 189

D Generalized rectification 191

D.1 Oriented epipolar geometry 191

D.2 Generalized Rectification 194

Chapter 1

Introduction

1.1 Scope of the work

Obtaining three dimensional (3D) models of scenes from images has been a long lasting research topic in computer vision. Many applications exist which require these models. Traditionally robotics and inspection applications were considered. In these cases accuracy was often the main concern. In this case expensive devices working only under controlled circumstances were the typical solutions that were used.

Nowadays however more and more interest comes from the multimedia and computer graphics communities. The evolution of computers is such that today even personal computers can display complex 3D models. Many computer games are located in large 3D worlds. The use of 3D models and environments on the Internet is becoming common practice. This evolution is however slowed down due to the difficulties of generating such 3D models. Although it is easy to generate simple 3D models, complex scenes are requiring a lot of effort. Furthermore existing objects or scenes are often considered. In these cases the effort required to recreate realistic 3D models is often prohibitive and the results are often disappointing.

A growing demand exists for systems which can *virtualize* existing objects or scenes. In this case the requirements are very different from the requirements encountered in previous applications. Most important is the visual quality of the 3D models. Also the boundary constraints are different. There is an important demand for easy acquisition procedures using off-the-shelf consumer products. This explains the success of the Quicktime VR technology which combines easy acquisition with fast rendering on low-end machines. Note however that in this case no 3D is extracted and that it is therefore only possible to *look around* and not to *walk around*.

In this dissertation it was investigated how far the limits of automatic acquisition of realistic 3D models could be pushed towards easy and flexible acquisition procedures. This has been done by developing a system which obtains dense metric 3D surface models from sequences of images taken with a hand-held camera. Due to the limitation in time of this project some choices had to be made. The system was built by combining existing state-of-the-art algorithms with new components developed within this project. Some of these components where extended or adapted to fit in the system.

In the research community a lot of effort was put in obtaining the calibration of the camera setup up to an arbitrary projective transformation from the images only and since many years a lot of work had been done to obtain dense correspondence maps for calibrated camera setups. There was however a missing link. Although the possibility of self-calibration (i.e. restricting the ambiguity on the calibration from projective to metric) had been shown, practical algorithms were not giving satisfying results. Additionally, existing theory and algorithms were restricted to constant camera parameters prohibiting the use of zoom and focusing capabilities available on most cameras.

In this context I decided to concentrate on the self-calibration aspect. Algorithms working well on real image sequences were required. It seemed also useful to investigate the possibilities of allowing varying camera parameters, especially a varying focal length so that the system could cope with zoom and focus.

1.2 3D models from images

In this section an overview of the literature is given. Some related research published after the start of this work is also discussed, but will be indicated as *recent* work. This is important since the evolution in some of the discussed areas has been impressive in the last few years.

Obtaining 3D models from image sequences is a difficult task. This comes from the fact that only very little information is available to start with. Both the scene geometry and the camera geometry are assumed unknown. Only very general assumptions are made, e.g. rigid scene, piecewise continuous surfaces, mainly diffuse reflectance characteristics for the scene and a pinhole camera model for the camera.

The general approach consists of separating the problem in a number of more manageable subproblems, which can then be solved by separate modules. Often interaction or feed-back is needed between these modules to extract the necessary information from the images. Certainly for the first steps when almost no information has been extracted, feedback is very important to verify the obtained hypotheses. Gradually more information and more certainty about this information is obtained. At later stages the coupling between the modules is less important although it can still improve the quality of the results.

Feature matching The first problem is the *correspondence problem*: Given a feature in an image, what is the corresponding feature (i.e. the projection of the same 3D feature) in the other image? This is an ill-posed problem and therefore is often very hard to solve. When some assumptions are satisfied, it is possible to automatically match points or other features between images. One of the most useful assumptions is that the images are not too different (i.e. same illumination, similar pose). In this case the coordinates of the features and the intensity distribution around the feature are similar in both images. This allows to restrict the search range and to match features through intensity cross-correlation.

It is clear that not all possible image features are suitable for matching. Often points are used since they are most easily handled by the other modules, but line segments [140, 23] or other features (such as regions) can also be matched. It is clear that not all points are suited for matching. Many points can be located in homogeneous regions were almost no information is available to differentiate between them. It is therefore important to use an interest point detector which extracts a certain number of points useful for matching. These points should clearly satisfy two criteria. The extraction of the points should be as much as possible independent of camera pose and illumination changes and the neighborhood of the selected points should contain as much information as possible to allow matching. Many interest point detectors exist (e.g. Harris [50], Deriche [24] or Förstner[43]). In [141] Schmid et al. concluded that the Harris corner detector gives the best results according to the two criteria mentioned above.

In fact the feature matching is often tightly coupled with the structure from motion estimation described in the next paragraph. Hypothetical matches are used to compute the scene and camera geometry. The obtained results are then used to drive the feature matching.

Structure from motion Researchers have been working for many years on the automatic extraction of 3D structure from image sequences. This is called the *structure from motion* problem: Given an image sequence of a rigid scene by a camera undergoing unknown motion, reconstruct the 3D geometry of the scene. To achieve this, the camera motion also has to be recovered simultaneously. When in addition the camera calibration is unknown as well, one speaks of *uncalibrated* structure from motion.

Early work on (calibrated) structure from motion focused on the two view problem [84, 168]. Starting from a set of corresponding features in the images, the camera motion and 3D scene structure could be recovered. Since then the research has shifted to the more difficult problem of longer image sequences. This allows to retrieve the scene geometry more accurately by taking advantage of redundancy. Some of the representative approaches are due to Azerbayejani et al. [5], Cui et al. [21], Spetsakis and Aloimonos [148] and Szeliski and Kang [156]. These methods make use of a full perspective camera model. Tomasi and Kanade [159] proposed a factorization approach based on the affine camera model (see [103] for more recent work). Recently Jacobs proposed a factorization method able to deal with missing data [68].

Uncalibrated structure from motion In this case the structure of the scene can only be recovered up to an arbitrary projective transformation. In the two view case the early work was done by Faugeras [36] and Hartley [51]. They obtained the fundamental matrix as an equivalent for the essential matrix. This matrix completely describes the projective structure of the two view geometry.

Since then many algorithms have been proposed to compute the fundamental matrix from point matches [57, 86, 13, 104]. Based on these methods, robust approaches were developed to obtain the fundamental matrix from real image data (see the work of Torr et al. [162, 163] and Zhang et al. [186]). These techniques use robust techniques like RANSAC [40] or LMedS [136] and feedback the results to the matcher to obtain more matches. These are then used to refine the solution.

A similar entity can also be obtained for three views. This is called the trifocal tensor. It describes the transfer for points (see Shashua [144]), lines (see Hartley [54]) or both (see Hartley [56]). In fact these trilinearities had already been discovered by Spetsakis and Aloimonos [148] in the calibrated case. Robust computation methods were also developed for the trifocal tensor (e.g. Torr and Zisserman [160, 161]). The properties of this tensor have been carefully studied (e.g. Shashua and Avidan [145]).

Relationships between more views have also been studied (see the work of Heyden [62], Triggs [165] and Faugeras and Mourrain [39]). See [98] for a recent tutorial on the subject. Recently Hartley [59] proposed a practical computation method for the quadrifocal tensor.

Up to now no equivalent solution to the factorization approach of Tomasi and Kanade [159] has been found for the uncalibrated structure from motion problem. Some ideas have been proposed [63, 154], but these methods are iterative or require part of the solution to start with. Another possibility consists of carrying out a nonlinear minimization over all the unknown parameters at once. This was for example proposed in the early paper of Mohr [96]. This is in fact the projective equivalent of what photogrammetrists call bundle adjustment [147]. A description of an efficient algorithm can be found in [79]. Bundle adjustment however requires a good initialization to start with.

The traditional approach for uncalibrated structure from motion sequences consists of putting up a reference frame from the two first views and then sequentially adding new views (see Beardsley et al. [9, 8] or also [79]). Recently a hierarchical approach has been proposed by Fitzgibbon and Zisserman [41] that builds up relations between image pairs, triplets, subsequences and finally the whole sequence.

Self-calibration Since projective structure is often not sufficient, researchers tried to develop methods to recover the metric structure of scenes obtained through uncalibrated structure from motion. The most popular approach consists of using some constraints on the intrinsic camera parameters of the camera. This is called self-calibration. In general fixed intrinsic camera parameters are assumed.

The first approach was proposed by Maybank and Faugeras [95] (see also [36]). It is based on the Kruppa equations [77]. The method was developed further in [87] and recently by Zeller in [183, 184]. This method only requires a pairwise calibration (i.e. the epipolar geometry). It uses the concept of the absolute conic which –besides the plane at infinity– is the only fixed entity for the group of Euclidean transformations [35]. Most other methods are also based on this concept of the absolute conic.

Hartley [53] proposed an alternative approach which obtains the metric calibration by minimizing the difference between the intrinsic camera parameters one tries to compute and the ones obtained through factorization of the camera projection matrices. A quasi-affine reconstruction is used as initialization. This initial reconstruction is obtained from the constraint that all 3D points seen in a view must be in front of the camera [52].

Since a few years several new methods have been proposed. Some of these are part of this work and will be presented in detail further on. Some other methods were developed in parallel. Triggs proposed a method based on the absolute (dual) quadric [166]. This is a disc-quadric (of planes) which encodes both the absolute conic and the plane at infinity. Heyden and Åström proposed a similar method [60].

Some researchers tried to take advantage of restricted motions to obtain simpler algorithms. Moons et al. [100, 99] designed a simple algorithm to obtain affine structure from a purely translating camera. Armstrong [2] incorporated this in a stratified approach to self-calibration. Hartley [54] proposed a self-calibration method for a purely rotating camera. Recently algorithms for planar motion were proposed by Armstrong, Zisserman and Hartley [3] and by Faugeras, Quan and Sturm [33]. In some of these cases ambiguities on the reconstruction exist. Zisserman et al. [189] recently proposed some ways to reduce this ambiguity by imposing, a posteriori, some constraints on the intrinsic parameters.

In some cases the motion is not general enough to allow for complete self-calibration. Recently Sturm established a complete catalogue of critical motion sequences for the case of constant intrinsic camera parameters [152, 150].

A more in-depth discussion of some existing self-calibration methods is given in Chapter 4.

Dense stereo matching The structure from motion algorithms only extract a restricted number of features. Although textured 3D models have been generated from this, the results are in general not very convincing. Often some important scene features are missed during matching resulting in incomplete models. Even when all important features are obtained the resulting models are often dented.

However once the structure from motion problem has been solved, the pose of the camera is known for all the views. In this case correspondence matching is simpler (since the epipolar geometry is known) and existing stereo matching algorithms can be used. This then allows to obtain a dense 3D surface model of the scene.

Many approaches exist for stereo matching. These approaches can be broadly classified into feature- and correlation-based approaches [29]. Some important feature based approaches were proposed by Marr and Poggio [89], Grimson [46], Pollard, Mayhew and Frisby [110] (all relaxation based methods), Gimmel’Farb [44] and Baker and Binford [6] and Ohta and Kanade [107] (using dynamic programming).

Successful correlation-based approaches were for example proposed by Okutomi and Kanade [108] or Cox et al. [20]. The latter was recently refined by Koch [72] and Falkenhagen [31, 30]. It is this last algorithm that is used in this work. Another approach based on optical flow was proposed by Proesmans et al. [134].

3D reconstruction systems It should be clear from the previous paragraphs that obtaining 3D models from an image sequence is not an easy task. It involves solving several complex subproblems.

One of the first systems was developed at CMU and is based on the Tomasi and Kanade factorization [159]. The “modeling from videotaping” approach however suffers from the restrictions of the affine camera model. In addition, since only matched feature points are used to generate the models the overall quality of the models is low.

The recent work of Beardsley et al. [9, 8] at Oxford was used in a similar way to obtain models. In this case the intrinsic camera parameters were assumed known as in the previous case. More recent work by Fitzgibbon and Zisserman [42] is using the self-calibration method described in Chapter 6 to deal with varying camera parameters.

Similar work was also done very recently at INRIA by Bougnoux [15, 14]. In this case, however, the system is an enhanced 3D modeling tool including algorithms for uncalibrated structure from motion and self-calibration. The correspondences of the 3D points which should be used for the model, however, have to be indicated by hand. The resulting models are therefore restricted to a limited number of planar patches.

Another recent approach developed by Debevec, Taylor and Malik [26, 158, 27] at Berkeley proved very successful in obtaining realistic 3D models from photographs. A mixed geometric- and image-based approach is used. The texture mapping is view dependent to enhance photorealism. An important restriction of this method is however the need for an approximate a priori model of the scene. This model is fitted semi-automatically to image features.

Shum, Han and Szeliski [146] recently proposed an interactive method for the construction of 3D models from panoramic images. In this case points, lines and planes are indicated in the panoramic image. By adding constraints on these entities (e.g. parallelism, coplanarity, verticality), 3D models can be obtained.

Finally some commercial systems exist which allow to generate 3D models from photographs (e.g. *PhotoModeler* [109]). These systems require a lot of interaction from the user (e.g. correspondences have to be indicated by hand) and some calibration information. The resulting models can be very realistic. It is however almost impossible to model complex shapes.

1.3 Main contributions

Before we enter a more detailed discussion of these topics, it seems useful to summarize the main contributions we believe are made through this work:

- A *stratified self-calibration* approach was proposed. Inspired by the successful stratified approaches for restricted motions, I have developed a similar approach for general motions. This was done based on a new constraint for self-calibration that I derived (i.e. *the modulus constraint*). This work was published in the papers [111, 123, 124, 122, 127] and technical reports [129, 130].

- An important contribution to the state-of-the-art was made by allowing *self-calibration in spite of varying camera parameters*. At first a restricted approach was derived which allowed the focal length to vary for single cameras [121, 131] and for stereo rigs [125, 126]. Later on, a general approach was proposed which could efficiently work with known, fixed and varying intrinsic camera parameters together with a pragmatic approach for a camera equipped with a zoom.

A theorem was derived which showed that for general motion sequences the minimal constraints that pixels are rectangular is sufficient to allow for self-calibration. This work was published in [112, 120] and in the technical report [128].

- A *complete system for automatic acquisition of metric 3D surface models from uncalibrated image sequences* was developed. The self-calibration techniques mentioned earlier were incorporated into this system allowing for an unprecedented flexibility in acquisition of 3D models from images. This was the first system to integrate uncalibrated structure from motion, self-calibration and dense stereo matching algorithms. This combination results in highly realistic 3D surface models obtained automatically from images taken with an uncalibrated hand-held camera, without restriction on zoom or focus. The complete system was described in [117, 115, 118, 119].

- The *acquisition flexibility* offered by this system makes new applications possible. As a test case our system was applied on a number of applications found in the area of archaeology and heritage preservation [116, 113, 114]. Some of these applications are only possible with a system as the one described in this dissertation.

1.4 Outline of the thesis

In Chapter 2 some basic concepts used throughout the text are presented. Projective geometry is introduced in this chapter. This is the natural mathematical framework to describe the projection of a scene onto an image. Some properties of transformations, conics and quadrics are given as well. Finally, the stratification of space in projective, affine, metric and Euclidean is described.

This introduction into the basic principles is continued in Chapter 3 where the camera model and image formation process are described. In this context some important multi-view relationships are also described.

Chapter 4 introduces the problem of self-calibration. The methods proposed by others are presented here, both general methods and methods requiring restricted motions. Some inherent problems or limitations of self-calibration are also discussed here.

In Chapter 5 a stratified approach to self-calibration is proposed. This approach is based on the modulus constraint. Some experiments compare this method to other state-of-the-art methods. Some additional applications of the modulus constraint are also given in this chapter.

Chapter 6 deals with self-calibration in the presence of varying intrinsic camera parameters. First, some theoretical considerations are presented. Then a flexible calibration method is proposed which can deal with known, fixed and varying intrinsic camera parameters. Based on this, a pragmatic approach is derived which works for a standard camera equipped with a zoom. Critical motion sequences are also discussed.

Chapter 7 presents the complete system for 3D model acquisition. Some problems with the actual system are also discussed and possible solutions are described.

In Chapter 8 results and applications of the system are presented. The system is applied to a highly sculptured temple surface, to old film footage, to the acquisition of plenoptic models, to an archaeological site and to some other examples. Hereby the flexibility and the potential of the approach is demonstrated.

The conclusions of our work are presented in Chapter 9. To enhance the readability of this text some more tedious derivations were placed in appendices.

Chapter 2

Projective geometry

. . . ὼστϵ καλλιον άποδϵξϵσθαι, ίσμϵν πον ὸτι τω όλω και παντι διοισϵι ὴμμϵνοζ τϵ γϵωμϵτραιζ και μη

“. . . experience proves that anyone who has studied geometry is infinitely quicker to grasp difficult subjects than one who has not.”
Plato - The Republic, Book 7, 375 B.C.

2.1 Introduction

The work presented in this thesis draws a lot on concepts of projective geometry. This chapter and the next one introduce most of the geometric concepts used in the rest of the text. This chapter concentrates on projective geometry and introduces concepts as points, lines, planes, conics and quadrics in two or three dimensions. A lot of attention goes to the stratification of geometry in projective, affine, metric and Euclidean layers. Projective geometry is used for its simplicity in formalism, additional structure and properties can then be introduced were needed through this hierarchy of geometric strata. This section was inspired by the introductions on projective geometry found in Faugeras’ book [34], in the book by Mundy and Zisserman (in [105]) and by the book on projective geometry by Semple and Kneebone [142].

2.2 Projective geometry

A point in projective n-space, $\mathcal{P}^n$, is given by a $(n+1)$-vector of coordinates $\mathtt{x} = [x_1 \ldots x_{n+1}]^\top$. At least one of these coordinates should differ from zero. These coordinates are called *homogeneous* coordinates. In the text the coordinate vector and the point itself will be indicated with the same symbol. Two points represented by $(n+1)$-vectors $\mathtt{x}$ and $\mathtt{y}$ are equal if and only if there exists a nonzero scalar λ such that $x_i = \lambda y_i$, for every i $(1 \leq i \leq n+1)$. This will be indicated by $\mathtt{x} \sim \mathtt{y}$.

Often the points with coordinate $x_{n+1} = 0$ are said to be *at infinity*. This is related to the affine space $\mathcal{A}$. This concept is explained more in detail in section 2.3.

A *collineation* is a mapping between projective spaces, which preserves collinearity (i.e. collinear points are mapped to collinear points). A collineation from $\mathcal{P}^m$ to $\mathcal{P}^n$ is mathematically represented by a $(m+1) \times (n+1)$-matrix $\mathbf{H}$. Points are transformed linearly: $\mathtt{x} \mapsto \mathtt{x}' \sim \mathbf{H}\mathtt{x}$. Observe that matrices $\mathbf{H}$ and $\lambda\mathbf{H}$ with λ a nonzero scalar represent the same collineation.

A *projective basis* is the extension of a coordinate system to projective geometry. A projective basis is a set of $n+2$ points such that no $n+1$ of them are linearly dependent. The set $\mathtt{e}_l = [0 \ldots 1 \ldots 0]^\top$ for every l $(1 \leq l \leq n+1)$, where 1 is in the lth position and $\mathtt{e}_{n+2} = [11 \ldots 1]^\top$ is the standard projective basis. A projective point of $\mathcal{P}^n$ can be described as a linear combination of any $n+1$ points of the standard basis. For example:

$$\mathtt{m} = \sum_{l=1}^{n+1} \lambda_l \mathtt{e}_l$$

It can be shown [36] that any projective basis can be transformed via a uniquely determined collineation into the standard projective basis. Similarly, if two set of points $\mathtt{m}_1, \ldots, \mathtt{m}_{n+2}$ and $\mathtt{m}'_1, \ldots, \mathtt{m}'_{n+2}$ both form a projective basis, then there exists a uniquely determined collineation $\mathbf{T}$ such that $\mathtt{m}'_l \sim \mathbf{T}\mathtt{m}_l$ for every l $(1 \leq l \leq n+2)$. This collineation $\mathbf{T}$ describes the change of projective basis. In particular, $\mathbf{T}$ is invertible.

2.2.1 The projective plane

The projective plane is the projective space $\mathcal{P}^2$. A point of $\mathcal{P}^2$ is represented by a 3-vector $\mathtt{m} = [x\,y\,w]^\top$. A line $\mathtt{l}$ is also represented by a 3-vector. A point $\mathtt{m}$ is located on a line $\mathtt{l}$ if and only if

$$\mathtt{l}^\top \mathtt{m} = 0 \ . \tag{2.1}$$

This equation can however also be interpreted as expressing that the line $\mathtt{l}$ passes through the point $\mathtt{m}$. This symmetry in the equation shows that there is no formal difference between points and lines in the projective plane. This is known as the principle of *duality*. A line $\mathtt{l}$ passing through two points $\mathtt{m}_1$ and $\mathtt{m}_2$ is given by their vector product $\mathtt{m}_1 \times \mathtt{m}_2$. This can also be written as

$$\mathtt{l} \sim [\mathtt{m}_1]_\times \mathtt{m}_2 \text{ with } [\mathtt{m}_1]_\times = \begin{bmatrix} 0 & w_1 & -y_1 \\ -w_1 & 0 & x_1 \\ y_1 & -x_1 & 0 \end{bmatrix} \ . \tag{2.2}$$

The dual formulation gives the intersection of two lines. All the lines passing through a specific point form a *pencil of lines*. If two lines $\mathtt{l}_1$ and $\mathtt{l}_2$ are distinct elements of the pencil, all the other lines can be obtained through the following equation:

$$\mathtt{l} \sim \lambda_1 \mathtt{l}_1 + \lambda_2 \mathtt{l}_2 \tag{2.3}$$

for some scalars λ_1 and λ_2. Note that only the ratio $\frac{\lambda_1}{\lambda_2}$ is important.

2.2.2 Projective 3-space

Projective 3D space is the projective space $\mathcal{P}^3$. A point of $\mathcal{P}^3$ is represented by a 4-vector $\mathtt{M} = [X\,Y\,Z\,W]^\top$. In $\mathcal{P}^3$ the dual entity of a point is a plane, which is also represented by a 4-vector. A point $\mathtt{M}$ is located on a plane Π if and only if

$$\Pi^\top \mathtt{M} = 0 \ . \tag{2.4}$$

A line can be given by the linear combination of two points $\lambda_1 \mathtt{M}_1 + \lambda_2 \mathtt{M}_2$ or by the intersection of two planes $\Pi_1 \cap \Pi_2$.

2.2.3 Transformations

Transformations in the images are represented by *homographies* of $\mathcal{P}^2 \to \mathcal{P}^2$. A homography of $\mathcal{P}^2 \to \mathcal{P}^2$ is represented by a 3×3-matrix $\mathbf{H}$. Again $\mathbf{H}$ and $\lambda\mathbf{H}$ represent the same homography for all nonzero scalars λ. A point is transformed as follows:

$$\mathtt{m} \mapsto \mathtt{m}' \sim \mathbf{H}\mathtt{m} \ . \tag{2.5}$$

The corresponding transformation of a line can be obtained by transforming the points which are on the line and then finding the line defined by these points:

$$\mathtt{l}'^\top \mathtt{m}' = \mathtt{l}^\top \mathbf{H}^{-1}\mathbf{H}\mathtt{m} = \mathtt{l}^\top \mathtt{m} = 0 \ . \tag{2.6}$$

From the previous equation the transformation equation for a line is easily obtained (with $\mathbf{H}^{-\top} = (\mathbf{H}^{-1})^\top = (\mathbf{H}^\top)^{-1}$):

$$\mathtt{l} \mapsto \mathtt{l}' \sim \mathbf{H}^{-\top}\mathtt{l} \tag{2.7}$$

Similar reasoning in $\mathcal{P}^3$ gives the following equations for transformations of points and planes in 3D space:

$$\mathtt{M} \mapsto \mathtt{M}' \sim \mathbf{T}\mathtt{M}, \tag{2.8}$$

$$\Pi \mapsto \Pi' \sim \mathbf{T}^{-\top}\Pi \tag{2.9}$$

where $\mathbf{T}$ is a 4×4-matrix.

2.2.4 Conics and quadrics

Conic A *conic* in $\mathcal{P}^2$ is the locus of all points $\mathtt{m}$ satisfying a homogeneous quadratic equation:

$$S(\mathtt{m}) = \mathtt{m}^\top \mathbf{C}\mathtt{m} = 0 \, , \tag{2.10}$$

where $\mathbf{C}$ is a 3×3 symmetric matrix only defined up to scale. A conic thus depends on five independent parameters.

Dual conic Similarly, the dual concept exists for lines. A *conic envelope* or *dual conic* is the locus of all lines $\mathtt{l}$ satisfying a homogeneous quadratic equation:

$$\mathtt{l}^\top \mathbf{C}^* \mathtt{l} = 0\,, \tag{2.11}$$

where $\mathbf{C}^*$ is a 3×3 symmetric matrix only defined up to scale. A dual conic thus also depends on five independent parameters.

Line-conic intersection Let $\mathtt{m}$ and $\mathtt{m}'$ be two points defining a line. A point on this line can then be represented by $\mathtt{m} + \lambda\mathtt{m}'$. This point lies on a conic S if and only if

$$S(\mathtt{m} + \lambda\mathtt{m}') = 0\,,$$

which can also be written as

$$S(\mathtt{m}) + 2\lambda S(\mathtt{m}, \mathtt{m}') + \lambda^2 S(\mathtt{m}')\,, \tag{2.12}$$

where

$$S(\mathtt{m}, \mathtt{m}') = \mathtt{m}^\top \mathbf{C}\mathtt{m}' = S(\mathtt{m}', \mathtt{m})$$

This means that a line has in general two intersection points with a conic. These intersection points can be real or complex and can be obtained by solving equation (2.12).

Tangent to a conic The two intersection points of a line with a conic coincide if the discriminant of equation (2.12) is zero. This can be written as

$$S(\mathtt{m}, \mathtt{m}') - S(\mathtt{m})S(\mathtt{m}') = 0 \ \ .$$

If the point $\mathtt{m}$ is considered fixed, this forms a quadratic equation in the coordinates of $\mathtt{m}'$ which represents the two tangents from $\mathtt{m}$ to the conic. If $\mathtt{m}$ belongs to the conic, $S(\mathtt{m}) = 0$ and the equation of the tangents becomes

$$S(\mathtt{m}, \mathtt{m}') = \mathtt{m}^\top \mathbf{C}\mathtt{m}' = 0 \ \ ,$$

which is linear in the coefficients of $\mathtt{m}'$. This means that there is only one tangent to the conic at a point of the conic. This tangent $\mathtt{l}$ is thus represented by :

$$\mathtt{l} \sim \mathbf{C}^\top \mathtt{m} = \mathbf{C}\mathtt{m} \tag{2.13}$$

Relation between conic and dual conic When $\mathtt{m}$ varies along the conic, it satisfies $\mathtt{m}^\top \mathbf{C}\mathtt{m}$ and thus the tangent line $\mathtt{l}$ to the conic at $\mathtt{m}$ satisfies $\mathtt{l}^\top \mathbf{C}^{-1} \mathtt{l} = 0$. This shows that the tangents to a conic $\mathbf{C}$ are belonging to a dual conic $\mathbf{C}^* \sim \mathbf{C}^{-1}$ (assuming $\mathbf{C}$ is of full rank).

Transformation of a conic/dual conic The transformation equations for conics and dual conics under a homography $\mathbf{H}$ can be obtained in a similar way to Section 2.2.3. Using equations (2.5) and (2.7) the following is obtained:

$$\begin{array}{rcl} \mathtt{m}'^{\top}\mathbf{C}'\mathtt{m}' & \sim & \mathtt{m}^{\top}\mathbf{H}^{\top}\mathbf{H}^{-\top}\mathbf{C}\mathbf{H}^{-1}\mathbf{H}\mathtt{m} = 0\,, \\ \mathtt{l}'^{\top}\mathbf{C}^{*\prime}\mathtt{l}' & \sim & \mathtt{l}^{\top}\mathbf{H}^{-1}\mathbf{H}\mathbf{C}^{*}\mathbf{H}^{\top}\mathbf{H}^{-\top}\mathtt{l} = 0\,, \end{array}$$

and thus

$$\mathbf{C} \mapsto \mathbf{C}' \sim \mathbf{H}^{-\top}\mathbf{C}\mathbf{H}^{-1} \tag{2.14}$$

$$\mathbf{C}^{*} \mapsto \mathbf{C}^{*\prime} \sim \mathbf{H}\mathbf{C}^{*}\mathbf{H}^{\top} \tag{2.15}$$

Observe that (2.14) and (2.15) also imply that $(\mathbf{C}')^{*} = (\mathbf{C}^{*})'$.

Quadric In projective 3-space $\mathcal{P}^3$ similar concepts exist. These are quadrics. A *quadric* is the locus of all points $\mathtt{M}$ satisfying a homogeneous quadratic equation:

$$\mathtt{M}^{\top}\mathbf{Q}\mathtt{M} = 0\,, \tag{2.16}$$

where $\mathbf{Q}$ is a 4×4 symmetric matrix only defined up to scale. A quadric thus depends on nine independent parameters.

Dual quadric Similarly, the dual concept exists for planes. A *dual quadric* is the locus of all planes Π satisfying a homogeneous quadratic equation:

$$\Pi^{\top}\mathbf{Q}^{*}\Pi = 0 \tag{2.17}$$

where $\mathbf{Q}^{*}$ is a 3×3 symmetric matrix only defined up to scale and thus also depends on nine independent parameters.

Tangent to a quadric Similar to equation (2.13), the tangent plane Π to a quadric $\mathbf{Q}$ through a point $\mathtt{M}$ of the quadric is obtained as

$$\Pi = \mathbf{Q}\mathtt{M}\ . \tag{2.18}$$

Relation between quadric and dual quadric When $\mathtt{M}$ varies along the quadric, it satisfies $\mathtt{M}^{\top}\mathbf{Q}\mathtt{M}$ and thus the tangent plane Π to $\mathbf{Q}$ at $\mathtt{M}$ satisfies $\Pi^{\top}\mathbf{Q}^{-1}\Pi = 0$. This shows that the tangent planes to a quadric $\mathbf{Q}$ are belonging to a dual quadric $\mathbf{Q}^{*} \sim \mathbf{Q}^{-1}$ (assuming $\mathbf{Q}$ is of full rank).

Transformation of a quadric/dual quadric The transformation equations for quadrics and dual quadrics under a homography $\mathbf{T}$ can be obtained in a similar way to Section 2.2.3. Using equations (2.8) and (2.9) the following is obtained

$$\begin{array}{rcl} \mathtt{M}'^{\top}\mathbf{Q}'\mathtt{M}' & \sim & \mathtt{M}^{\top}\mathbf{T}^{\top}\mathbf{T}^{-\top}\mathbf{Q}\mathbf{T}^{-1}\mathbf{T}\mathtt{M} = 0 \\ \Pi'^{\top}\mathbf{Q}^{*\prime}\Pi' & \sim & \Pi^{\top}\mathbf{T}^{-1}\mathbf{T}\mathbf{Q}^{*}\mathbf{T}^{\top}\mathbf{T}^{-\top}\Pi = 0 \end{array}$$

and thus

$$\mathbf{Q} \mapsto \mathbf{Q}' \sim \mathbf{T}^{-\top}\mathbf{Q}\mathbf{T}^{-1} \tag{2.19}$$

$$\mathbf{Q}^* \mapsto \mathbf{Q}^{*\prime} \sim \mathbf{T}\mathbf{Q}^*\mathbf{T}^{\top} \tag{2.20}$$

Observe again that $(\mathbf{Q}')^* = (\mathbf{Q}^*)'$.

2.3 The stratification of 3D geometry

Usually the world is perceived as a Euclidean 3D space. In some cases (e.g. starting from images) it is not possible or desirable to use the full Euclidean structure of 3D space. It can be interesting to only deal with the more restricted and thus simpler structure of projective geometry. An intermediate layer is formed by the affine geometry. These structures can be thought of as different geometric strata which can be overlaid on the world. The simplest being projective, then affine, next metric and finally Euclidean structure.

This concept of stratification is closely related to the groups of transformations acting on geometric entities and leaving invariant some properties of configurations of these elements. Attached to the projective stratum is the group of projective transformations, attached to the affine stratum is the group of affine transformations, attached to the metric stratum is the group of similarities and attached to the Euclidean stratum is the group of Euclidean transformations. It is important to notice that these groups are subgroups of each other, e.g. the metric group is a subgroup of the affine group and both are subgroups of the projective group.

An important aspect related to these groups are their invariants. An *invariant* is a property of a configuration of geometric entities that is not altered by any transformation belonging to a specific group. Invariants therefore correspond to the measurements that one can do considering a specific stratum of geometry. These invariants are often related to geometric entities which stay unchanged – at least as a whole – under the transformations of a specific group. These entities will play a very important role in this text. Recovering them allows to upgrade the structure of the geometry to a higher level of the stratification.

In the following paragraphs the different strata of geometry are discussed. The associated groups of transformations, their invariants and the corresponding invariant structures are presented. This idea of stratification can be found back in [142] and [35].

2.3.1 Projective stratum

The first stratum is the projective one. It is the less structured one and has therefore the least number of invariants and the largest group of transformations associated with it. The group of projective transformations or collineations is the most general group of linear transformations.

As seen in the previous chapter a projective transformation of 3D space can be represented by a 4×4 invertible matrix

$$\mathbf{T}_P \sim \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \tag{2.21}$$

This transformation matrix is only defined up to a nonzero scale factor and has therefore 15 degrees of freedom.

Relations of incidence, collinearity and tangency are projectively invariant. The cross-ratio is an invariant property under projective transformations as well. It is defined as follows: Assume that the four points $\mathtt{M}_1, \mathtt{M}_2, \mathtt{M}_3$ and $\mathtt{M}_4$ are collinear. Then they can be expressed as $\mathtt{M}_i = \mathtt{M} + \lambda_i \mathtt{M}'$ (assume none is coincident with $\mathtt{M}'$). The cross-ratio is defined as

$$\{\mathtt{M}_1, \mathtt{M}_2; \mathtt{M}_3, \mathtt{M}_4\} = \frac{\lambda_1 - \lambda_3}{\lambda_1 - \lambda_4} : \frac{\lambda_2 - \lambda_3}{\lambda_2 - \lambda_4} \ . \tag{2.22}$$

The cross-ratio is not depending on the choice of the reference points $\mathtt{M}$ and $\mathtt{M}'$ and is invariant under the group of projective transformations of $\mathcal{P}^3$. A similar cross-ratio invariant can be derived for four lines intersecting in a point or four planes intersecting in a common line.

The cross-ratio can in fact be seen as the coordinate of a fourth point in the basis of the first three, since three points form a basis for the projective line $\mathcal{P}^1$. Similarly, two invariants could be obtained for five coplanar points; and, three invariants for six points, all in general position.

2.3.2 Affine stratum

The next stratum is the affine one. In the hierarchy of groups it is located in between the projective and the metric group. This stratum contains more structure than the projective one, but less than the metric or the Euclidean strata. Affine geometry differs from projective geometry by identifying a special plane, called the *plane at infinity*.

This plane is usually defined by $W = 0$ and thus $\Pi_\infty = [0\,0\,0\,1]^\top$. The projective space can be seen as containing the affine space under the mapping $\mathcal{A}^3 \rightarrow \mathcal{P}^3 : [X\,Y\,Z]^\top \mapsto [X\,Y\,Z\,1]^\top$. This is a one-to-one mapping. The plane $W = 0$ in $\mathcal{P}^3$ can be seen as containing the limit points for $\|\mathtt{M}\| \rightarrow \infty$, since these points are $[\frac{X}{\|\mathtt{M}\|}\,\frac{Y}{\|\mathtt{M}\|}\,\frac{Z}{\|\mathtt{M}\|}\,\frac{1}{\|\mathtt{M}\|}]^\top \sim [X_\infty\,Y_\infty\,Z_\infty\,0]$. This plane is therefore called the plane at infinity Π_∞. Strictly speaking, this plane is not part of the affine space, the points contained in it can't be expressed through the usual non-homogeneous 3-vector coordinate notation used for affine, metric and Euclidean 3D space.

An *affine transformation* is usually presented as follows:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix} \text{ with } \det(a_{ij}) \neq 0$$

Using homogeneous coordinates, this can be rewritten as follows $\mathtt{M}' \sim \mathbf{T}_A \mathtt{M}$ with

$$\mathbf{T}_A \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2.23}$$

An affine transformation counts 12 independent degrees of freedom. It can easily be verified that this transformation leaves the plane at infinity Π_∞ unchanged (i.e. $\Pi_\infty \sim \mathbf{T}_A^{-\top} \Pi_\infty$ or $\mathbf{T}_A^\top \Pi_\infty \sim \Pi_\infty$). Note, however, that the position of points in the plane at infinity can change under an affine transformation, but that all these points stay within the plane Π_∞.

All projective properties are a fortiori affine properties. For the (more restrictive) affine group parallelism is added as a new invariant property. Lines or planes having their intersection in the plane at infinity are called *parallel*. A new invariant property for this group is the *ratio of lengths along a certain direction*. Note that this is equivalent to a cross-ratio with one of the points at infinity.

From projective to affine Up to now it was assumed that these different strata could simply be overlaid onto each other, assuming that the plane at infinity is at its canonical position (i.e. $\Pi_\infty = [0\,0\,0\,1]^\top$). This is easy to achieve when starting from a Euclidean representation. Starting from a projective representation, however, the structure is only determined up to an arbitrary projective transformation. As was seen, these transformations do – in general – not leave the plane at infinity unchanged.

Therefore, in a specific projective representation, the plane at infinity can be anywhere. In this case upgrading the geometric structure from projective to affine implies that one first has to find the position of the plane at infinity in the particular projective representation under consideration.

This can be done when some affine properties of the scene are known. Since parallel lines or planes are intersecting in the plane at infinity, this gives constraints on the position of this plane. In Figure 2.1 a projective representation of a cube is given. Knowing this is a cube, three vanishing points can be identified. The plane at infinity is the plane containing these 3 vanishing points.

Ratios of lengths along a line define the point at infinity of that line. In this case the points $\mathtt{M}_0$, $\mathtt{M}_1$, $\mathtt{M}_2$ and the cross-ratio $\{\mathtt{M}_1, \mathtt{M}_2; \mathtt{M}_0, \mathtt{M}_\infty\}$ are known, therefore the point $\mathtt{M}_\infty$ can be computed.

Once the plane at infinity Π_∞ is known, one can upgrade the projective representation to an affine one by applying a transformation which brings the plane at infinity to its canonical position. Based on (2.9) this equation should therefore satisfy

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \sim \mathbf{T}^{-\top} \Pi_\infty \text{ or } \mathbf{T}^\top \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \sim \Pi_\infty \tag{2.24}$$

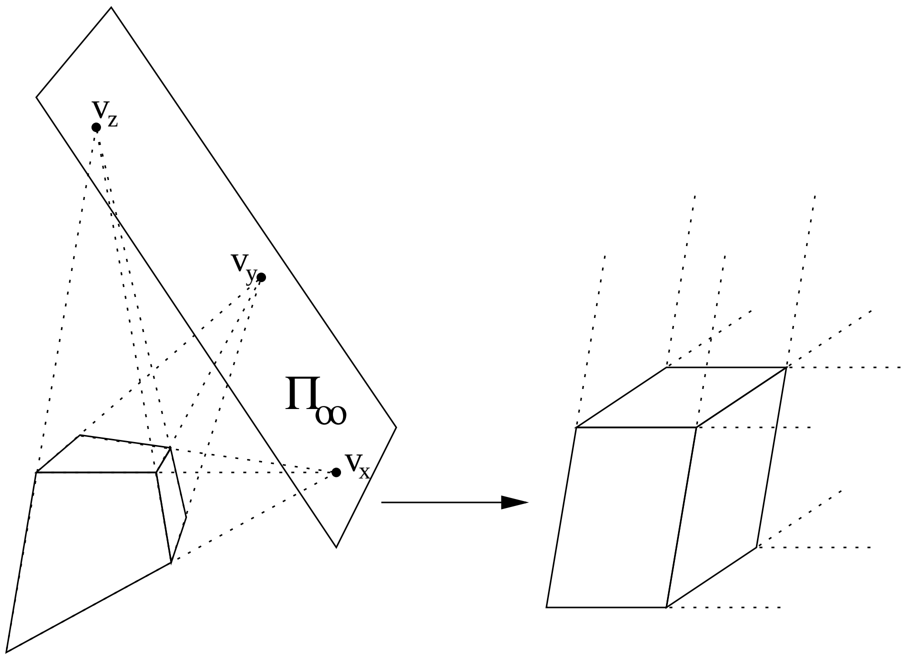

Figure 2.1: *Projective (left) and affine (right) structures which are equivalent to a cube under their respective ambiguities. The vanishing points obtained from lines which are parallel in the affine stratum constrain the position of the plane at infinity in the projective representation. This can be used to upgrade the geometric structure from projective to affine.*

This determines the fourth row of $\mathbf{T}$. Since, at this level, the other elements are not constrained, the obvious choice for the transformation is the following

$$\mathbf{T}_{PA} \sim \begin{bmatrix} \mathbf{I}_{3\times 3} & 0_3 \\ \pi_\infty^\top & 1 \end{bmatrix} \tag{2.25}$$

with π_∞ the first 3 elements of Π_∞ when the last element is scaled to 1. It is important to note, however, that every transformation of the form

$$\begin{bmatrix} \mathbf{A} & 0_3 \\ \pi_\infty^\top & 1 \end{bmatrix} \text{ with } \det \mathbf{A} \neq 0 \tag{2.26}$$

maps Π_∞ to $[0\,0\,0\,1]^\top$.

2.3.3 Metric stratum

The metric stratum corresponds to the group of similarities. These transformations correspond to Euclidean transformations (i.e. orthonormal transformation + translation) complemented with a scaling. When no absolute yardstick is available, this is the highest level of geometric structure that can be retrieved from images. This property is crucial for special effects since it enables the possibility to use scale models in movies.

A metric transformation can be represented as follows:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \sigma \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_{14} \\ t_{24} \\ t_{34} \end{bmatrix} \tag{2.27}$$

with r_{ij} the coefficients of an orthonormal matrix. The coefficients r_{ij} are related by 6 independent constraints $\sum_{k=1}^{3} r_{ik} r_{jk} = \delta_{ij}, (1 \leq i \leq j; 1 \leq j \leq 3)$ with δ_{ij} the Kronecker delta[1]. This corresponds to the matrix relation that $\mathbf{R}^\top\mathbf{R} = \mathbf{R}\mathbf{R}^\top = \mathbf{I}$ and thus $\mathbf{R}^{-1} = \mathbf{R}^\top$. Recall that $\mathbf{R}$ is a rotation matrix if and only if $\mathbf{R}\mathbf{R}^\top = \mathbf{I}$ and $\det\mathbf{R} = 1$. In particular, an orthonormal matrix only has 3 degrees of freedom. Using homogeneous coordinates, (2.27) can be rewritten as $\mathtt{M}' \sim \mathbf{T}_M \mathtt{M}$, with

$$\mathbf{T}_M \sim \begin{bmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_X \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_Y \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.28}$$

A metric transformation therefore counts 7 independent degrees of freedom, 3 for orientation, 3 for translation and 1 for scale.

In this case there are two important new invariant properties: *relative lengths* and *angles*. Similar to the affine case, these new invariant properties are related to an

---

[1] The Kronecker delta is defined as follows $\begin{cases} \delta_{ij} = 1 \text{ for } i = j \\ \delta_{ij} = 0 \text{ for } i \neq j \end{cases}$.

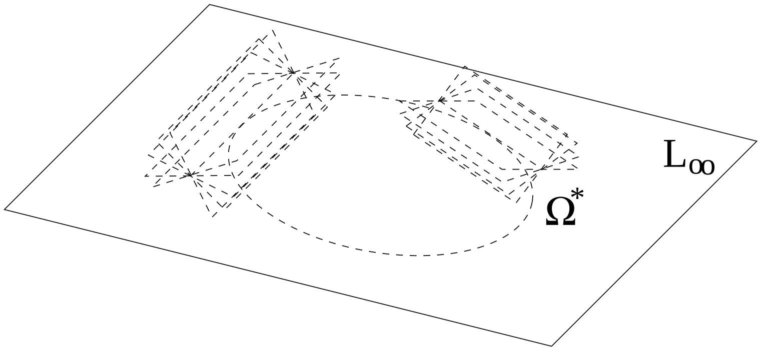

Figure 2.2: *The absolute conic* Ω *and the absolute dual quadric* Ω^* *in 3D space.*

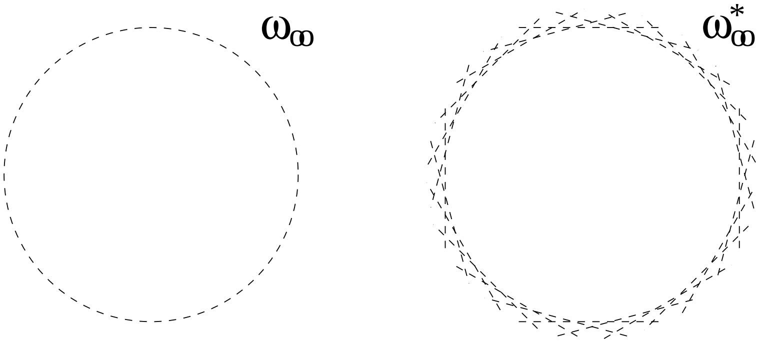

Figure 2.3: *The absolute conic* ω_∞ *and dual absolute conic* ω^*_∞ *represented in the purely imaginary part of the plane at infinity* Π_∞

invariant geometric entity. Besides leaving the plane at infinity unchanged similarity transformations also transform a specific conic into itself, i.e. the *absolute conic*. This geometric concept is more abstract than the plane at infinity. It could be seen as an imaginary circle located in the plane at infinity. In this text the absolute conic is denoted by Ω. It will be seen that it is often more practical to represent this entity in 3D space by its dual entity Ω^*. When only the plane at infinity is under consideration, ω_∞ and ω^*_∞ are used to represent the absolute conic and the dual absolute conic (these are 2D entities). Figure 2.2 and Figure 2.3 illustrate these concepts. The canonical form for the absolute conic Ω is:

$$\Omega : X^2 + Y^2 + Z^2 = 0 \text{ and } W = 0 \tag{2.29}$$

Note that two equations are needed to represent this entity. The associated dual entity, the absolute dual quadric Ω^*, however, can be represented as a single quadric. The canonical form is:

$$\Omega^* \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} . \tag{2.30}$$

Note that $\mathtt{\Pi}_\infty = [0\,0\,0\,1]^\top$ is the null space of Ω^*. Let $\mathtt{M}_\infty \sim [X\,Y\,Z\,0]^\top$ be a point of the plane at infinity, then that point in the plane at infinity is easily parameterized as $\mathtt{m}_\infty \sim [X\ Y\ Z]^\top$. In this case the absolute conic can be represented as a 2D conic:

$$\omega_\infty \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \omega^*_\infty \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{2.31}$$

According to (2.28), applying a similarity transformation to $\mathtt{M}_\infty$ results in $\mathtt{m}_\infty \mapsto \mathtt{m}'_\infty \sim \sigma \mathbf{R} \mathtt{m}_\infty$. Using equations (2.14),(2.15) and (2.20), it can now be verified that a similarity transformation leaves the absolute conic and its associated entities unchanged:

$$\begin{bmatrix} \mathbf{I}_{3\times3} & \mathtt{0}_3 \\ \mathtt{0}_3^\top & 0 \end{bmatrix} \sim \begin{bmatrix} \sigma\mathbf{R} & \mathtt{t} \\ \mathtt{0}_3^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3\times3} & \mathtt{0}_3 \\ \mathtt{0}_3^\top & 0 \end{bmatrix} \begin{bmatrix} \sigma\mathbf{R} & \mathtt{t} \\ \mathtt{0}_3^\top & 1 \end{bmatrix}^\top \tag{2.32}$$

and

$$\mathbf{I}_{3\times3} \sim \sigma^{-1}\mathbf{R}^{-\top}\mathbf{I}_{3\times3}\mathbf{R}^{-1}\sigma^{-1} \qquad\qquad \mathbf{I}_{3\times3} \sim \sigma\mathbf{R}\mathbf{I}_{3\times3}\mathbf{R}^\top\sigma \tag{2.33}$$

Inversely, it is easy to prove that the projective transformations which leave the absolute quadric unchanged form the group of similarity transformations (the same could be done for the absolute conic and the plane at infinity):

$$\begin{bmatrix} \mathbf{I}_{3\times3} & \mathtt{0}_3 \\ \mathtt{0}_3^\top & 0 \end{bmatrix} \sim \begin{bmatrix} \mathbf{A} & \mathtt{b} \\ \mathtt{c}^\top & d \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3\times3} & \mathtt{0}_3 \\ \mathtt{0}_3^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}^\top & \mathtt{c} \\ \mathtt{b}^\top & d \end{bmatrix} \sim \begin{bmatrix} \mathbf{A}\mathbf{A}^\top & \mathbf{A}\mathtt{c} \\ \mathtt{c}^\top\mathbf{A}^\top & \mathtt{c}^\top\mathtt{c} \end{bmatrix}$$

Therefore $\mathbf{A}\mathbf{A}^\top \sim \mathbf{I}_{3\times3}$ and $\mathtt{c} = \mathtt{0}_3$ which are exactly the constraints for a similarity transformation.

Angles can be measured using Laguerre's formula (see for example [142]). Assume two directions are characterized by their vanishing points $\mathtt{v}$ and $\mathtt{v}'$ in the plane at infinity (i.e. the intersection of a line with the plane at infinity indicating the direction). Compute the intersection points $\mathtt{j}$ and $\mathtt{j}'$ between the absolute conic and the line through the two vanishing points. The following formula based on the cross-ratio then gives the angle (with $i = \sqrt{-1}$):

$$\alpha = \frac{1}{2i} \log\{\mathtt{v}_1, \mathtt{v}_2; \mathtt{j}, \mathtt{j}'\} \tag{2.34}$$

From projective or affine to metric In some cases it is needed to upgrade the projective or affine representation to metric. This can be done by retrieving the absolute conic or one of its associated entities. Since the conic is located in the plane at infinity,

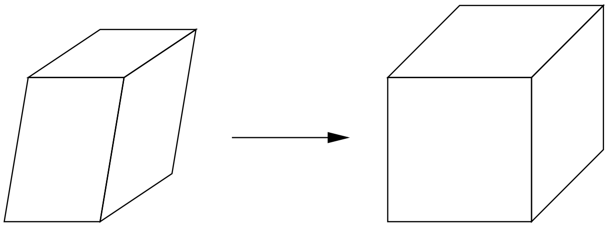

Figure 2.4: *Affine (left) and metric (right) representation of a cube. The right angles and the identical lengths in the different directions of a cube give enough information to upgrade the structure from affine to metric.*

it is easier to retrieve it once this plane has been identified (i.e. the affine structure has been recovered). It is, however, possible to retrieve both entities at the same time. The absolute quadric Ω^* is especially suited for this purpose, since it encodes both entities at once.

Every known angle or ratio of lengths imposes a constraint on the absolute conic. If enough constraints are at hand, the conic can uniquely be determined. In Figure 2.4 the cube of Figure 2.1 is further upgraded to metric (i.e. the cube is transformed so that obtained angles are orthogonal and the sides all have equal length).

Once the absolute conic has been identified, the geometry can be upgraded from projective or affine to metric by bringing it to its canonical (metric) position. In Section 2.3.2 the procedure to go from projective to affine was explained. Therefore, we can restrict ourselves here to the upgrade from affine to metric. In this case, there must be an affine transformation which brings the absolute conic to its canonical position; or, inversely, from its canonical position to its actual position in the affine representation. Combining (2.23) and (2.20) yields

$$\Omega^* \sim \begin{bmatrix} \mathbf{A} & \mathtt{a} \\ 0_3^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3\times 3} & 0_3 \\ 0_3^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A}^\top & 0_3 \\ \mathtt{a}^\top & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{A}^\top & 0_3 \\ 0_3^\top & 0 \end{bmatrix} \tag{2.35}$$

Under these circumstances the absolute conic and its dual have the following form (assuming the standard parameterization of the plane at infinity, i.e. $W = 0$):

$$\omega_\infty = \mathbf{A}^{-\top}\mathbf{A}^{-1} \text{ and } \omega_\infty^* = \mathbf{A}\mathbf{A}^\top \tag{2.36}$$

One possible choice for the transformation to upgrade from affine to metric is

$$\mathbf{T}_{AM} = \begin{bmatrix} \mathbf{A}^{-1} & 0_3 \\ 0_3^\top & 0 \end{bmatrix} \tag{2.37}$$

where a valid $\mathbf{A}$ can be obtained from Ω^* by Cholesky factorization. Combining (2.25) and (2.37) the following transformation is obtained to upgrade the geometry from projective to metric at once

$$\mathbf{T}_{PM} = \mathbf{T}_{AM}\mathbf{T}_{PA} = \begin{bmatrix} \mathbf{A}^{-1} & 0_3 \\ \pi_\infty & 1 \end{bmatrix} \tag{2.38}$$

2.3.4 Euclidean stratum

For the sake of completeness, Euclidean geometry is briefly discussed. It does not differ much from metric geometry as we have defined it here. The difference is that the scale is fixed and that therefore not only relative lengths, but *absolute lengths* can be measured. Euclidean transformations have 6 degrees of freedom, 3 for orientation and 3 for translation. A Euclidean transformation has the following form

$$\mathbf{T}_E \sim \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_X \\ r_{21} & r_{22} & r_{23} & t_Y \\ r_{31} & r_{32} & r_{33} & t_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.39}$$

with r_{ij} representing the coefficients of an orthonormal matrix, as described previously. If $\mathbf{R}$ is a rotation matrix (i.e. det$\mathbf{R} = 1$) then, this transformation represents a rigid motion in space.

2.3.5 Overview of the different strata

The properties of the different strata are briefly summarized in Table 2.1 . The different geometric strata are presented. The number of degrees of freedom, transformations and the specific invariants are given for each stratum. Figure 2.5 gives an example of an object which is equivalent to a cube under the different geometric ambiguities. Note from the figure that for purposes of visualization at least a metric level should be reached (i.e. is perceived as a cube).

2.4 Conclusion

In this chapter some concepts of projective geometry were presented. These will allow us, in the next chapter, to described the projection from a scene into an image and to understand the intricate relationships which relate multiple views of a scene. Based on these concepts methods can be conceived that inverse this process and obtain 3D reconstructions of the observed scenes. This is the main subject of this thesis.

ambiguity	DOF	transformation	invariants
projective	15	$\mathbf{T}_P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$	cross-ratio
affine	12	$\mathbf{T}_A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	relative distances along direction parallelism *plane at infinity*
metric	7	$\mathbf{T}_M = \begin{bmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_x \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_y \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	relative distances angles *absolute conic*
Euclidean	6	$\mathbf{T}_E = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$	absolute distances

Table 2.1: *Number of degrees of freedom, transformations and invariants corresponding to the different geometric strata (the coefficients r_{ij} form orthonormal matrices)*

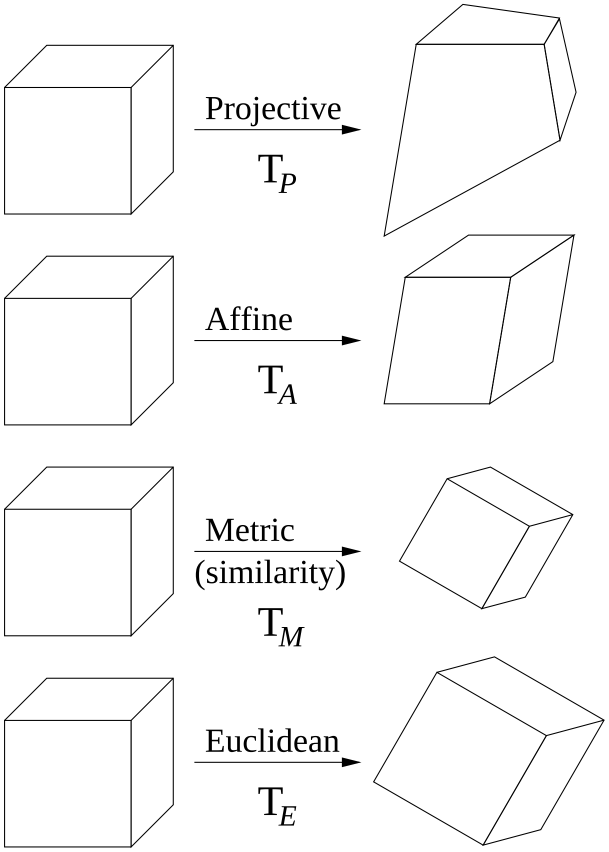

Figure 2.5: *Shapes which are equivalent to a cube for the different geometric ambiguities*

Chapter 3

Camera model and multiple view geometry

3.1 Introduction

Before discussing how 3D information can be obtained from images it is important to know how images are formed. First, the camera model is introduced; and then some important relationships between multiple views of a scene are presented.

3.2 The camera model

In this work the perspective camera model is used. This corresponds to an ideal pinhole camera. The geometric process for image formation in a pinhole camera has been nicely illustrated by Dürer (see Figure 3.1). The process is completely determined by choosing a perspective projection center and a retinal plane. The projection of a scene point is then obtained as the intersection of a line passing through this point and the center of projection $\mathtt{C}$ with the retinal plane $\mathcal{R}$.

Most cameras are described relatively well by this model. In some cases additional effects (e.g. radial distortion) have to be taken into account (see Section 3.2.5).

3.2.1 A simple model

In the simplest case where the projection center is placed at the origin of the world frame and the image plane is the plane $Z = 1$, the projection process can be modeled as follows:

$$x = \frac{X}{Z} \quad y = \frac{Y}{Z} \tag{3.1}$$

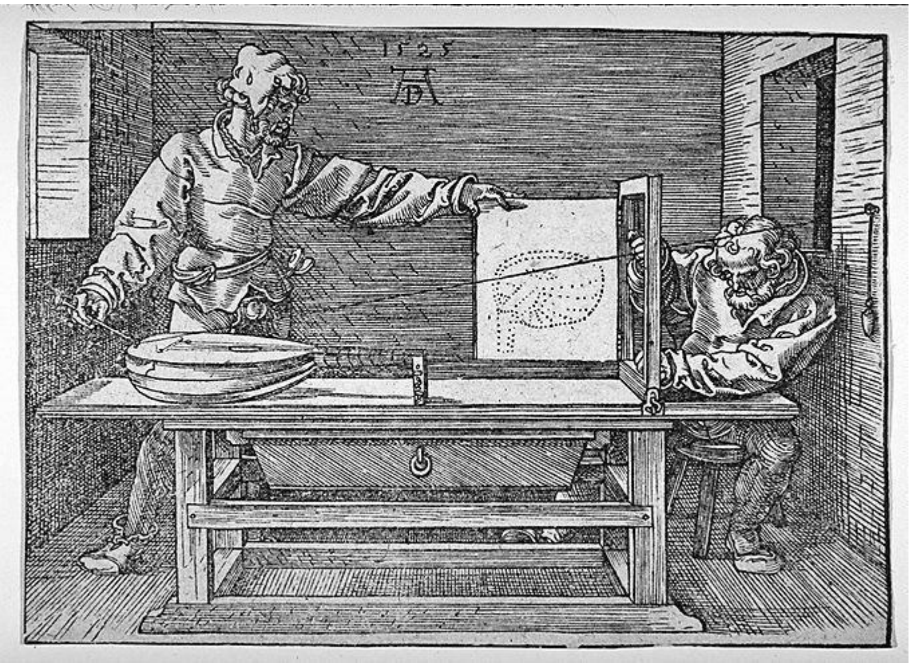

Figure 3.1: *Man Drawing a Lute (The Draughtsman of the Lute), woodcut 1525, Albrecht Dürer.*

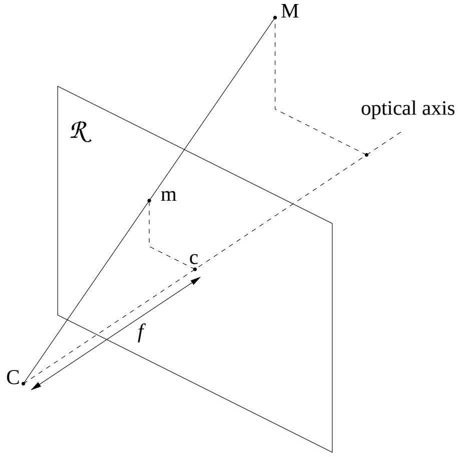

Figure 3.2: *Perspective projection*

For a world point (X, Y, Z) and the corresponding image point (x, y). Using the homogeneous representation of the points a linear projection equation is obtained:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.2}$$

This projection is illustrated in Figure 3.2. The optical axis passes through the center of projection `C` and is orthogonal to the retinal plane $\mathcal{R}$. It's intersection with the retinal plane is defined as the principal point `c`.

3.2.2 Intrinsic calibration

With an actual camera the focal length f (i.e. the distance between the center of projection and the retinal plane) will be different from 1, the coordinates of equation (3.2) should therefore be scaled with f to take this into account.

In addition the coordinates in the image do not correspond to the physical coordinates in the retinal plane. With a CCD camera the relation between both depends on the size and shape of the pixels and of the position of the CCD chip in the camera. With a standard photo camera it depends on the scanning process through which the images are digitized.

The transformation is illustrated in Figure 3.3. The image coordinates are obtained through the following equations:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{p_x} & (\tan\alpha)\frac{f}{p_y} & c_x \\ & \frac{f}{p_y} & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x_{\mathcal{R}} \\ y_{\mathcal{R}} \\ 1 \end{bmatrix}$$

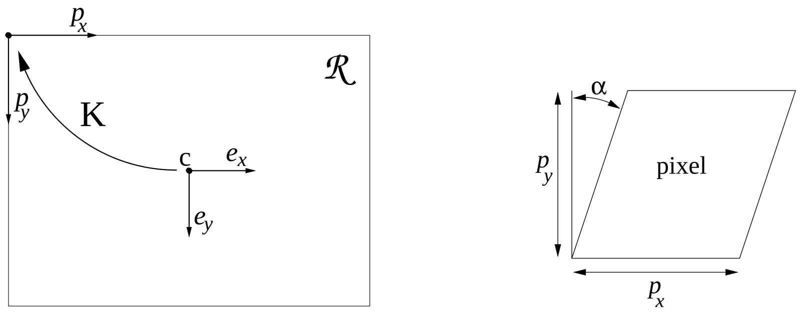

Figure 3.3: *From retinal coordinates to image coordinates*

where p_x and p_y are the width and the height of the pixels, $\mathtt{c} = [c_x\, c_y\, 1]^\top$ is the principal point and α the skew angle as indicated in Figure 3.3. Since only the ratios $\frac{f}{p_x}$ and $\frac{f}{p_y}$ are of importance the simplified notations of the following equation will be used in the remainder of this text:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x_{\mathcal{R}} \\ y_{\mathcal{R}} \\ 1 \end{bmatrix} \tag{3.3}$$

with f_x and f_y being the focal length measured in width and height of the pixels, and s a factor accounting for the skew due to non-rectangular pixels. The above upper triangular matrix is called the *calibration matrix* of the camera; and the notation **K** will be used for it. So, the following equation describes the transformation from retinal coordinates to image coordinates.

$$\mathtt{m} = \mathbf{K}\mathtt{m}_{\mathcal{R}} \quad . \tag{3.4}$$

For most cameras the pixels are almost perfectly rectangular and thus s is very close to zero. Furthermore, the principal point is often close to the center of the image. These assumptions can often be used, certainly to get a suitable initialization for more complex iterative estimation procedures.

For a camera with fixed optics these parameters are identical for all the images taken with the camera. For cameras which have zooming and focusing capabilities the focal length can obviously change, but also the principal point can vary. An extensive discussion of this subject can for example be found in the work of Willson [181, 179, 180, 182].

3.2.3 Camera motion

Motion of scene points can be modeled as follows

$$\mathtt{M}' = \begin{bmatrix} \mathbf{R} & \mathtt{t} \\ 0_3^\top & 1 \end{bmatrix} \mathtt{M} \tag{3.5}$$

with $\mathbf{R}$ a rotation matrix and $\mathtt{t} = [t_x t_y t_z]^\top$ a translation vector.

The motion of the camera is equivalent to an inverse motion of the scene and can therefore be modeled as

$$\mathtt{M}' = \begin{bmatrix} \mathbf{R}^\top & -\mathbf{R}^\top \mathtt{t} \\ 0_3^\top & 1 \end{bmatrix} \mathtt{M}, \tag{3.6}$$

with $\mathbf{R}$ and $\mathtt{t}$ indicating the motion of the camera.

3.2.4 The projection matrix

Combining equations (3.2), (3.3) and (3.6) the following expression is obtained for a camera with some specific intrinsic calibration and with a specific position and orientation:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}^\top & \text{-}\mathbf{R}^\top \mathtt{t} \\ 0_3^\top & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

which can be simplified to

$$\mathtt{m} \sim \mathbf{K}[\mathbf{R}^\top \text{-}\mathbf{R}^\top \mathtt{t}]\mathtt{M} \tag{3.7}$$

or even

$$\mathtt{m} \sim \mathbf{P}\mathtt{M}\ . \tag{3.8}$$

The 3×4 matrix $\mathbf{P}$ is called the *camera projection matrix.*

Using (3.8) the plane corresponding to a back-projected image line $\mathtt{l}$ can also be obtained: Since $\mathtt{l}^\top \mathtt{m} \sim \mathtt{l}^\top \mathbf{P}\mathtt{M} \sim \Pi^\top \mathtt{M}$,

$$\Pi \sim \mathbf{P}^\top \mathtt{l} \tag{3.9}$$

The transformation equation for projection matrices can be obtained as described in paragraph 2.2.3. If the points of a calibration grid are transformed by the same transformation as the camera, their image points should stay the same:

$$\mathtt{m} \sim \mathbf{P}'\mathtt{M}' \sim \mathbf{P}\mathbf{T}^{-1}\mathbf{T}\mathtt{M} \sim \mathbf{P}\mathtt{M} \tag{3.10}$$

and thus

$$\mathbf{P} \mapsto \mathbf{P}' \sim \mathbf{P}\mathbf{T}^{-1} \tag{3.11}$$

The projection of the outline of a quadric can also be obtained. For a line in an image to be tangent to the projection of the outline of a quadric, the corresponding plane should be on the dual quadric. Substituting equation (3.9) in (2.17) the following constraint $\mathtt{l}^\top \mathbf{P}\mathbf{Q}^* \mathbf{P}^\top \mathtt{l} = 0$ is obtained for $\mathtt{l}$ to be tangent to the outline. Comparing this result with the definition of a conic (2.10), the following projection equation is obtained for quadrics (this results can also be found in [71]). :

$$\mathbf{C}^* \sim \mathbf{P}\mathbf{Q}^* \mathbf{P}^\top\ . \tag{3.12}$$

Relation between projection matrices and image homographies

The homographies that will be discussed here are collineations from $\mathcal{P}^2 \rightarrow \mathcal{P}^2$. A homography $\mathbf{H}$ describes the transformation from one plane to another. A number of special cases are of interest, since the image is also a plane. The projection of points of a plane into an image i can be described through a homography $\mathbf{H}_{\Pi i}$. The matrix representation of this homography is dependent on the choice of the projective basis in the plane.

As an image is obtained by perspective projection, the relation between points $\mathtt{M}_\Pi$ belonging to a plane Π in 3D space and their projections $\mathtt{m}_{\Pi i}$ in the image is mathematically expressed by a homography $\mathbf{H}_{\Pi i}$. The matrix of this homography is found as follows. If the plane Π is given by $\Pi \sim [\pi^\top\ 1]^\top$ and the point $\mathtt{M}_\Pi$ of Π is represented as $\mathtt{M}_\Pi \sim [\mathtt{m}_\Pi^\top\ 1]^\top$, then $\mathtt{M}_\Pi$ belongs to Π if and only if $0 = \Pi^\top \mathtt{M}_\Pi = \pi^\top \mathtt{m}_\Pi + 1$. Hence,

$$\mathtt{M}_\Pi \sim \begin{bmatrix} \mathtt{m}_\Pi \\ 1 \end{bmatrix} = \begin{bmatrix} \mathtt{m}_\Pi \\ -\pi^\top \mathtt{m}_\Pi \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3\times 3} \\ -\pi^\top \end{bmatrix} \mathtt{m}_\Pi \ . \tag{3.13}$$

Now, if the camera projection matrix is $\mathbf{P}_i = [\mathbf{A}_i|\mathtt{a}_i]$, then the projection $\mathtt{m}_{\Pi i}$ of $\mathtt{M}_\Pi$ onto the image is

$$\begin{aligned} \mathtt{m}_{\Pi i} \sim \mathbf{P}_i \mathtt{M}_\Pi &= [\mathbf{A}_i|\mathtt{a}_i] \begin{bmatrix} \mathbf{I}_{3\times 3} \\ -\pi^\top \end{bmatrix} \mathtt{m}_\Pi \\ &= [\mathbf{A}_i - \mathtt{a}_i \pi^\top] \mathtt{m}_\Pi \ . \end{aligned} \tag{3.14}$$

Consequently, $\mathbf{H}_{\Pi i} \sim \mathbf{A}_i - \mathtt{a}_i \pi^\top$.

Note that for the specific plane $\Pi_{\mathtt{REF}} = [0\,0\,0\,1]^\top$ the homographies are simply given by $\mathbf{H}_{\mathtt{REF}i} \sim \mathbf{A}_i$.

It is also possible to define homographies which describe the transfer from one image to the other for points and other geometric entities located on a specific plane. The notation $\mathbf{H}^{\Pi}_{ij}$ will be used to describe such a homography from view i to j for a plane Π. These homographies can be obtained through the following relation $\mathbf{H}^{\Pi}_{ij} = \mathbf{H}_{\Pi j} \mathbf{H}_{\Pi i}^{-1}$ and are independent to reparameterizations of the plane (and thus also to a change of basis in $\mathcal{P}^3$).

In the metric and Euclidean case, $\mathbf{A}_i = \mathbf{K}_i \mathbf{R}_i^\top$ and the plane at infinity is $\Pi_\infty = [0001]^\top$. In this case, the homographies for the plane at infinity can thus be written as:

$$\mathbf{H}^{\infty}_{ij} = \mathbf{K}_i \mathbf{R}_{ij}^\top \mathbf{K}_i^{-1} \, , \tag{3.15}$$

where $\mathbf{R}_{ij} = \mathbf{R}_i^\top \mathbf{R}_j$ is the rotation matrix that describes the relative orientation from the j^{th} camera with respect top the i^{th} one.

In the projective and affine case, one can assume that $\mathbf{P}_1 = [\mathbf{I}_{3\times 3}|0_3]$ (since in this case $\mathbf{K}_i$ is unknown). In that case, the homographies $\mathbf{H}_{\Pi 1} \sim \mathbf{I}_{3\times 3}$ for all planes; and thus, $\mathbf{H}^{\mathtt{REF}}_{1i} = \mathbf{H}_{\mathtt{REF}i}$. Therefore $\mathbf{P}_i$ can be factorized as

$$\mathbf{P}_i = [\mathbf{H}^{\mathtt{REF}}_{1i}|\mathtt{e}_{1i}] \tag{3.16}$$

where $\mathtt{e}_{1i}$ is the projection of the center of projection of the first camera (in this case, $[0\,0\,0\,1]^\top$) in image i. This point $\mathtt{e}_{1i}$ is called the *epipole*, for reasons which will become clear in Section 3.3.1.

Note that this equation can be used to obtain $\mathbf{H}_{1i}^{\mathtt{REF}}$ and $\mathtt{e}_{1i}$ from $\mathbf{P}_i$, but that due to the unknown relative scale factors $\mathbf{P}_i$ can, in general, not be obtained from $\mathbf{H}_{1i}^{\mathtt{REF}}$ and $\mathtt{e}_{1i}$. Observe also that, in the affine case (where $\Pi_\infty = [0001]^\top$), this yields $\mathbf{P}_i = [\mathbf{H}_{1i}^{\infty}|\mathtt{e}_{1i}]$.

Combining equations (3.14) and (3.16), one obtains

$$\mathbf{H}_{1i}^{\Pi} = \mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi^\top \tag{3.17}$$

This equation gives an important relationship between the homographies for all possible planes. Homographies can only differ by a term $\mathtt{e}_{1i}[\mathtt{1} - \pi']^\top$. This means that in the projective case the homographies for the plane at infinity are known up to 3 common parameters (i.e. the coefficients of π_∞ in the projective space). Equations (3.17) and (3.15) will play an important role in Chapter 5.

Equation (3.16) also leads to an interesting interpretation of the camera projection matrix:

$$\begin{aligned}
\mathtt{m}_1 &\sim [\mathbf{I}_{3\times 3}|\mathtt{0}_3]\begin{bmatrix}\mathtt{m}\\ 1\end{bmatrix} = \mathtt{m} && (3.18)\\
\mathtt{m}_i &\sim [\mathbf{H}_{1i}^{\mathtt{REF}}|\mathtt{e}_{1i}]\begin{bmatrix}\mathtt{m}\\ 1\end{bmatrix} = \mathbf{H}_{1i}^{\mathtt{REF}}\mathtt{m} + \mathtt{e}_{1i} && (3.19)\\
&= \lambda\mathbf{H}_{1i}^{\mathtt{REF}}\mathtt{m}_1 + \mathtt{e}_{1i} = \mathbf{P}_i(\lambda\begin{bmatrix}\mathtt{m}_1\\ 0\end{bmatrix} + \begin{bmatrix}\mathtt{0}_3\\ 1\end{bmatrix}) && (3.20)
\end{aligned}$$

In other words, a point can thus be parameterized as being on the line through the optical center of the first camera (i.e. $[0001]^\top$) and a point in the reference plane $\Pi_{\mathtt{REF}}$. This interpretation is illustrated in Figure 3.4.

3.2.5 Deviations from the camera model

The perspective camera model describes relatively well the image formation process for most cameras. However, when high accuracy is required or when low-end cameras are used, additional effects have to be taken into account.

The failures of the optical system to bring all light rays received from a point object to a single image point or to a prescribed geometric position should then be taken into account. These deviations are called aberrations. Many types of aberrations exist (e.g. astigmatism, chromatic aberrations, spherical aberrations, coma aberrations, curvature of field aberration and distortion aberration). It is outside the scope of this work to discuss them all. The interested reader is referred to the work of Willson [181] and to the photogrammetry literature [147].

Many of these effects are negligible under normal acquisition circumstances. Radial distortion, however, can have a noticeable effect for shorter focal lengths. Radial distortion is a linear displacement of image points radially to or from the center of the

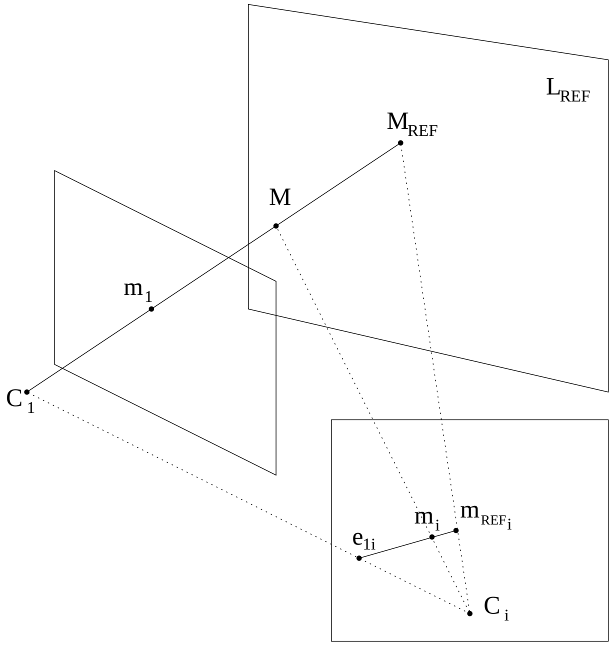

Figure 3.4: *A point* $\mathtt{M}$ *can be parameterized as* $\mathtt{C}_1 + \lambda \mathtt{M}_{\mathtt{REF}}$. *Its projection in another image can then be obtained by transferring* $\mathtt{m}_1$ *according to* $\Pi_{\mathtt{REF}}$ *(i.e. with* $\mathbf{H}_{1i}^{\mathtt{REF}}$*) to image* i *and applying the same linear combination with the projection* $\mathtt{e}_{1i}$ *of* $\mathtt{C}_1$ *(i.e.* $\mathtt{m}_i \sim \mathtt{e}_{1i} + \lambda \mathbf{H}_{1i}^{\mathtt{REF}} \mathtt{m}_1$*).*

image, caused by the fact that objects at different angular distance from the lens axis undergo different magnifications.

It is possible to cancel most of this effect by warping the image. The coordinates in undistorted image plane coordinates (x, y) can be obtained from the observed image coordinates (x_o, y_o) by the following equation:

$$\begin{array}{l} x = x_o + (x_o - c_x)(K_1 r^2 + K_2 r^4 + \ldots) \\ y = y_o + (y_o - c_y)(K_1 r^2 + K_2 r^4 + \ldots) \end{array} \tag{3.21}$$

where K_1 and K_2 are the first and second parameters of the radial distortion and

$$r = (x_o - c_x)^2 + (y_o - c_y)^2 \ .$$

Note that it can sometimes be necessary to allow the center of radial distortion to be different from the principal point [182].

When the focal length of the camera changes (through zoom or focus) the parameters K_1 and K_2 will also vary. In a first approximation this can be modeled as follows:

$$\begin{array}{l} x = x_o + (x_o - c_x)(K_{f1}\frac{r^2}{f^2} + K_{f2}\frac{r^4}{f^4} + \ldots) \\ y = y_o + (y_o - c_y)(K_{f1}\frac{r^2}{f^2} + K_{f2}\frac{r^4}{f^4} + \ldots) \end{array} \tag{3.22}$$

Due to the changes in the lens system this is only an approximation, except for digital zooms where (3.22) is exact.

3.3 Multi view geometry

Different views of a scene are not unrelated. Several relationships exist between two, three or more images. These are very important for the calibration and reconstruction from images. Many insights in these relationships have been obtained in recent years.

3.3.1 Two view geometry

In this section the following question will be addressed: *Given an image point in one image, does this restrict the position of the corresponding image point in another image?* It turns out that it does and that this relationship can be obtained from the calibration or even from a set of prior point correspondences.

Although the exact position of the scene point $\mathtt{M}$ is not known, it is bound to be on the line of sight of the corresponding image point $\mathtt{m}$. This line can be projected in another image and the corresponding point $\mathtt{m}'$ is bound to be on this projected line $\mathtt{l}'$. This is illustrated in Figure 3.5. In fact all the points on the plane Π defined by the two projection centers and $\mathtt{M}$ have their image on $\mathtt{l}'$. Similarly, all these points are projected on a line $\mathtt{l}$ in the first image. $\mathtt{l}$ and $\mathtt{l}'$ are said to be in *epipolar correspondence* (i.e. the corresponding point of every point on $\mathtt{l}$ is located on $\mathtt{l}'$, and vice versa).

Every plane passing through both centers of projection $\mathtt{C}$ and $\mathtt{C}'$ results in such a set of corresponding epipolar lines, as can be seen in Figure 3.6. All these lines pass

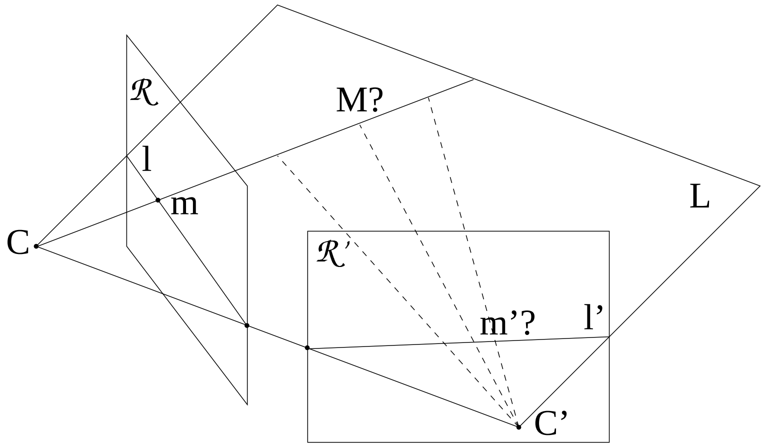

Figure 3.5: *Correspondence between two views. Even when the exact position of the 3D point* $\mathtt{M}$ *corresponding to the image point* $\mathtt{m}$ *is not known, it has to be on the line through* $\mathtt{C}$ *which intersects the image plane in* $\mathtt{m}$. *Since this line projects to the line* $\mathtt{l}'$ *in the other image, the corresponding point* $\mathtt{m}'$ *should be located on this line. More generally, all the points located on the plane defined by* $\mathtt{C}$, $\mathtt{C}'$ *and* $\mathtt{M}$ *have their projection on* $\mathtt{l}$ *and* $\mathtt{l}'$.

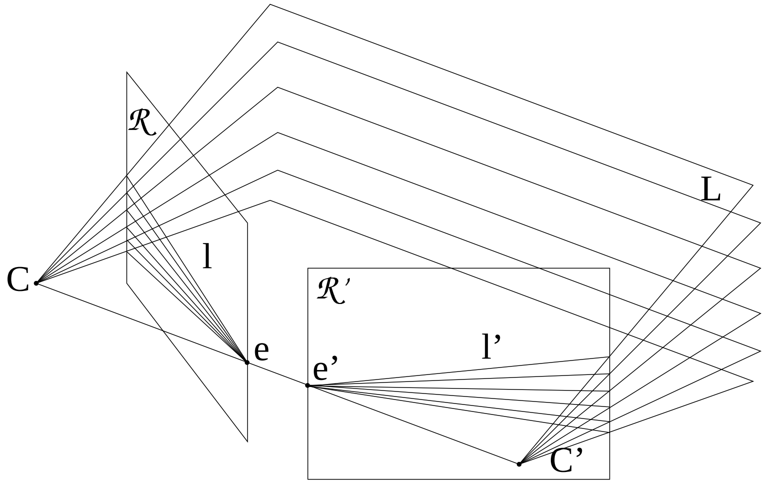

Figure 3.6: *Epipolar geometry. The line connecting* $\mathtt{C}$ *and* $\mathtt{C}'$ *defines a bundle of planes. For every one of these planes a corresponding line can be found in each image, e.g. for* Π *these are* $\mathtt{l}$ *and* $\mathtt{l}'$. *All 3D points located in* Π *project on* $\mathtt{l}$ *and* $\mathtt{l}'$ *and thus all points on* $\mathtt{l}$ *have their corresponding point on* $\mathtt{l}'$ *and vice versa. These lines are said to be in epipolar correspondence. All these epipolar lines must pass through* $\mathtt{e}$ *or* $\mathtt{e}'$, *which are the intersection points of the line* $\mathtt{CC}'$ *with the retinal planes* $\mathcal{R}$ *and* $\mathcal{R}'$ *respectively. These points are called the epipoles.*

through two specific points $\mathtt{e}$ and $\mathtt{e}'$. These points are called the *epipoles*, and they are the projection of the center of projection in the opposite image.

This epipolar geometry can also be expressed mathematically. The fact that a point $\mathtt{m}$ is on a line $\mathtt{l}$ can be expressed as $\mathtt{l}^\top \mathtt{m} = 0$. The line passing trough $\mathtt{m}$ and the epipole $\mathtt{e}$ is

$$\mathtt{l} \sim [\mathtt{e}]_\times \mathtt{m}\,, \tag{3.23}$$

with $[\mathtt{e}]_\times$ the antisymmetric 3×3 matrix representing the vectorial product with $\mathtt{e}$.

From (3.9) the plane Π corresponding to $\mathtt{l}$ is easily obtained as $\Pi \sim \mathbf{P}^\top \mathtt{l}$ and similarly $\Pi \sim \mathbf{P}'^\top \mathtt{l}'$. Combining these equations gives:

$$\mathtt{l}' \sim \left(\mathbf{P}'^\top\right)^\dagger \mathbf{P}^\top \mathtt{l} \equiv \mathbf{H}^{-\top} \mathtt{l} \tag{3.24}$$

with † indicating the Moore-Penrose pseudo-inverse. The notation $\mathbf{H}^{-\top}$ is inspired by equation (2.7). Substituting (3.23) in (3.24) results in

$$\mathtt{l}' \sim \mathbf{H}^{-\top} [\mathtt{e}]_\times \mathtt{m}\ .$$

Defining $\mathbf{F} = \mathbf{H}^{-\top} [\mathtt{e}]_\times$, we obtain

$$\mathtt{l}' \sim \mathbf{F}\mathtt{m}\,, \tag{3.25}$$

and thus,

$$\mathtt{m}'^\top \mathbf{F} \mathtt{m} = 0\ . \tag{3.26}$$

This matrix $\mathbf{F}$ is called the *fundamental matrix*. These concepts were introduced by Faugeras [36] and Hartley [51]. Since then many people have studied the properties of this matrix (e.g. [85, 86]) and a lot of effort has been put in robustly obtaining this matrix from a pair of uncalibrated images [162, 163, 186].

Having the calibration, $\mathbf{F}$ can be computed and a constraint is obtained for corresponding points. When the calibration is not known equation (3.26) can be used to compute the fundamental matrix $\mathbf{F}$. Every pair of corresponding points gives one constraint on $\mathbf{F}$. Since $\mathbf{F}$ is a 3×3 matrix which is only determined up to scale, it has $3 \times 3 - 1$ unknowns. Therefore 8 pairs of corresponding points are sufficient to compute $\mathbf{F}$ with a linear algorithm.

Note from (3.25) that $\mathbf{F}\mathtt{e} = 0$, because $[\mathtt{e}]_\times \mathtt{e} = 0$. Thus, rank $\mathbf{F} = 2$. This is an additional constraint on $\mathbf{F}$ and therefore 7 point correspondences are sufficient to compute $\mathbf{F}$ through a nonlinear algorithm. In Section 7.3.1 the robust computation of the fundamental matrix from images will be discussed in more detail.

Relation between the fundamental matrix and image homographies

There also exists an important relationship between the homographies $\mathbf{H}^\Pi_{ij}$ and the fundamental matrices $\mathbf{F}_{ij}$. Let $\mathtt{m}_i$ be a point in image i. Then $\mathtt{m}_j \sim \mathbf{H}^\Pi_{ij} \mathtt{m}_i$ is the corresponding point for the plane Π in image j. Therefore, $\mathtt{m}_j$ is located on the corresponding epipolar line; and,

$$(\mathbf{H}^\Pi_{ij} \mathtt{m}_i)^\top \mathbf{F}_{ij} \mathtt{m}_i = 0 \tag{3.27}$$

should be verified. Moreover, equation (3.27) holds for every image point $\mathtt{m}_i$. Since the fundamental matrix maps points to corresponding epipolar lines, $\mathbf{F}_{ij}\mathtt{m}_i \sim \mathtt{e}_{ij} \times \mathtt{m}_j$ and equation (3.27) is equivalent to $\mathtt{m}_j^\top [\mathtt{e}_{ij}]_\times \mathbf{H}_{ij}^{\Pi} \mathtt{m}_i = 0$. Comparing this equation with $\mathtt{m}_j^\top \mathbf{F}_{ij} \mathtt{m}_i = 0$, and using that these equations must hold for all image points $\mathtt{m}_i$ and $\mathtt{m}_j$ lying on corresponding epipolar lines, it follows that:

$$\mathbf{F}_{ij} \sim [\mathtt{e}_{ij}]_\times \mathbf{H}_{ij}^{\Pi} \ . \tag{3.28}$$

Let $\mathtt{l}_j$ be a line in image j and let Π be the plane obtained by back-projecting $\mathtt{l}_j$ into space. If $\mathtt{m}_{\Pi i}$ is the image of a point of this plane projected in image i, then the corresponding point in image j must be located on the corresponding epipolar line (i.e. $\mathbf{F}_{ij}\mathtt{m}_{\Pi i}$). Since this point is also located on the line $\mathtt{l}_j$ it can be uniquely determined as the intersection of both (if these lines are not coinciding): $\mathtt{l}_j \times \mathbf{F}_{ij}\mathtt{m}_{\Pi i}$. Therefore, the homography $\mathbf{H}_{ij}^{\Pi}$ is given by $[\mathtt{l}_j]_\times \mathbf{F}_{ij}$. Note that, since the image of the plane Π is a line in image j, this homography is not of full rank. An obvious choice to avoid coincidence of $\mathtt{l}_j$ with the epipolar lines, is $\mathtt{l}_j \sim \mathtt{e}_{ij}$ since this line does certainly not contain the epipole (i.e. $\mathtt{e}_{ij}^\top \mathtt{e}_{ij} \neq 0$). Consequently,

$$[\mathtt{e}_{ij}]_\times \mathbf{F}_{ij} \tag{3.29}$$

corresponds to the homography of a plane. By combining this result with equations (3.16) and (3.17) one can conclude that it is always possible to write the projection matrices for two views as

$$\begin{array}{rcl} \mathbf{P}_1 & = & [\mathbf{I}_{3\times 3} \,|\, 0_3] \\ \mathbf{P}_2 & = & [[\mathtt{e}_{12}]_\times \mathbf{F}_{12} - \mathtt{e}_{12}\pi^\top \,|\, \mathtt{e}_{12}] \end{array} \tag{3.30}$$

Note that this is an important result, since it means that a projective camera setup can be obtained from the fundamental matrix which can be computed from 7 or more matches between two views. Note also that this equation has 4 degrees of freedom (i.e. the 3 coefficients of π and the arbitrary relative scale between $\mathbf{F}_{12}$ and $\mathtt{e}_{12}$). Therefore, this equation can only be used to instantiate a new frame (i.e. an arbitrary projective representation of the scene) and not to obtain the projection matrices for all the views of a sequence (i.e. compute $\mathbf{P}_3, \mathbf{P}_4, \ldots$). How this can be done is explained in Section 7.3.3.

3.3.2 Three view geometry

Considering three views it is, of course, possible to group them in pairs and to get the two view relationships introduced in the last section. Using these pairwise epipolar relations, the projection of a point in the third image can be predicted from the coordinates in the first two images. This is illustrated in Figure 3.7. The point in the third image is determined as the intersection of the two epipolar lines. This computation, however, is not always very well conditioned. When the point is located in the trifocal plane (i.e. the plane going through the three centers of projection), it is completely undetermined.

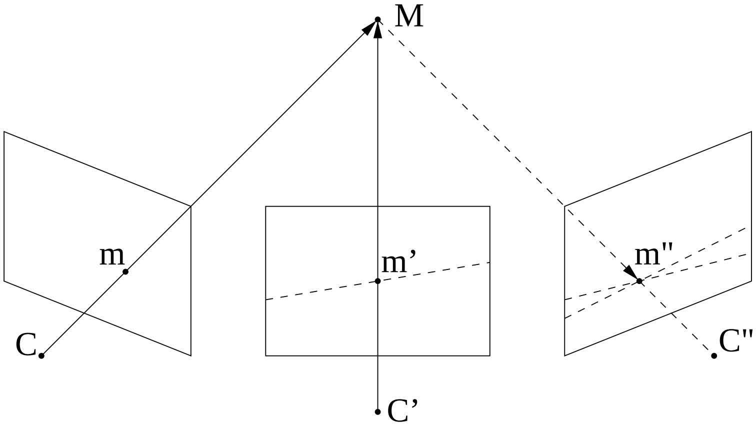

Figure 3.7: *Relation between the image of a point in three views. The epipolar lines of points* $\mathtt{m}$ *and* $\mathtt{m}'$ *could be used to obtain* $\mathtt{m}''$*. This does, however, not exhaust all the relations between the three images. For a point located in the trifocal plane (i.e. the plane defined by* $\mathtt{C}, \mathtt{C}'$ *and* $\mathtt{C}''$*) this would not give a unique solution, although the 3D point could still be obtained from its image in the first two views and then be projected to* $\mathtt{m}''$*. Therefore, one can conclude that in the three view case not all the information is described by the epipolar geometry. These additional relationships are described by the trifocal tensor.*

Fortunately, there are additional constraints between the images of a point in three views. When the centers of projection are not coinciding, a point can always be reconstructed from two views. This point then projects to a unique point in the third image, as can be seen in Figure 3.7, even when this point is located in the trifocal plane. For two views, no constraint is available to restrict the position of corresponding lines. Indeed, back-projecting a line forms a plane, the intersection of two planes always results in a line. Therefore, no constraint can be obtained from this. But, having three views, the image of the line in the third view can be predicted from its location in the first two images, as can be seen in Figure 3.8. Similar to what was derived for two views, there are multi linear relationships relating the positions of points and/or lines in three images [148]. The coefficients of these multi linear relationships can be organized in a tensor which describes the relationships between points [144] and lines [54] or any combination thereof [56]. Several researchers have worked on methods to compute the trifocal tensor (e.g. see [160, 161]).

The trifocal tensor $\mathbf{T}$ is a $3 \times 3 \times 3$ tensor. It contains 27 parameters, only 18 of which are independent due to additional nonlinear constraints. The trilinear relationship for a point is given by the following equation[1]:

$$m_i(m'_j m''_k T_{i33} - m''_k T_{ij3} - m'_j T_{i3k} + T_{ijk}) = 0 \tag{3.31}$$

Any triplet of corresponding points should satisfy this constraint.

---

[1] The Einstein convention is used (i.e. indices that are repeated should be summed over).

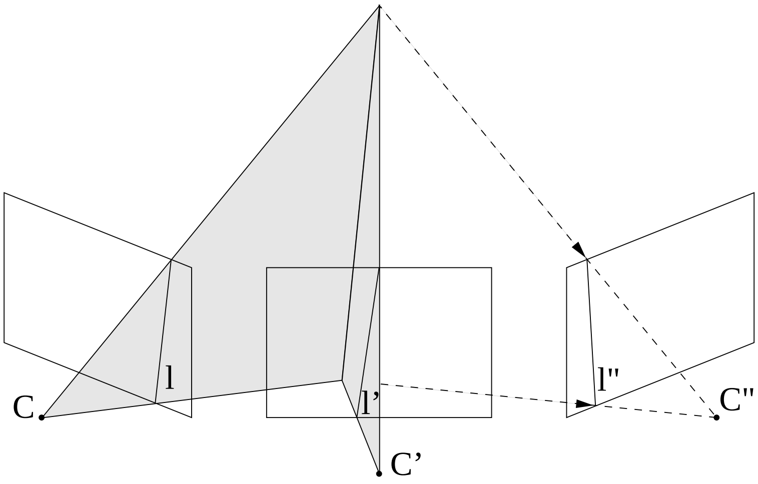

Figure 3.8: *Relation between the image of a line in three images. While in the two view case no constraints are available for lines, in the three view case it is also possible to predict the position of a line in a third image from its projection in the other two. This transfer is also described by the trifocal tensor.*

A similar constraint applies for lines. Any triplet of corresponding lines should satisfy:

$$l_i \sim l'_j l''_k T_{ijk}$$

3.3.3 Multi view geometry

Many people have been studying multi view relationships [62, 165, 39]. Without going into detail we would like to give some intuitive insights to the reader. For a more in depth discussion the reader is referred to [98].

An image point has 2 degrees of freedom. But n images of a 3D point do not have $2n$ degrees of freedom, but only 3. So, there must be $2n-3$ independent constraints between them. For lines, which also have 2 degrees of freedom in the image, but 4 in 3D space, n images of a line must satisfy $2n-4$ constraints.

Some more properties of these constraints are explained here. A line can be backprojected into space linearly (3.9). A point can be seen as the intersection of two lines. To correspond to a real point or line the planes resulting from the backprojection must all intersect in a single point or line. This is easily expressed in terms of determinants, i.e. $|\Pi_1\Pi_2\Pi_3\Pi_4| = 0$ for points and that all the 3×3 subdeterminants of $[\Pi_1\Pi_2\Pi_3]$ should be zero for lines. This explains why the constraints are multi linear, since this is a property of columns of a determinant. In addition no constraints combining more than 4 images exist, since with 4-vectors (i.e. the representation of the planes) maximum 4×4 determinants can be obtained. The twofocal (i.e. the fundamental matrix) and the trifocal tensors have been discussed in the previous paragraphs, recently Hartley [59] proposed an algorithm for the practical computation of the quadrifocal tensor.

3.4 Conclusion

In this chapter some important concepts were introduced. A geometric description of the image formation process was given and the camera projection matrix was introduced. Some important relationships between multiple views were also derived. The insights obtained by carefully studying these properties have shown that it is possible to retrieve a relative calibration of a two view camera setup from point matches only. This is an important result which will be exploited further on to obtain a 3D reconstruction starting from the images.

Chapter 4

Self-calibration

4.1 Introduction

One of the main contributions of this work is on the subject of self-calibration. Before discussing the specific algorithms that were developed, the general concepts and several methods developed by others will be discussed.

This chapter is organized as follows. First it is showed that without additional constraints a reconstruction obtained from images is only determined up to an arbitrary projective transformation. Before discussing the possibility of self-calibration the traditional approaches for calibration are briefly reviewed. The main part of this chapter is then dedicated to the subject of self-calibration. The most important existing approaches for general motion are reviewed. The work of Faugeras/Maybank/Luong/Zeller [37, 85, 87, 183], Hartley [53], Heyden and Åström [60], Triggs [166] and Pollefeys and Van Gool [124] is presented. Some specific methods for restricted motions are also discussed. The cases of pure translation [173, 100], pure rotation [55] and planar motion [3, 33] are presented. Finally, the subject of motion sequence which do not allow self-calibration is discussed. The work of Sturm [150] on critical motion sequences is presented and some new results are added. The last section summarizes this chapter.

4.2 Projective ambiguity

Suppose a set of images of a static scene is given. If the calibration, position and orientation of the cameras are known, it is possible to reconstruct the observed scene points. Two (or more) corresponding image points (i.e. points corresponding to the same scene point) can be used to obtain the reconstruction of a scene point. This reconstructed point is computed as the intersection of the rays of sight corresponding to the image points. This reconstruction is uniquely determined in space.

In the uncalibrated case the camera calibration, orientation and position are unknown. In the most general case the following assumptions are made:

- The intrinsic camera parameters are unknown. Different camera(s) (settings) could be used for every view. Therefore, the parameters of the calibration matrix $\mathbf{K}$ are unknown and can be different for every view. This means, in general, 5 unknown parameters per view.
- The extrinsic camera parameters are also unknown. The position and the orientation of the camera are unknown and can be different for every view. This means 6 other unknown parameters for every view.
- The only assumption about the scene is that it is static. Every point has 3 degrees of freedom.

Due to the first two assumptions $\mathbf{P}$ is unconstrained. The first 3×3 part is determined by the product of $\mathbf{K}$ and $\mathbf{R}$. It can be proven (using the properties of the QR-decomposition) that any matrix can be obtained up to scale. It is clear that the last column, being given by $-\mathbf{R}^\top \mathtt{t}$, is also unconstrained. From the three assumptions it follows that the points $\mathtt{M}$ are unconstrained (except that the same $\mathtt{M}$ has to be used for all the views). In conclusion both $\mathbf{P}$ and $\mathtt{M}$ are unconstrained.

Let us assume that a reconstruction was obtained, which corresponds to the original scene up to a projective transformation. This reconstruction consists of both camera projection matrices and scene points. How this can be achieved will be explained in Chapter 7. Assume for now that a valid reconstruction could be obtained for some image points. This means that for the image points $\mathtt{m}_{il}$ (with i referring to the image in which the point appears and l indicating from which scene point it was projected) a reconstruction $\{\mathbf{P}_i, \mathtt{M}_l\}$ was obtained. This reconstruction must satisfy the following equation:

$$\mathtt{m}_{il} \sim \mathbf{P}_i \mathtt{M}_l, \qquad \forall i, l \ .$$

In this case, however, also the following equation will be satisfied:

$$\mathtt{m}_{il} \sim (\mathbf{P}_i \mathbf{T}^{-1})(\mathbf{T}\mathtt{M}_l) \sim \mathbf{P}'_i \mathtt{M}'_l, \qquad \forall i, l\,,$$

with $\mathbf{P}'_i = \mathbf{P}_i \mathbf{T}^{-1}$ and $\mathtt{M}'_l = \mathbf{T}\mathtt{M}_l$ where $\mathbf{T}$ is an arbitrary projective transformation. This means that $\{\mathbf{P}'_i, \mathtt{M}'_l\}$ is also a possible reconstruction. Without additional constraints the reconstruction therefore is only determined up to an arbitrary projective transformation. This is called a *projective reconstruction* of the scene. Only the projective stratum of the geometry is retrieved.

Although this can be sufficient for some applications [135], for many applications a projective reconstruction is not usable. For visualization, for example, at least a metric representation is needed. The information that is needed to update the projective reconstruction to a metric stratum are metric quantities of the scene, or, some calibration information about the camera. The latter can consist of constraints on the intrinsic or extrinsic parameters of the camera.

A projective reconstruction is determined up to an arbitrary projective transformation that has 15 degrees of freedom. For the metric case the ambiguity transformation only has 7 degrees of freedom. This means that 8 independent constraints should, in general, be sufficient to perform an upgrade of the calibration from projective to metric. How this can be achieved is described in the next sections.

4.3 Calibration

In this section some existing calibration approaches are briefly discussed. These can be based on Euclidean or metric knowledge about the scene or about the camera or its motion. One approach consists of first computing a projective reconstruction and then upgrading it a posteriori to a metric (or Euclidean) reconstruction by imposing some constraints. The traditional approaches however immediately go for a metric (or Euclidean) reconstruction.

4.3.1 Scene knowledge

The knowledge of (relative) distances or angles in the scene can be used to obtain information about the metric structure. One of the easiest means to calibrate the scene at a metric level is the knowledge of the relative position of 5 or more points in general position. Assume the points $\mathtt{M}'_l$ are the metric coordinates of the reconstructed points $\mathtt{M}_l$, then the transformation $\mathbf{T}$ which upgrades the reconstruction from projective to metric can be obtained from the following equations

$$\mathtt{M}'_l \sim \mathbf{T}\mathtt{M}_l \text{ or } \lambda_l \mathtt{M}'_l = \mathbf{T}\mathtt{M}_l \tag{4.1}$$

which can be rewritten as linear equations by eliminating λ_l. Boufama et al. [12] investigated how some Euclidean constraints could be imposed on an uncalibrated reconstruction. The constraints they dealt with are known 3D points, points on a ground plane, vertical alignment and known distances between points. Bondyfalat and Bougnoux [11] recently proposed a method in which the constraints are first processed by a geometric reasoning system so that a minimal representation of the scene is obtained. These constraints can be incidence, parallelism and orthogonality. This minimal representation is then fed to a constrained bundle adjustment.

The traditional approach taken by photogrammetrists [16, 47, 147, 48] consists of immediately imposing the position of known control points during reconstruction. These methods use bundle adjustment [17] which is a global minimization of the reprojection error. This can be expressed through the following criterion:

$$\mathcal{C}_{bundle} = \sum_{i=1}^{n} \sum_{l \in I_i} \left((x_{li} - \mathbf{P}_i(\mathtt{M}_l))^2 + (y_{li} - \mathbf{P}_i(\mathtt{M}_l))^2 \right) \tag{4.2}$$

where I_i is the set of indices corresponding to the points seen in view i and $\mathbf{P}_i(\mathtt{M}_l)$ described the projection of a point $\mathtt{M}_l$ with camera $\mathbf{P}_i$ taking all distortions into account. Note that $\mathtt{M}_l$ is known for control points and unknown for other points. It is clear that this approach results in a huge minimization problem and that, even if the special structure of the Jacobian is taken into account, it is computationally very expensive.

Calibration object In the case of a calibration object, the parameters of the camera are estimated using an object with known geometry. The known calibration can then be used to immediately obtain metric reconstructions.

Many approaches exist for this type of calibration. Most of these methods consist of a two step procedure where a calibration is obtained first for a simplified (linear) model and then a more complex model, taking distortions into account, is fitted to the measurements. The difference between the methods mainly lies in the type of calibration object that is expected (e.g. planar or not) or the complexity of the camera model that is used. Some existing techniques are Faugeras and Toscani [32], Weng, Cohen and Herniou [177], Tsai [169, 170] (see also the implementation by Willson [181]) and Lenz and Tsai [81].

4.3.2 Camera knowledge

Knowledge about the camera can also be used to restrict the ambiguity on the reconstruction from projective to metric or even beyond. Different parameters of the camera can be known. Both knowledge about the extrinsic parameters (i.e. position and orientation) as the intrinsic parameters can be used for calibration.

Extrinsic parameters Knowing the relative position of the viewpoints, is equivalent to knowing the relative position of 3D points. Therefore, the relative position of 5 viewpoints in general position suffices to obtain a metric reconstruction. This is the principle behind the omnirig [143] recently proposed by Shashua. I used a similar approach in [126, 125] to cancel the effect of a change in focal length for stereo rigs (see Section 5.4.3).

It is less obvious to deal with the orientation parameters, except when the intrinsic parameters are also known (see below).

Intrinsic parameters If the intrinsic camera parameters are known, it is possible to obtain a metric reconstruction. This calibration can for example be obtained through off-line calibration with a calibration object. In the minimal case of 2 views and 5 points multiple solutions can exist [38], but in general a unique solution is easily found. Traditional structure from motion algorithms assume known intrinsic parameters and obtain metric reconstructions out of it (e.g. [84, 168, 5, 21, 148, 156]).

Intrinsic and extrinsic parameters When both intrinsic and extrinsic camera parameters are known, the full camera projection matrix is determined. In this case a Euclidean reconstruction is immediately obtained by back-projecting the points.

In the case of known relative position and orientation of the cameras, the first view can be aligned with the world frame without loss of generality. If only the (relative) orientation and the intrinsic parameters are known, the first 3×3 part of the camera projection matrices is known and it is still possible to linearly obtain the transformation which upgrades the projective reconstruction to metric.

4.4 Self-calibration

In many cases the specific values of the intrinsic or extrinsic camera parameters are not known. Often there are, however, some restrictions on these parameters. Using these restrictions to achieve a metric calibration is called *self-calibration* or auto-calibration. The traditional self-calibration problem is more restricted. In that case it is assumed that all intrinsic camera parameters are unknown but constant and that the motion of the camera is unrestricted. This corresponds to an unknown camera which is freely moved around (e.g. hand-held). This problem has been addressed by many researchers [37, 85, 87, 183, 53, 60, 166, 124, 122, 123] and will be discussed more in detail in Section 4.4.1.

In some practically important cases, however, the motion of the camera is restricted. This knowledge can often be exploited to design simpler algorithms. But these are not always able to retrieve all the desired parameters, since restricted motion sequences do not always contain enough information to uniquely determine the metric stratum of the reconstruction. Some interesting classes of motions, which will be discussed further on, are pure translations, pure rotations and planar motion.

At the end of this section the problem of critical motion sequences is introduced. This is mainly based on the work of Sturm [152, 153, 150]. Some new results are also presented.

4.4.1 General motions

Many methods exist for self-calibration, but they can easily be divided into just a few classes. A first class starts from a projective reconstruction and tries to find the absolute conic as the only conic which satisfies all the constraints imposed on its image. Typically, this means that these images have to be identical, since they are immediately related to the intrinsic camera parameters, which are assumed constant.

A second class of methods also uses the absolute conic, but restricts the constraints to the epipolar geometry. The advantage is that only the fundamental matrices are needed. On the other hand, this method suffers from several important disadvantages.

Besides these two classes some methods exist which factorize the projection matrices and impose the intrinsic camera parameters to be constant.

In the following paragraphs these different classes are discussed separately. For every class a few specific methods that are representative for that class are presented.

The image of the absolute conic

One of the most important concepts for self-calibration is the absolute conic and its projection in the images. Since it is invariant under Euclidean transformations, its relative position to a moving camera is constant. For constant intrinsic camera parameters its image will therefore also be constant. This is similar to someone who has the impression that the moon is following him when driving on straight road. Note that the absolute conic is more general, because it is not only invariant to translations but also to rotations and reflections.

It can be seen as a calibration object which is naturally present in all the scenes. It was seen in 2.3.3 that once this object is localized, it can be used to upgrade the reconstruction to metric. It is, however, not so simple to find this object back in the scene. The only difference with other proper virtual conics[1] is that its relative position towards the camera is always unchanged. So, if the relative position of another proper virtual conic also stays unchanged for a specific image sequence, there is no way to differentiate the real absolute conic from the other candidate. This problem has been studied in depth by Sturm [152, 150] and will be discussed in Section 4.4.3 and Section 6.4. The relationship between the absolute conic and its projection in an image is easily obtained using the dual quadric projection equation (3.12) on the dual absolute quadric:

$$\omega_i^* \sim \mathbf{P}_i \Omega^* \mathbf{P}_i^\top \ . \tag{4.3}$$

For a Euclidean representation of the world this results in (see equation (2.30)):

$$\omega_i^* \sim \mathbf{K}_i[\mathbf{R}_i^\top \,|\, \text{-}\mathbf{R}_i^\top \mathtt{t}_i] \begin{bmatrix} \mathbf{I}_{3\times 3} & 0_3 \\ 0_3^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_i \\ \text{-}\mathtt{t}_i^\top \mathbf{R}_i \end{bmatrix} \mathbf{K}_i^\top = \mathbf{K}_i \mathbf{K}_i^\top \tag{4.4}$$

This equation is very useful, because it immediately relates the intrinsic camera parameters to the (dual) image of the absolute conic.

In the case of a projective representation of the world the absolute quadric Ω^* will not be at its standard position, but will have the following form according to equation (2.20): $\Omega^* = \mathbf{T}\Omega_M^* \mathbf{T}^\top$ with $\mathbf{T}$ being the transformation from the metric to the projective representation. But, since the images were obtained in a Euclidean world, the image ω_i still satisfies (4.4). If Ω^* is retrieved, it is possible to upgrade the geometry from projective to metric through the procedure explained in Section 2.3.3.

The image of the absolute conic can also be transferred from one image to another through the homography of its supporting plane (i.e. the plane at infinity):

$$\omega_j \sim \mathbf{H}_{ij}^{\infty\,-\top} \omega_i \mathbf{H}_{ij}^{\infty\,-1} \text{ or } \omega_j^* \sim \mathbf{H}_{ij}^{\infty} \omega_i^* \mathbf{H}_{ij}^{\infty\,\top} \ . \tag{4.5}$$

From affine to metric When the intrinsic camera parameters are constant (i.e. $\omega_i = \omega_j$) and the homography of the plane at infinity $\mathbf{H}_{ij}^{\infty\,\top}$ is known, then equation (4.5) can be reduced to a set of linear equations in the coefficients of ω or ω^* [53]. By enforcing the equality of the determinants on both sides of the equation, an exact equality is obtained (i.e. not up to scale). This can be achieved by scaling $\mathbf{H}_{ij}^{\infty}$ properly:

$$\omega^* = \mathbf{H}_{ij}^{\infty} \omega^* \mathbf{H}_{ij}^{\infty\,\top} \text{ when } |\mathbf{H}_{ij}^{\infty}| = 1 \ . \tag{4.6}$$

This equation therefore allows to easily upgrade an affine reconstruction to metric. Figure 4.1 illustrates the concept of the absolute conic and its projection in the images.

---

[1] A proper virtual conic is a non-degenerate conic (i.e. rank **C**=3) which has no real points.

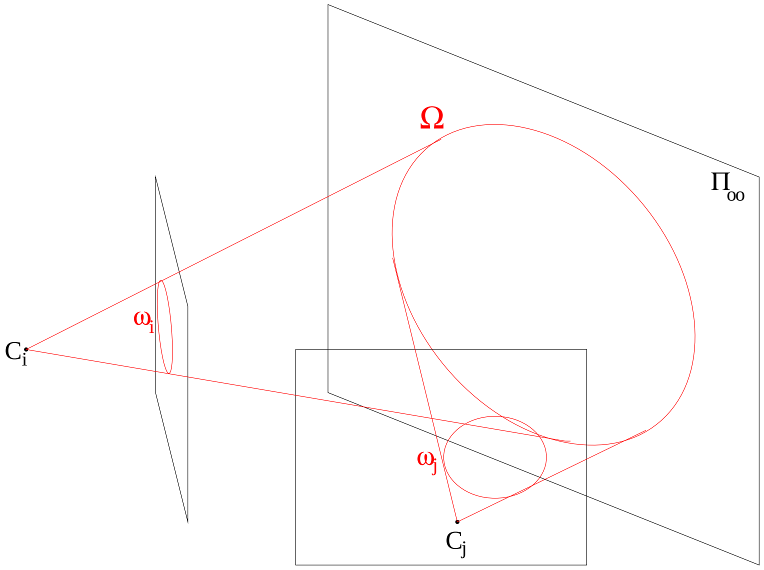

Figure 4.1: *The absolute conic and its projection in the images*

Triggs' method The use of the absolute quadric for self-calibration was proposed by Triggs in [166]. He proposes to use equation (4.3) to determine the absolute quadric by imposing that $\omega_i = \omega$ is constant. The most important problem with this equation is the presence of the unknown scale factors:

$$\omega^* = \lambda_i \mathbf{P}_i \Omega^* \mathbf{P}_i^\top \ . \tag{4.7}$$

These can be eliminated by taking ratios of components and cross-multiplication:

$$[\omega^*]_{kl} [\mathbf{P}_i \Omega^* \mathbf{P}_i^\top]_{k'l'} - [\omega^*]_{k'l'} [\mathbf{P}_i \Omega^* \mathbf{P}_i^\top]_{kl} = 0 \tag{4.8}$$

where $[.]_{kl}$ denotes the entry on row k and column l of the matrix between the square brackets. There are 15 equations of this type per view, 5 of which are independent [166]. The number of unknowns is 5 for ω^* and 8 for Ω^* if the rank 3 constraint is enforced.

Triggs proposed two methods to solve these equations. The first one uses a non-linear constraint minimization algorithm to minimize the residual to the equations (4.8) in an algebraic least-squares sense while enforcing rank 3 for Ω^*. In this case, 3 views, in general, are sufficient to obtain a metric reconstruction.

The second approach is quasi-linear. The coefficients of ω^* and Ω^* can be reorganized in vectors $\bar{\omega}^*$ and $\bar{\Omega}^*$ of respectively 6 and 10 coefficients. The equations (4.8) are linear in the coefficients of $[\omega^* \Omega^*] = \bar{\omega}^* \bar{\Omega}^{*\top}$ and have rank 15 with regard to these 59 unknowns (since everything is only determined up to scale). So, in this case, 4 images are needed to obtain the dual absolute quadric. $\bar{\omega}^*$ and $\bar{\Omega}^*$ can be obtained from $[\omega^* \Omega^*]$ by taking the left and right singular vectors associated with the biggest singular value (closest rank 1 approximation of $[\omega^*\Omega^*]$). When reconstructing Ω^*, the closest rank 3 approximation should be taken (putting the smallest singular value to zero).

Triggs reported that the nonlinear method should be preferred despite the need for (approximate) initialization, since it is faster, more accurate, and more robust.

Heyden and Åström's method Heyden and Åström were the first to propose a method based on the dual absolute quadric [60], but they did not give this geometric interpretation to their constraints. There are two important differences with [166]. The first one being that $\mathbf{P}_1$ is forced to $[\mathbf{I}_{3\times3} \,|\, 0_3]$. The second important difference is that the scale factors are seen as additional unknowns instead of eliminating them.

The authors are looking for a transformation which brings the projective camera projection matrices to a metric equivalent:

$$\mathbf{P}_i \mathbf{T}_{PM}^{-1} \sim \mathbf{K}[\mathbf{R}_i^\top \,|\, \text{-}\mathbf{R}_i^\top \mathtt{t}_i] \tag{4.9}$$

Assuming $\mathbf{P}_1 = [\mathbf{I}_{3\times3} \,|\, 0_3]$, $\mathbf{R}_1 = \mathbf{I}_{3\times3}$ and $\mathtt{t}_1 = 0_3$, there must exist such a transformation of the form

$$\mathbf{T}_{PM}^{-1} = \begin{bmatrix} \mathbf{K} & 0_3 \\ \mathtt{a}^\top & 1 \end{bmatrix} \; . \tag{4.10}$$

Since the last column of equation (4.9) has three independent unknowns (i.e. $\mathtt{t}_i$), it is not going to be of any help and can better be left out:

$$\mathbf{P}_i \begin{bmatrix} \mathbf{K} \\ \mathtt{a}^\top \end{bmatrix} \sim \mathbf{K}\mathbf{R}_i^\top \; . \tag{4.11}$$

The rotation component can also be eliminated by multiplying each side of equation (4.11) with its transpose.

$$\mathbf{P}_i \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathbf{K}^\top\mathtt{a} \\ \mathtt{a}^\top\mathbf{K} & \mathtt{a}^\top\mathtt{a} \end{bmatrix} \mathbf{P}_i^\top \sim \mathbf{K}\mathbf{R}\mathbf{R}^\top\mathbf{K}^\top \sim \mathbf{K}\mathbf{K}^\top \; . \tag{4.12}$$

This equation could have been obtained immediately from (4.3) and (4.4) by imposing $\mathbf{P}_1 = [\mathbf{I}_{3\times3} \,|\, 0_3]$. The advantage of fixing $\mathbf{P}_1$ is that the number of unknowns is restricted to 8 instead of 5+8 in [166].

The method proposed in [60] is now briefly described. Instead of eliminating the unknown scale factors, they are regarded as independent unknowns, yielding the following equations:

$$\mathbf{K}\mathbf{K}^\top - \lambda_i \mathbf{P}_i \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathtt{a}\mathbf{K}^\top \\ \mathbf{K}\mathtt{a}^\top & \mathtt{a}^\top\mathtt{a} \end{bmatrix} \mathbf{P}_i^\top = \mathbf{0}_{3\times3} \; . \tag{4.13}$$

Note that this equation is trivially satisfied for $\mathbf{P}_1$ with $\lambda_1 = 1$. The problem can be formulated as an optimization problem using the following goal function:

$$\mathcal{C}_{H\&A}(\mathbf{K}, \mathtt{a}, \lambda_i) = \sum_{i=2}^{n} \left\| \mathbf{K}\mathbf{K}^\top - \lambda_i \mathbf{P}_i \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathtt{a}\mathbf{K}^\top \\ \mathbf{K}\mathtt{a}^\top & \mathtt{a}^\top\mathtt{a} \end{bmatrix} \mathbf{P}_i^\top \right\|_F \; , \tag{4.14}$$

where $\|.\|_F$ denotes the Frobenius norm. An important disadvantage of this method is the introduction of an additional unknown λ_i for every view. This seems to cause convergence problems for longer image sequences.

This method does not treat all the images alike. Implicitly it is assumed that there is no error in the first image since $\mathbf{P}_1$ is assumed perfectly known.

Alternative method In [124] I proposed a related method for self-calibration. Instead of using the absolute quadric, the absolute conic and the plane at infinity are used explicitly. It is, however, shown that the obtained constraints are algebraically equivalent. The advantage being the possibility to deal more easily with the scale factors. This method is used in the next chapter as a refinement step in the stratified approach.

When the plane at infinity is known, equation (4.5) can be used to determine the dual image of the absolute conic ω^*:

$$\lambda_i \omega^* = \mathbf{H}^\infty_{1i} \omega^* \mathbf{H}^{\infty\top}_{1i} \quad . \tag{4.15}$$

Hartley proposed in [53] to eliminate the scale factors by scaling the determinants $|\mathbf{H}^\infty_{1i}| = 1$. Observing that the determinant on the left- and right-hand side of equation (4.15) must be equal results in $\lambda_i = 1$. Therefore equation (4.15) results in linear equations in the elements of ω^*.

When the position of the plane at infinity (i.e. the affine calibration) is not known, this equation can not be used immediately. From (3.17) it is known that the homography of the plane at infinity can be expressed as follows[2]:

$$\mathbf{H}^\infty_{1i} = \mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty^\top \quad . \tag{4.16}$$

Filling this in in (4.15) the following nonlinear equation is obtained:

$$\lambda_i \omega^* = [\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty^\top]\omega^*[\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty^\top]^\top \quad . \tag{4.17}$$

The following equation imposes that the determinants for the left and the right-hand side of (4.17) are equal:

$$\lambda_i^3 |\omega^*| = |\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty|.|\omega^*|.|\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty| \, , \tag{4.18}$$

with $|*|$ representing the determinant. This allows us to obtain a closed form expression for λ_i:

$$\lambda_i = |\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty|^{\frac{2}{3}} \quad . \tag{4.19}$$

Using expression (4.19), equation (4.17) can now be used to determine π_∞ and ω^* and thus also $\mathbf{K}$. It is proposed to use the following criterion for minimization:

$$\mathcal{C} = \sum_{i=2}^{n} \left\| \mathbf{K}\mathbf{K}^\top - \lambda_i [\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty^\top]\mathbf{K}\mathbf{K}^\top[\mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi_\infty^\top]^\top \right\|_F \quad . \tag{4.20}$$

---

[2] Note that the special case $\Pi_\infty = [\pi_\infty 0]$ which can't be expressed through this parameterization, can immediately be discarded, since this would mean that the first camera was placed at infinity in the real world.

An alternative to the use of equation (4.19) is to normalize both parts of equation (4.17) to a Frobenius norm of 1. If we define the matrix operator $\mathbf{F}(\mathbf{A}) \equiv \frac{\mathbf{A}}{\|\mathbf{A}\|_F}$, then the following criterion is obtained:

$$\mathcal{C}' = \sum_{i=2}^{n} \left\| \mathbf{F}(\mathbf{K}\mathbf{K}^\top) - \mathbf{F}([\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]\mathbf{K}\mathbf{K}^\top[\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]^\top) \right\|_F \quad . \tag{4.21}$$

It can be shown that the equations proposed by Heyden and Åström [60] are equivalent with the alternative constraints proposed in [124]. Starting from equation (4.17),

$$\lambda_i \mathbf{K}\mathbf{K}^\top = [\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]\mathbf{K}\mathbf{K}^\top[\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]^\top \tag{4.22}$$

and rewriting this equation, (4.12) can easily be obtained (using $\mathbf{a} = -\mathbf{K}^\top\pi_\infty$):

$$\begin{aligned} \lambda_i \mathbf{K}\mathbf{K}^\top &= [\mathbf{H}_{1i}^{\mathtt{REF}} \,|\, \mathtt{e}_{1i}] \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & -\mathbf{K}\mathbf{K}^\top\pi_\infty \\ -\pi_\infty^\top\mathbf{K}\mathbf{K}^\top & \pi_\infty\mathbf{K}\mathbf{K}^\top\pi_\infty \end{bmatrix} \begin{bmatrix} \mathbf{H}_{1i}^{\mathtt{REF}\top} \\ \mathbf{e}_{1i}^\top \end{bmatrix} \\ &= \mathbf{P} \begin{bmatrix} \mathbf{K}\mathbf{K}^\top & \mathbf{K}\mathbf{a} \\ \mathbf{a}^\top\mathbf{K}^\top & ||\mathbf{a}||^2 \end{bmatrix} \mathbf{P}^\top \quad . \end{aligned} \tag{4.23}$$

Since both constraints are algebraically equivalent, this technique also suffers from a bias towards the first image. This bias can be eliminated by not relating the dual images of the absolute conic ω_i^* to its image in the first image ω_1^*, but directly to the dual absolute conic ω_∞^* itself. In a camera centered world frame (i.e. with $\mathbf{P}_1 = [\mathbf{I}_{3\times3} \,|\, 0_3]$) the plane at infinity can easily be parameterized so that the homographies from image 1 to i (i.e. $\mathbf{H}_{1i}^\infty$) and the homographies which map the plane at infinity in image i (i.e. $\mathbf{H}_{\infty i}$) are the same. In this case the equivalent to equation (4.22) is

$$\lambda_i \mathbf{K}\mathbf{K}^\top = [\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]\omega_\infty^*[\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]^\top \,, \tag{4.24}$$

which results in the following criterion to minimize:

$$\mathcal{C}''(\mathbf{K}, \pi_\infty, \omega_\infty^*) = \begin{array}{r} \left\| \mathbf{F}(\mathbf{K}\mathbf{K}^\top) - \mathbf{F}(\omega_\infty^*) \right\|_F \\ + \sum_{i=2}^{n} \left\| \mathbf{F}(\mathbf{K}\mathbf{K}^\top) - \mathbf{F}([\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]\omega_\infty^*[\mathbf{H}_{1i}^{\mathtt{REF}} - \mathtt{e}_{1i}\pi_\infty^\top]^\top) \right\|_F \end{array} \tag{4.25}$$

The Kruppa equations

The first method that was proposed for self-calibration [95, 37] is based on the Kruppa equations [77]. These equations are similar to the equations (4.3) and (4.5), but are restricted to the epipolar geometry. The Kruppa equations impose that the epipolar lines which correspond to the epipolar planes tangent to the absolute conic, should be tangent to its projection in both images. This is illustrated in Figure 4.2.

The Kruppa equation can be derived starting from equation (4.5) which is equivalent to

$$\mathtt{l}^\top\mathbf{K}\mathbf{K}^\top\mathtt{l} = 0 \Leftrightarrow \mathtt{l}^\top\mathbf{H}_{ij}^\infty\mathbf{K}\mathbf{K}^\top\mathbf{H}_{ij}^{\infty\top}\mathtt{l} = 0, \,\forall \mathtt{l} \quad . \tag{4.26}$$

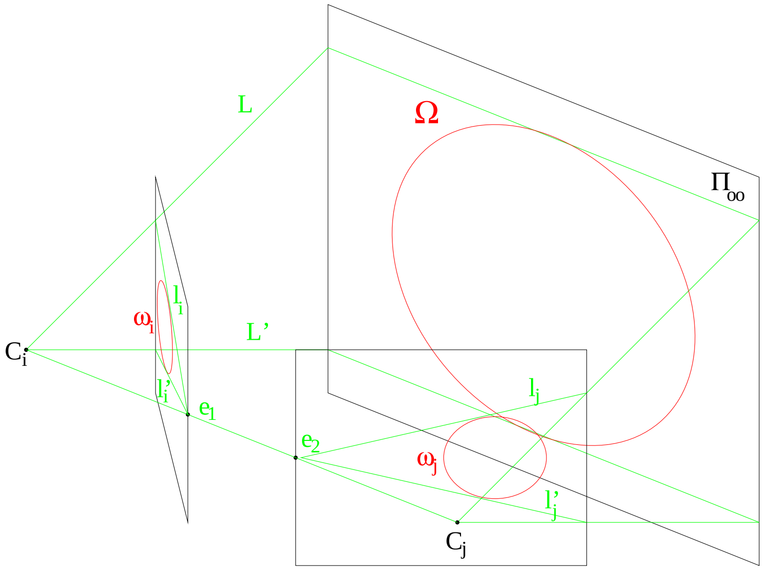

Figure 4.2: *The Kruppa equations impose that the image of the absolute conic satisfies the epipolar constraint. In both images the epipolar lines corresponding to the two planes through* $\mathtt{C}_i$ *and* $\mathtt{C}_j$ *tangent to* Ω *must be tangent to the images* ω_i *and* ω_j.

If one is only interested in epipolar lines, l should be parameterized as $[\mathtt{e}_{ij}]_\times \mathtt{m}$ and therefore one obtains:

$$\mathtt{m}^\top[\mathtt{e}_{ij}]_\times^\top \mathbf{K}\mathbf{K}^\top[\mathtt{e}_{ij}]_\times\mathtt{m} = 0 \Leftrightarrow \mathtt{m}^\top[\mathtt{e}_{ij}]_\times^\top \mathbf{H}^\infty_{ij}\mathbf{K}\mathbf{K}^\top\mathbf{H}^{\infty\top}_{ij}[\mathtt{e}_{ij}]_\times\mathtt{m} = 0,\ \forall\mathtt{m}; \tag{4.27}$$

which, using equation (3.28), yields

$$[\mathtt{e}_{ij}]_\times^\top \mathbf{K}\mathbf{K}^\top[\mathtt{e}_{ij}]_\times \sim \mathbf{F}_{ij}\mathbf{K}\mathbf{K}^\top\mathbf{F}_{ij}^\top \tag{4.28}$$

From the 5 equations obtained here only 2 are independent [183]. Scale factors can be eliminated by cross-multiplication. It is possible to use a minimization criterion similar to equation (4.21) to solve the self-calibration problem:

$$\mathcal{C}_K(\mathbf{K}) = \sum_{i=2}^{n}\sum_{j=i}^{n}\left\|\mathbf{F}(\mathbf{K}\mathbf{K}^\top) - \mathbf{F}(\mathbf{F}_{ij}\mathbf{K}\mathbf{K}^\top\mathbf{F}_{ij}^\top)\right\|_F \tag{4.29}$$

An interesting feature of this self-calibration technique is that no consistent projective reconstruction should be available, only pairwise epipolar calibration. This can be very useful is some cases where it is hard to relate all the images into a single projective frame.

The price that is paid for this advantage is that 3 of the 5 absolute conic transfer equations are used to eliminate the dependence on the position of the plane at infinity.

This explains why this method performs poorly compared to others when a consistent projective reconstruction can be obtained (see Section 5.3.4).

The Kruppa equations do not enforce directly a consistent supporting plane Π_∞ for the absolute conic. In other words, if one would estimate the position of the plane at infinity once $\mathbf{K}$ has been found, a slightly different solution would be found for every pair of images. In fact some specific degeneracies exist for the Kruppa equations (see Sturm [150]).

Hartley's method

A few years ago Hartley proposed an alternative self-calibration method [53]. This method is not based on the absolute conic, but directly uses a QR-decomposition of the camera projection matrices. Hartley derives an equation similar to (4.11):

$$\mathbf{P}_i \begin{bmatrix} \mathbf{I} \\ \pi_\infty^\top \end{bmatrix} \mathbf{K} \sim \mathbf{K}_i \mathbf{R}_i \tag{4.30}$$

where $\mathbf{K}$ and π_∞ are the unknowns. It is proposed to compute $\mathbf{K}_i$ through QR-decomposition of the left-hand side of equation (4.30). The following equation should be roughly satisfied for the solution:

$$\mathbf{K} \approx \mathbf{K}_i \text{ or } \mathbf{K}^{-1}\mathbf{K}_i \approx \mathbf{I} \tag{4.31}$$

The proposed minimization criterion is the following:

$$\mathcal{C}_H(\mathbf{K}, \pi_\infty) = \sum_{i=2}^{n} \|\alpha_i \mathbf{K}^{-1}\mathbf{K}_i - \mathbf{I}\|_F \tag{4.32}$$

where α_i is chosen such that the sums of squares of diagonal elements of both terms are equal. Note that this criterion is very close to the following alternatives

$$\mathcal{C}'_H(\mathbf{K}, \pi_\infty) = \sum_{i=2}^{n} \left\|\mathbf{F}(\mathbf{K}^{-1}\mathbf{K}_i) - \mathbf{F}(\mathbf{I})\right\|_F \tag{4.33}$$

or even

$$\mathcal{C}''_H(\mathbf{K}, \pi_\infty) = \sum_{i=2}^{n} \|\mathbf{F}(\mathbf{K}_i) - \mathbf{F}(\mathbf{K})\|_F \tag{4.34}$$

which are closer in notation to the other criteria used in this chapter.

The main difference between Hartley's method and the other is that the rotational component is eliminated through QR-decomposition instead of through multiplication by the transpose.

4.4.2 Restricted motions

For self-calibration some restricted motions can be very interesting. Restricted motions can result in simpler algorithms; but, on the other hand, it is not always possible to retrieve all the calibration parameters from these motions. Some specific methods which take advantage of restricted motions are discussed in the following paragraphs. The occurrence of degenerate cases for self-calibration due to restricted motions will be discussed in section 4.4.3.

Pure translation

Van Gool et al. [173] (see also Moons et al. [100]) proposed a simple algorithm to obtain an affine reconstruction in the case of pure camera translation. In this case, the metric camera projection matrices can be chosen as $\mathbf{P}_1 = \mathbf{K}[\mathbf{I}\,|\,0\,]$ and $\mathbf{P}_2 = \mathbf{K}[\mathbf{I}\,|\,-\mathtt{t}]$ and thus the projection of a point $\mathtt{M} = [\mathtt{m}^\top 1]^\top$ can be described as follows:

$$\begin{aligned} \lambda_1 \mathtt{m}_1 &= \mathbf{K}[\mathbf{I}\,|\,0\,]\mathtt{M} && (4.35) \\ \lambda_2 \mathtt{m}_2 &= \mathbf{K}[\mathbf{I}\,|\,-\mathtt{t}]\mathtt{M} \\ &= \lambda_1 \mathtt{m}_1 + \lambda \mathtt{e} \\ & && (4.36) \end{aligned}$$

with $\mathtt{e} \sim -\mathbf{K}\mathtt{t}$ being the epipole in the first image. This means that corresponding points and the epipoles must be collinear. Therefore, two pairs of corresponding points are sufficient to retrieve the epipole. An affine reconstruction is easily obtained by noting that the following camera projection matrices

$$\mathbf{P}_1 = [\mathbf{I}_{3\times3}\,|\,0_3] \text{ and } \mathbf{P}_2 = [\mathbf{I}_{3\times3}\,|\,\mathtt{e}] \quad (4.37)$$

can be obtained from the metric ones through an affine transformation and thus represent an affine reconstruction.

Note that in this case $\mathbf{H}_{12}^{\infty} = \mathbf{I}_{3\times3}$ and that equation (4.5) is therefore trivially satisfied. Consequently, no constraints can be obtained to restrict $\mathbf{K}$ and to obtain a metric reconstruction.

When besides pure translation additional – more general – motions are available, a metric reconstruction can easily be obtained. Combining the results of Moons [100] and Hartley [55], Armstrong [2] proposed a stratified approach to metric reconstruction. A pure translation is carried out to obtain an affine reconstruction. The pose for the additional views is computed towards this affine reconstruction. These views are therefore also affinely calibrated. In this case the homographies for the plane at infinity $\mathbf{H}_{ij}^{\infty}$ are readily available and can be used in equation (4.5).

Pure rotation

Another interesting type of motion is pure rotation. However, in this case, no parallax will occur, since no translation is involved. In this case the fundamental matrix and the epipole can not be determined. On the other hand, the image displacements can be described by a homography, as shown next.

For a camera undergoing a pure rotation around its center of projection, the camera projection matrices can be written as $\mathbf{P}_1 = \mathbf{K}[\mathbf{I}_{3\times3}\,|\,0_3]$ and $\mathbf{P}_i = \mathbf{K}[\mathbf{R}_i\,|\,0_3]$. The following image points are obtained for a scene point $\mathtt{M} = [\mathtt{m}^\top \, d]^\top$:

$$\left.\begin{array}{l} \mathtt{m}_1 \sim \mathbf{K}\mathtt{m} \\ \mathtt{m}_i \sim \mathbf{K}\mathbf{R}_i\mathtt{m} \end{array}\right\} \Rightarrow \mathtt{m}_i \sim \mathbf{K}\mathbf{R}_i\mathbf{K}^{-1}\mathtt{m}_1 \tag{4.38}$$

independently of d. Therefore, the homography $\mathbf{H}_{1i} = \mathbf{K}\mathbf{R}_i\mathbf{K}^{-1}$ describes the image motion for all points. Note that this homography is also valid for the points at infinity and thus $\mathbf{H}^\infty_{1i} = \mathbf{H}_{1i} = \mathbf{K}\mathbf{R}_i\mathbf{K}^{-1}$. Equation (4.5) can therefore easily be used to determine the dual image of the absolute conic and the calibration parameters $\mathbf{K}$.

Planar motion

A general motion can be described as a rotation around and a translation along a screw axis. In the case of planar motions, only a rotation around a screw axis takes place. This means that the translation is in a plane which is orthogonal to the rotation axis. A sequence of motions is planar if every motion is planar and if all the rotation axes are parallel.

Armstrong et al. [3] (see also [4]) proposed to use this type of motion for self-calibration. Recently Faugeras et al. [33] have proposed a similar approach based on 1D projective cameras which live in the trifocal plane.

An important concept in self-calibration are fixed entities. For general motion there are two: the plane at infinity and the absolute conic. For sequences of planar motions more entities are fixed. The point at infinity of the rotation axis $\mathtt{V}$ and the horizon $\mathtt{H}$ (i.e. the vanishing line of the plane of the motion). Since intersections of fixed entities must also be fixed entities, the two points of the horizon which are on the absolute conic must be fixed entities too. These intersection points are called the circular points $\mathtt{I}$ and $\mathtt{J}$ of the plane. Together these 3 points are located in the plane at infinity Π_∞ and – if known – can be used to locate the plane at infinity.

Since vanishing points corresponding to orthogonal directions must lie on each others polar with respect to the absolute conic, and, because the polar of a point on this conic is the tangent through that point, the absolute conic must be tangent to the lines $\mathtt{l}_{\mathtt{vi}}$ and $\mathtt{l}_{\mathtt{vj}}$ in $\mathtt{i}$ and $\mathtt{j}$ respectively [142]. This is illustrated in Figure 4.3. This defines 4 constraints on the absolute conic which has 5 degrees of freedom. One additional constraint therefore suffices to uniquely identify the absolute conic ω_∞. Knowing that the pixels are rectangular (no skew) could be used, but is typically degenerate, because one of the camera axis often is parallel to the rotation axis. The aspect ratio, however, would be very useful to fix the last degree of freedom.

Since – under the considered motions – these entities are fixed in space, their images should be fixed too. If these images can be extracted, the scene points can be obtained through back-projection and triangulation.

To design a practical algorithm for this, it is needed to extract these fixed entities from the images. The horizon is easily extracted as the image of the plane through all the centers of projection.

The horopter [94] is defined as the set of points which have the same image in two views. Its image is given by the conic defined by the symmetric part of the fundamental matrix. In the case of a planar motion, it consists of a two-line conic [4]; one of

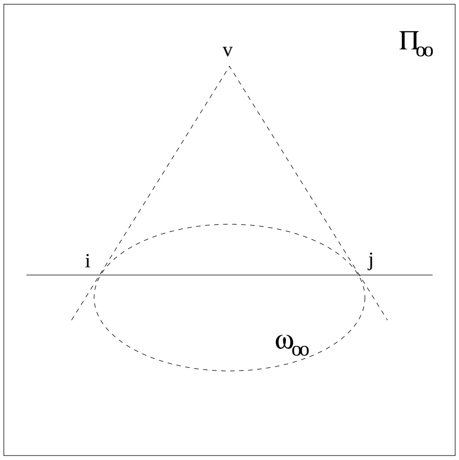

Figure 4.3: *The points* `v`, `i` *and* `j` *are the fixed points of the plane at infinity. Since the direction of* `v` *is orthogonal on the direction defined by* `i` *and* `j` *and these two points are located on the absolute conic* ω_∞, *this conic must be tangent to the lines* $\mathtt{l_{vi}}$ *and* $\mathtt{l_{vj}}$.

them being the horizon, the other being the image of the rotation axis. They are easily differentiated, because the epipoles should be located on the horizon. The vanishing point `v` is located at the intersection of the images of all the rotation axes.

Both circular points `i` and `j` are located on the horizon. They must have the same image in all three images. Let us consider `i`. Taking two arbitrary points $\mathtt{h}_1$ and $\mathtt{h}_2$ of the horizon, the image of this circular point can be parameterized as $\mathtt{i}(\lambda_i) = \mathtt{h}_1 + \lambda_i \mathtt{h}_2$. This image must satisfy the trifocal point-transfer equation (3.31):

$$\mathtt{i}(\lambda_i)_k \left\{ \mathtt{i}(\lambda_i)_l \mathtt{i}(\lambda_i)_m \mathbf{T}_{k33} - \mathtt{i}(\lambda_i)_m \mathbf{T}_{kl3} - \mathtt{i}(\lambda_i)_l \mathbf{T}_{k3m} + \mathbf{T}_{klm} \right\} = 0 \qquad (4.39)$$

This equation is a cubic expression in λ_i. It has in general three solutions, all real or one real and two complex conjugated. In our case, two will be complex, corresponding to `i` and `j`.

4.4.3 Critical motion sequences

It was noticed very soon that not all motion sequences are suited for self-calibration. Some obvious cases are the restricted motions described in the previous section (i.e. pure translation, pure rotation and planar motion). There are, however, more motion sequences which do not lead to unique solutions for the self-calibration problem. This means that at least two reconstructions are possible which satisfy all constraints on the camera parameters for all the images of the sequence and which are not related by a similarity transformation.

Several researchers realized this problem and mentioned some specific cases or did a partial analysis of the problem [166, 183, 127]. Sturm [152, 153] provided a complete catalogue of critical motion sequences (CMS) for constant intrinsic parameters. Additionally, he identified specific degeneracies of some algorithms [150].

Due to the importance of this work in the context of self-calibration and to the new results which are based on this analysis (e.g. Theorem 4.2 and Section 6.4), an extended review of this work is given here.

The absolute conic is a proper virtual conic (PVC) [10]. Problems occur when the absolute conic is not the only PVC which has a fixed projection in all the images. In that case, there is no way to determine which one corresponds to the real absolute conic and the motion sequence is said to be *critical* with respect to self-calibration.

The main idea in Sturm's approach is to turn the problem around and to look for the motion sequences which leave a specific PVC unchanged. The PVC's can be classified in a few different types.

Potential absolute conic on Π_∞ First one should consider the PVC's on the plane at infinity. These can have either a triple eigenvalue, a double and a single one or three distinct eigenvalues. Only the absolute conic itself corresponds to the first case, which is therefore not critical. In the second case, the eigenspace corresponds to a plane and a line orthogonal to that plane. An arbitrary rotation around that line, or, a rotation of 180^o around a line in the plane which is incident to the other line, will leave the PVC unchanged, as does, of course, also the identity transformation. In the third case, the eigenspace consists of three mutually orthogonal lines. Besides the identity transformation, only rotations of 180^o around one of these lines leave the PVC unchanged.

From these cases a first class of CMS is obtained.

CMS-Class 1: Motion sequences for which all rotations are by an arbitrary angle about an axis parallel with a specific line or by 180^o about an axis perpendicular to this line.

Sturm identifies several subclasses which correspond to more restricted motions, but therefore to more potential absolute conics.

CMS-Class 1.1 Motion sequences for which all rotations are by an arbitrary angle about an axis parallel with a specific line.

CMS-Class 1.2 Motion sequences for which all rotations are by 180^o about mutually orthogonal axes.

CMS-Class 1.3 Motion sequences for which all rotations are by 180^o about some specific axis.

CMS-Class 1.4 Motion sequences for which no rotation takes place at all (i.e. pure translations).

Potential absolute conic not on Π_∞ In this case, Sturm proposes to start from a PVC and from a specific camera pose and to look for all the other camera poses which have the same image for the PVC under consideration. The PVC can be both a circle or an ellipse. As an image of a PVC has a projection cone associated to it, one can just as well look at possible poses for this cone so that it contains the PVC.

The possible orientations for the camera depend on the Euclidean normal form of the projection cone[3]. Any rotation around the vertex of an absolute cone leaves the cone globally unchanged. Arbitrary rotations about the main axis of a circular cone or rotations by 180^o about an axis perpendicular but incident with the main axis leave this cone unchanged. Rotations by 180^o degrees around an axis of an elliptic cone leave it unchanged.

The different combinations of cone types and conic types should be considered. These result in a number of CMS classes.

CMS-Class 2 For a *circle* and an *elliptic cone* infinitely many positions are possible. These are located on two parallel circles. At each position 4 orientations are possible.

CMS-Class 3 For an *ellipse* and an *elliptic cone* similar to Class 2, but located on a degree 12 curve [150]. At each position 4 orientations are possible.

A *circle* and a *circular cone* result in 2 possible positions. At each position the camera may rotate freely about the line joining the projection centers or by 180^o about a line orthogonal to it. This is not classified as a separate class since the combination of a circle with an absolute cone will result in a more general class. This class could be seen as Class 5.1.

CMS-Class 4 An *ellipse* and a *circular cone* result in 4 positions in which the camera may freely rotate about the main axis of the cone and by 180^o about an axis perpendicular to it.

An *ellipse* combined with an *absolute cone* is not feasible since all planar sections of this cone are circles.

CMS-Class 5 The combination of a *circle* with an *absolute cone* results in two possible positions with arbitrary orientations, since the orthogonal projection of the vertex on the supporting plane of the circle must coincide with the center of the circle.

Sturm identified some more specific subclasses which are practically important. An overview of these different classes and subclasses is given in Table 4.1. In Figure 4.4 some practically important CMS are illustrated.

To get some intuitive geometric insight in these concepts, the reader is referred to Appendix C (an example is shown in Figure 6.5).

Practical use of the CMS classification The classification of all possible critical motion sequences is very important and it can be used to avoid critical motions when acquiring an image sequence on which one intends to use self-calibration. In some cases, however, an uncalibrated image sequence is available from which a metric reconstruction of the recorded scene is expected. In this case, it is not always clear what can be achieved nor if the motion sequence is critical or not.

Sturm [153] showed that the recovered motion sequence for any reconstruction satisfying the fixed absolute conic image constraint would consist of rigid motions (i.e. Euclidean motions). This result is also valid for critical motion sequences, where

---

[3] There are three types of imaginary cones: an *absolute cone* has one triple eigenvalue, a *circular cone* has a single and a double eigenvalue, and an *elliptic cone* has three single eigenvalues. In the case of a circular cone the main axis is the one associated with the single eigenvalue.

Class	Description	$\#\mathtt{t}$	$\#\mathbf{R}$	$\#\Omega$
1		∞^3	$2 \times \infty$	∞
1.1		∞^3	∞	∞
1.1.1	planar motions	∞^2	∞	∞
1.4	pure translations	∞^3	1	∞^5
2		$2 \times \infty$	4	∞
2.1	orbital motions	∞	1	∞^2
3		8	4	4
4		4	$2 \times \infty$	3
5		2	∞^3	2
5.2	pure rotations	1	∞^3	∞^3

Table 4.1: *Classes of critical motion sequences.* $\#\mathtt{t}$ *and* $\#\mathbf{R}$ *represent respectively the number of different positions and different orientations at each position of which a critical motion sequence of the specific class can consist.* $\#\omega_\infty$ *indicates the minimum level of ambiguity on the absolute conic for this class of critical motion sequences.*

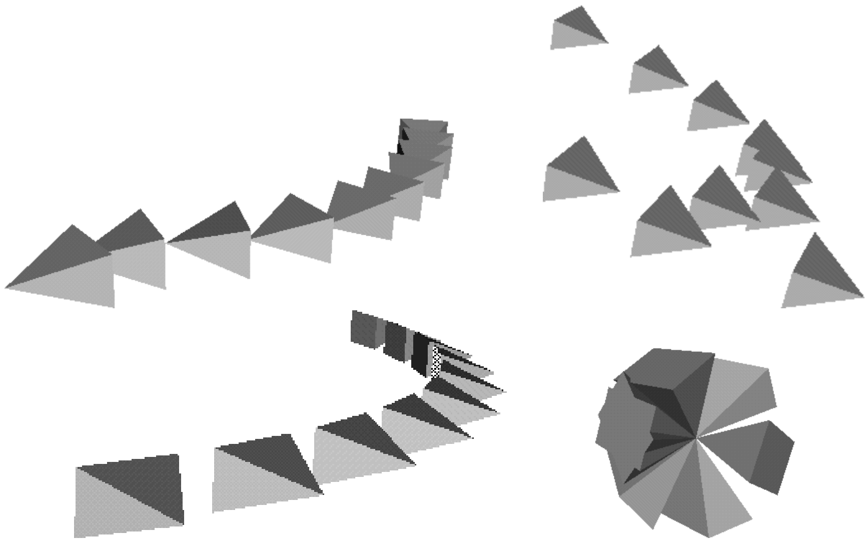

Figure 4.4: *Illustration of critical motion sequences. The cameras are represented by small pyramids. Planar motion (top left), pure translation (top right), orbital motion (lower left) and pure rotation (lower right).*

the recovered motion sequence would be in the same CMS class as the original sequence. This is an important observation, because it allows to identify critical motions sequences and to determine the ambiguity on the reconstruction from any valid instantiation of the reconstruction. In that case more specific algorithms can be called or additional constraints can be brought in to reduce the ambiguity [189]. Here we give a simpler proof than in [153]. This proof is based on the disc quadric representation.

Theorem 4.1 *Let S be a motion sequence that is critical with respect to the dual quadric* Φ^**, and let* $\mathbf{P}_{Ei}$ *be the original projection matrices of the frames in S. Let* $\mathbf{T}$ *be any projective transformation mapping* Φ^* *to* Ω^* *and* $\mathbf{P}_{Pi} = \mathbf{P}_{Ei}\mathbf{T}^{-1}$ *be the projection matrices transformed by* $\mathbf{T}$*. There exists a Euclidean transformation between any pair of* $\mathbf{P}_{Pi}$*.*

Proof: From S being a critical motion sequence with respect to Φ^*, it follows that there must exist a $\mathbf{K}_P$ for which

$$\mathbf{K}_P\mathbf{K}_P^\top \sim \phi_i^* \sim \mathbf{P}_{Ei}\Phi^*\mathbf{P}_{Ei}^\top$$

Since $\Phi^* \sim \mathbf{T}^{-1}\Omega^*\mathbf{T}^{-\top}$ and $\mathbf{P}_{Pi} = \mathbf{P}_{Ei}\mathbf{T}^{-1}$, one gets

$$\mathbf{K}_P\mathbf{K}_P^\top \sim \mathbf{P}_{Pi}\Omega^*\mathbf{P}_{Pi}^\top$$

Defining $\mathbf{H}_{Pi}$ as the left 3×3 part of $\mathbf{P}_{Pi}$ this yields

$$\mathbf{K}_P\mathbf{K}_P^\top \sim \mathbf{H}_{Pi}\mathbf{H}_{Pi}^\top \text{ or } \mathbf{I} \sim \mathbf{K}_P^{-1}\mathbf{H}_{Pi}\mathbf{H}_{Pi}^\top\mathbf{K}_P^{-\top}$$

Since the matrix $\mathbf{K}_P^{-1}\mathbf{H}_{Pi}$ satisfies the orthonormality constraints, it must correspond to some rotation matrix, say $\mathbf{R}_{Pi}$. Thus $\mathbf{H}_{Pi} = \mathbf{K}_P\mathbf{R}_{Pi}$. Therefore, it is always possible to write the projection matrices $\mathbf{P}_{Pi}$ as follows:

$$\mathbf{P}_{Pi} = \mathbf{K}_P[\mathbf{R}_{Pi}^\top \,|\, \text{-}\mathbf{R}_{Pi}^\top\mathtt{t}_{Pi}]$$

□

This means that the sequence S_P consisting of $\mathbf{P}_{Pi}$ is Euclidean and has, after transformation by $\mathbf{T}$, the same set of potential absolute conics. Since the different classes of CMS given in [152] can't be transformed into each other through a projective transformation, the sequence S_P will be a CMS of the same class as S. Therefore, one can conclude that any reconstruction being a solution to the self-calibration problem allows us to identify the class of CMS of the original sequence and thus also all ambiguous reconstructions.

A question that was not answered though, is: *What can still be done with an ambiguous reconstruction?* The answer is given by the next theorem.

But let us first define $C(S)$ as being the set of potential absolute quadrics for the motion sequences S. Let us also define the transformation of S as the sets of the transformed elements.

$$S = \{\left[\begin{array}{cc} \mathbf{R} & \mathtt{t} \\ \mathtt{0}_3^\top & 1 \end{array}\right], \ldots\} \quad \rightarrow \quad \mathbf{T}S = \{\mathbf{T}\left[\begin{array}{cc} \mathbf{R} & \mathtt{t} \\ \mathtt{0}_3^\top & 1 \end{array}\right]\} \tag{4.40}$$

Theorem 4.2 *Let S be a critical motion sequence and $\mathbf{P}_{Ei}$ the corresponding projection matrices. Let Φ^* be an arbitrary element of $C(S)$ and let $\mathbf{T}$ be an arbitrary projective transformation mapping Φ^* to Ω^*. Let $S_P = \mathbf{T}S$ and $\mathbf{P}_{Pi} = \mathbf{P}_{Ei}\mathbf{T}^{-1}$. Let M represent a Euclidean motion for which $C(S_P \cup M) = C(S_P)$ and let $\mathbf{P}_{Pnew}$ be the corresponding projection matrix. Then there exists a Euclidean transformation between $\mathbf{P}_{Enew} = \mathbf{P}_{Pnew}\mathbf{T}$ and any other $\mathbf{P}_{Ei}$.*

Proof: From $\Omega^* \in C(S)$, it follows that $\mathbf{T}\Omega^*\mathbf{T}^\top \in C(S_P)$. Since it is assumed that $C(S_P \cup M) = C(S_P)$, it follows that Theorem 4.1 can be applied to the sequence $S_P \cup M$, with the dual quadric $\mathbf{T}\Omega^*\mathbf{T}^\top$, the transformation $\mathbf{T}^{-1}$ and $\{\mathbf{P}_{P1}, \ldots, \mathbf{P}_{Pn}, \mathbf{P}_{Pnew}\}$ as so-called original projection matrices. □

This theorem allows us to conclude that it is possible to generate correct new views, even starting from an ambiguous reconstruction. In this case, we should, however, restrict the motion of the virtual camera to the type of the critical motion sequence recovered in the reconstruction. For example, if we have acquired a model by doing a planar motion on the ground plane and thus rotating around vertical axes, then we should not move the camera outside this plane nor rotate around non-vertical axes. But, if we restrict our virtual camera to this critical motion, then all these motions will correspond to Euclidean motions in the real world and no distortion will be present in the images (except, of course, for modeling errors). Note that the recovered camera parameters should be used (i.e. the ones obtained by factorizing $\mathbf{P}_{Pi}$ in $\mathbf{K}_P[\mathbf{R}_{Pi}^\top \,|\, \text{-}\mathbf{R}_{Pi}^\top \mathtt{t}_{Pi}]$).

4.5 Conclusion

In this chapter different aspects of self-calibration were discussed. First, some general concepts were introduced. Different methods which allows to solve the classical self-calibration problem were briefly presented and discussed. All these methods assume that the motion is sufficiently general so that a unique solution can be obtained.

In some cases the motion of the camera is restricted. On the one hand, this can lead to simpler self-calibration algorithms. The cases of pure translations, pure rotations and planar motions were discussed. On the other hand, restricted motions can cause problems in the sense that in these cases a unique solution to the self-calibration problem can not always be found.

Due to the consequences which these restricted motion sequences have for self-calibration, this problem was discussed in detail. An interesting theorem was derived which tells us that even for critical motion sequences it is possible to generate correct new views, however not from all possible poses.

Chapter 5

Stratified self-calibration

5.1 Introduction

In recent years several methods were proposed to obtain the calibration of a camera from correspondences between several views of the same scene. Most of these were presented in the previous chapter. These methods are based on the rigidity of the scene and on the constancy of the intrinsic camera parameters. Most existing methods start from the projective calibration and then immediately try to solve for the intrinsic parameters. However, they all have to cope with the affine parameters (i.e. the position of the plane at infinity).

Faugeras et al. [37] eliminated these affine parameters yielding two Kruppa equations for each pair of views. A more robust approach was proposed by Zeller and Faugeras [183]. Heyden and Åström [60] and Triggs [166] proposed methods based on the absolute quadric. Hartley [53] does a minimization on all eight parameters to obtain metric projection matrices. Most of these methods encounter problems as they have to solve for many parameters at once from nonlinear equations.

This problem prompted a stratified approach, where starting from a projective reconstruction an affine reconstruction is obtained first and used as the initialization towards metric reconstruction. A similar method has been proposed by Armstrong et al. [2] based on the work of Moons et al. [100]. But this method needs a pure translation which is for example not easy to ensure with a hand-held camera. Successful stratified approaches have also been proposed for the self-calibration of fixed stereo rigs by Zisserman et al. [188], Devernay and Faugeras [28] and more recently by Horaud and Csurka [65].

A first general approach for a single camera based on the modulus constraint needed at least four views [122] to obtain the self-calibration. The method was developed further in [123] and [111]. This enhanced method is more robust and can obtain the metric calibration of a camera setup from only three images. This chapter also discusses the possibility of using only two views of the scene. It will be shown that combining constraints on the intrinsic camera parameters with characteristics inferred from the scene can solve the calibration where none of them separately could.

Finally the modulus constraint can also be used to obtain the self-calibration in spite of a varying focal length [121].

An alternative approach to self-calibration was recently proposed [120] and will be discussed in the next chapter. In cases where all the intrinsic camera parameters except the focal length are (approximately) known, a solution can be obtained through a linear algorithm (allowing a varying focal length). This solution is then used as an initialization for a non-linear algorithm which allows varying/constant/known camera parameters. This algorithm therefore offers more flexibility, but it also fails in certain cases. The main reason for this is the linear parameterization for Ω^* which does not enforce the rank 3 constraint and therefore suffers from more general classes of critical motions. In the cases where both methods are applicable and succeed in their initialization phase the final results will in general be identical since the refinement step is the same for both methods.

This chapter is organized as follows. In section 5.2 the modulus constraint is derived. Section 5.3 explains how the self-calibration problem can be solved using the modulus constraint. In section 5.4 two other applications of the modulus constraint are presented. In this section a similar method for stereo rigs is also presented [125, 126].

5.2 The modulus constraint

A stratified approach to self-calibration first requires a method to identify the plane at infinity. The property of the homographies for this plane –called infinity homographies in the remainder of this text– derived in this paragraph will be used for this purpose. The infinity homography from view i to j can be written as a function of the metric entities of equation (3.15) or explicitly as functions of projective entities and the position of the plane at infinity in the specific projective reference frame (see equation (3.17)). Both representations are given in the following equation:

$$\mathbf{H}^\infty_{ij} \sim \mathbf{H}^\infty_{1j}\mathbf{H}^{\infty\,-1}_{1i} \sim \underbrace{\mathbf{K}\mathbf{R}_j^\top\mathbf{R}_i^{-\top}\mathbf{K}^{-1}}_{metric} \sim \underbrace{(\mathbf{H}_{1j} - \mathsf{e}_{1j}\pi_\infty^\top)(\mathbf{H}_{1i} - \mathsf{e}_{1i}\pi_\infty^\top)^{-1}}_{projective} \quad . \tag{5.1}$$

From the Euclidean representation it follows that $\mathbf{H}^\infty_{ij}$ is conjugated[1] with a rotation matrix (up to a scale factor) which implies that the 3 eigenvalues of $\mathbf{H}^\infty_{ij}$ must have the same moduli, hence the modulus constraint [121, 122]. Note from equation (5.1) that this property requires the intrinsic camera parameters to be constant.

This can be made explicit by writing down the characteristic equation for the infinity homography:

$$\det(\mathbf{H}^\infty_{ij} - \lambda\mathbf{I}) = l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = 0 \tag{5.2}$$

It can be shown that the following condition is a necessary condition for the roots of equation (5.2) to have equal moduli (see Appendix A.1):

$$l_3 l_1^3 = l_2^3 l_0 \tag{5.3}$$

---

[1] Matrices $\mathbf{A}$ and $\mathbf{B}$ are conjugated if $\mathbf{A} = \mathbf{C}\mathbf{B}\mathbf{C}^{-1}$ for some matrix $\mathbf{C}$. Conjugated matrices have the same eigenvalues.

This yields a constraint on the 3 affine parameters contained in π_∞ by expressing l_3, l_2, l_1, l_0 as a function of these parameters (using the projective representation in equation (5.1)). The inverse in the projective representation of $\mathbf{H}_{ij}^\infty$ can be avoided by using the constraint $\det\left(\mathbf{H}_{1j}^\infty - \lambda \mathbf{H}_{1i}^\infty\right) = 0$ which is equivalent to equation (5.2). Factorizing this expression using the multi-linearity of determinants, l_3, l_2, l_1, l_0 turn out to be *linear* in the elements of π_∞ (see Appendix A.2). Therefore between every pair of views a modulus constraint is obtained:

$$\mathcal{M}_{ij} : {l_3}_{ij} {l_1}_{ij}^3 - {l_2}_{ij}^3 {l_0}_{ij} = 0 \tag{5.4}$$

This results in a polynomial equation of degree four in the coefficients of π_∞. In general three of these constraints should only leave a finite number of solutions – not more than 64. With more equations only one possible solution should persist except for some critical motion sequences (e.g. sequences with rotation about parallel axes). In the case of planar motion the stratified approach introduces an additional ambiguity (i.e. the plane at infinity can not be identified uniquely in this case). See the work of Sturm [150] for a more complete discussion of this problem.

5.3 Self-calibration from constant intrinsic parameters

In section 5.2 a constraint on the location of the plane at infinity was derived for every pair of views. These constraints can thus be used for self-calibration. Once the affine structure has been recovered, it is easy to upgrade the structure to metric using the concepts of section 4.4.1.

5.3.1 Finding the plane at infinity

A minimum of three views is needed to allow for self-calibration using the modulus constraint. Three views yield the following constraints $\mathcal{M}_{12}, \mathcal{M}_{13}$ and $\mathcal{M}_{23}$. This third constraint $\mathcal{M}_{23}$ is in general independent of the two first constraints. The fact that $\mathbf{H}_{12}$ and $\mathbf{H}_{13}$ are both conjugated with rotation matrices (i.e. have the same eigenvalues as rotation matrices) does not imply that this is also the case for $\mathbf{H}_{23}$ since the eigenvectors of $\mathbf{H}_{12}$ and $\mathbf{H}_{13}$ could be different.

In the minimal case of three constraints $(\mathcal{M}_{12}, \mathcal{M}_{13}, \mathcal{M}_{23})$ for three unknown coefficients of π_∞ the most adequate method to solve for these unknowns is to use continuation. This is a numerical method which finds all the solutions of a system of polynomial equations (for more details see [102]). Having 3 polynomial equations of degree 4 a maximum of 64 solutions can be found.

Many of these solutions can be ruled out. Since only real solutions are of interest, all complex solutions can be discarded. In addition it should be verified that the eigenvalues of $\mathbf{H}_{ij}^\infty$ correspond to those of a rotation matrix (see Appendix A.1).

For the remaining solutions the intrinsic camera parameters can be computed using the method proposed in section 5.3.2. Only the solution yielding (quasi)-constant parameters should be taken into account. If more than one solution persist the most plausible one is chosen (no skew, aspect ratio around one, principal point around the middle of the image). This approach worked well for over 90% of the experiments (see Section 5.3.4).

When more views are at hand it is important to take advantage of all the available constraints to obtain a more accurate and robust solution. This can be done by combining the different constraints into a least squares criterion which can be minimized through nonlinear optimization:

$$\mathcal{C}_{MC} = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{M}_{ij}^2 \quad . \tag{5.5}$$

The recommended method for initialization of the minimization is the continuation method, but trying out a few random starting values often works in practice and allows a simpler algorithm. Alternatives consist of using Hartley's quasi-affine reconstruction [53] or to compute an approximate position for the plane at infinity from a priori knowledge on the intrinsic camera parameters.

5.3.2 Finding the absolute conic and refining the calibration results

Once the plane at infinity has been obtained, equation (4.6) can be used to compute the absolute conic as described in section 4.4.1. This is a linear method.

These results can be refined through a non-linear minimization step. The method described in [124] or in Section 4.4.1 can be used for this purpose. It is proposed to minimize criterion (4.25):

$$\mathcal{C}''(\mathbf{K}, \pi_\infty, \omega_\infty^*) = \begin{array}{c} \left\| \mathbf{F}(\mathbf{K}\mathbf{K}^\top) - \mathbf{F}(\omega_\infty^*) \right\|_F \\ + \sum_{i=2}^{n} \left\| \mathbf{F}(\mathbf{K}\mathbf{K}^\top) - \mathbf{F}([\mathbf{H}_{1i}^{\Pi} + \mathsf{e}_{1i}\pi_\infty^\top]\omega_\infty^*[\mathbf{H}_{1i}^{\Pi} + \mathsf{e}_{1i}\pi_\infty^\top]^\top) \right\|_F \end{array}$$

The implementation presented in this text uses a Levenberg-Marquard algorithm for the minimization.

Besides the stratified approach based on the modulus constraint (with and without refinement), the simulations of Section 5.3.4 were also carried out using some methods presented in the previous chapter. The criterion of equation (4.29) was used for the Kruppa equations [36] and the criterion of equation (4.25) for the methods of [60, 166]. In these cases the intrinsic camera parameters were initialized from a close guess.

5.3.3 Stratified self-calibration algorithm

step 1: projective calibration

step 1.1: sequentially compute the projective camera matrices for all the views (see [9] or Section 7.3).

step 2: affine calibration

step 2.1: formulate the modulus constraint $\mathcal{M}_{ij}$ for all pairs of views

step 2.2a: find a set of initial solutions through continuation

step 2.2b: (for $n > 3$) solve the set of equations $\mathcal{M}_{ij}$ through minimization of criterion $\mathcal{C}_{MC}$ (see eq. (5.5))

step 2.3: compute the affine projection matrices

$$\mathbf{P}_{Ai} = \mathbf{P}_i \mathbf{T}_{PA}^{-1} \text{ with } \mathbf{T}_{PA} = \begin{bmatrix} \mathbf{I}_{3\times3} & 0_3 \\ \pi_\infty^\top & 1 \end{bmatrix}$$

step 3: metric calibration

step 3.1: compute the dual image of the absolute conic from

$$\mathbf{K}\mathbf{K}^\top \sim \mathbf{H}_{\infty 1i}\mathbf{K}\mathbf{K}^\top\mathbf{H}_{\infty 1i}^\top$$

step 3.2: find the intrinsic parameters $\mathbf{K}$ through Cholesky factorization

step 3.3: compute the metric projection matrices

$$\mathbf{P}_{Mi} = \mathbf{P}_{Ai} \mathbf{T}_{AM}^{-1} \text{ with } \mathbf{T}_{AM} = \begin{bmatrix} \mathbf{K}^{-1} & 0_3 \\ 0_3^\top & 1 \end{bmatrix}$$

step 4: refined calibration

step 4.1 refine the results through nonlinear minimization of $\mathcal{C}''$ (see eq. (4.25))

step 4.2 compute the refined metric projection matrices

$$\tilde{\mathbf{P}}_{Mi} = \mathbf{P}_i \mathbf{T}_{PM}^{-1} \text{ with } \mathbf{T}_{AM} = \begin{bmatrix} \mathbf{K}^{-1} & 0_3 \\ \pi_\infty^\top & 1 \end{bmatrix}$$

5.3.4 Some simulations

The simulations were carried out on sequences ranging from 3 to 36 views. The scene consisted of 200 points which were positioned on 2 orthogonal 10×10 grids and then perturbed. For the calibration matrix the canonical form $\mathbf{K} = \mathbf{I}$ was chosen. The distance to the scene was chosen in such a way that the viewing conditions corresponded to a standard 35mm camera. The views were spaced 10^o apart, then a random perturbation was applied to their position and orientation. An example of such a sequence can be seen in Figure 5.1.

The scene points were projected onto the images. Gaussian noise with an equivalent standard deviations of 0, 0.2, 0.5, 1 and 2 pixels for 500×500 images was added to these projections. Hundred sequences were generated for every parameter combination. The projective reconstruction was obtained with a variant of the factorization method proposed in [154]. The self-calibration was carried out using the method proposed in this chapter. At first the modulus constraint was used to identify the plane at infinity, then the absolute conic was located. The results were then refined using the method of paragraph 5.3.2.

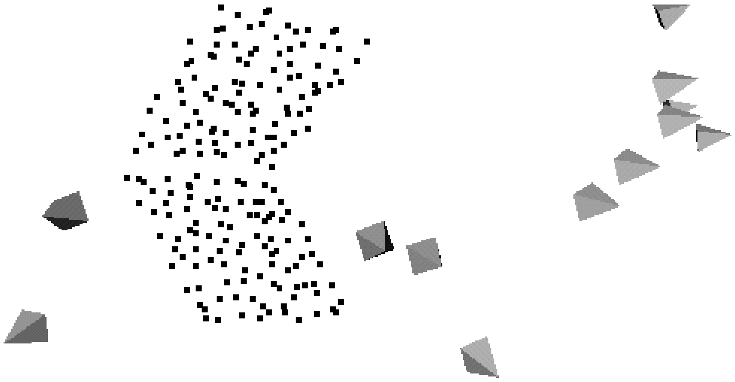

Figure 5.1: *Example of a sequence used for the simulations. Camera viewpoints are represented by small pyramids.*

For sake of comparison two alternative methods were included in the tests. The first method is based on the Kruppa equations [37], the second on a method similar to the methods using the absolute quadric [166, 60]. This second method in fact consists of immediately starting the refinement step of Section 5.3.2 on some prior guess. See Section 4.4.1 for more details on these methods. Both these algorithms were initialized with values close to the exact solution. The focal length was initialized randomly within a factor of 2, the aspect ratio, the principal point and the skew were only slightly perturbed (within ± 0.1) as proposed in [166, 60]. For all these implementations care was taken of normalizing the data according to the recommendations of Hartley [57].

The results of the experiments can be seen in Figure 5.2. The 2D back-projection error and the relative 3D error were computed. Since in a number of cases the algorithms fail to obtain satisfying results, the median values of the errors are shown. In addition the failure rate is given.

In Figure 5.3 some more results are given for the stratified approach. These results were obtained from 1000 experiments carried out on standard sequences of 12 views (0.5 pixels noise). The different subfigures are histograms of the obtained values for the different calibration parameters f_x, f_y, u_x, u_y, s, the layout was inspired by equation (3.4). The height of the peaks in the lower-right subfigure indicates how often the algorithm could find a solution (in this case 982 and 997 resp.). The left part gives the result obtained immediately with the modulus constraint and the right part after refinement. From this figure one can conclude that the calibration results are good and that there is no bias in the estimation. The refinement procedure clearly improves the results.

Notice that for short image sequences the 2D error is small while the 3D error is big. Using longer image sequences (and thus a wider angle between the extreme views) causes the 2D error to increase since an error on the intrinsic camera parameters can be compensated by distorting the structure, but only when the angle between the

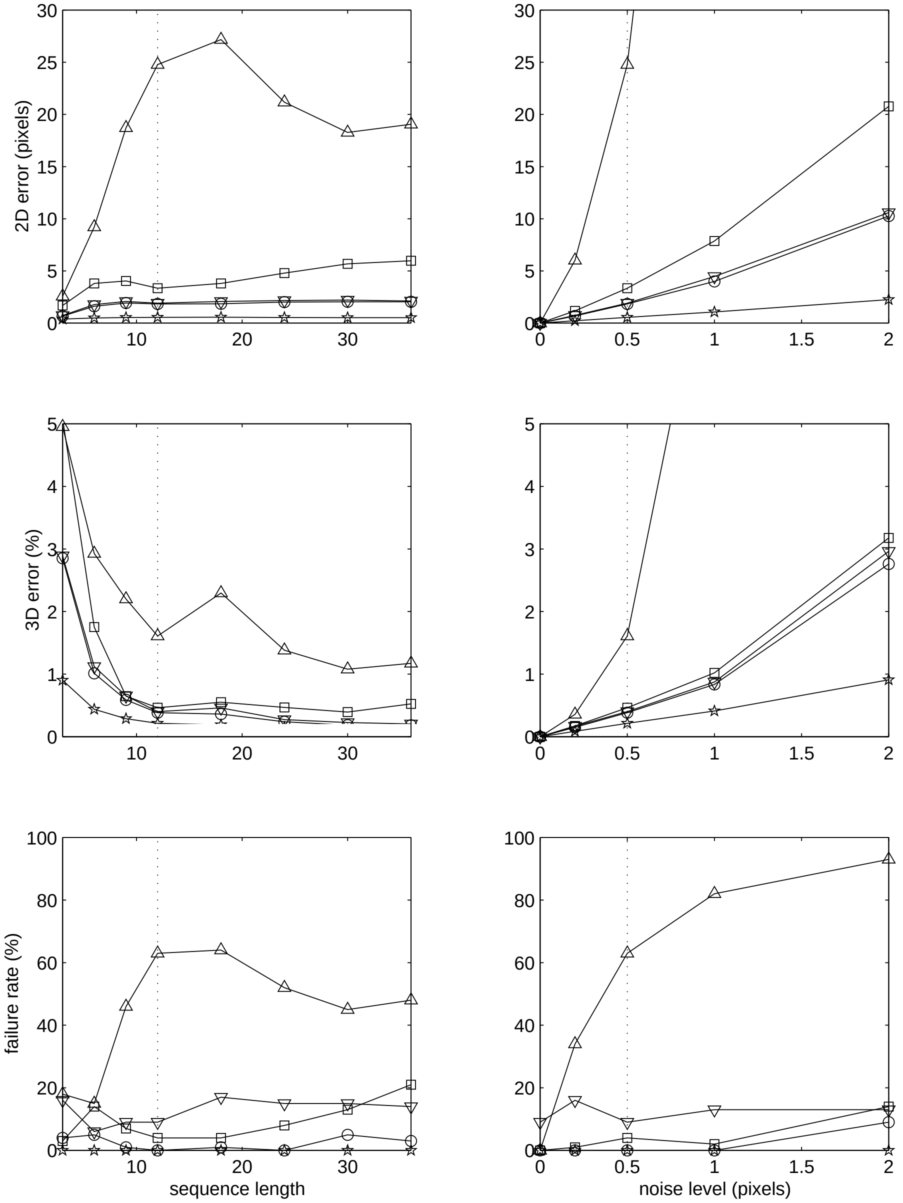

Figure 5.2: *Results of self-calibration experiment for image sequences varying from 3 to 36 views (left; noise 0.5 pixels) and noise ranging from 0.0 to 2.0 pixels (right; 12 views). The 2D reprojection error (top), the relative 3D error (middle) and the failure rate (bottom) are given. The median of the RMS error for 100 experiments was used as error measure. The algorithms are the modulus constraint* □*, modulus constraints with refinement* ∘*, absolute conic* ∇*, Kruppa equations* △*. As a reference the error is also given for the projective reconstruction* ⋆*.*

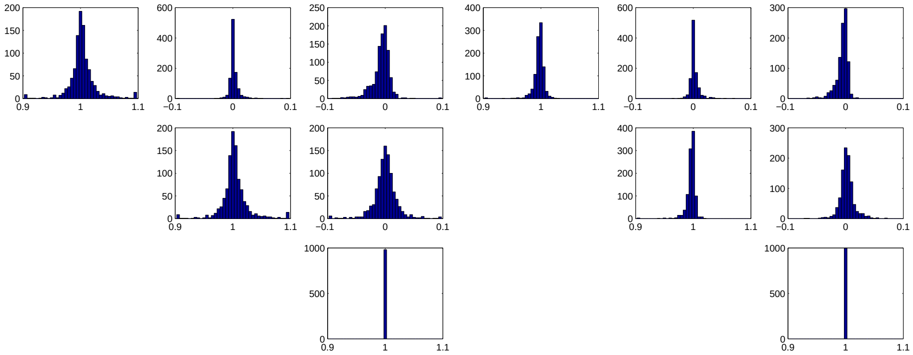

Figure 5.3: *Computed intrinsic camera parameters for 1000 sequences of 12 views (0.5 pixels noise). The results were obtained with the modulus constraint (left) and after refinement (right).*

views is small. This also explains the large 3D error for short image sequences.

The algorithm based on the modulus constraint gives good results, especially when the refinement procedure is carried out. Immediately minimizing the criterion of equation (4.25) on a prior guess of the intrinsic parameters gives similar results except that this method fails more frequently. The results for the Kruppa equations are poor. This algorithm fails often and the error is much higher than for the other algorithms.

These results tend to indicate that self-calibration results are more dependent on a good estimation of the plane at infinity than on the image of the absolute conic (i.e. the intrinsic camera parameters). In fact even when the modulus constraint itself does not converge to a satisfying calibration, the results are often good enough to successfully initialize the refinement procedure. On the other hand starting from a prior guess the minimization sometimes even fails in the absence of noise, indicating that it was started in the attraction basin of a local minimum. This is an argument in favor of a stratified approach to self-calibration. In addition the modulus constraint does not require any prior guess for the intrinsic camera parameters.

The Kruppa equations are similar to the equations used for refinement except that the position of the plane at infinity is eliminated from it. In the presence of noise this implies that for every image pair a different plane at infinity is implicitly used. All the other algorithms which are used here explicitly impose one consistent plane at infinity. An advantage of the Kruppa equations is that only the fundamental matrices are needed (i.e. pairwise calibration) and not a consistent projective frame over all the views. This also explains the observation of Luong [86] that the calibration results obtained with the Kruppa equations improved once the explicit camera poses were computed, since in this step a consistent frame was introduced.

The success of the stratified approach proposed here is consistent with the very good calibration results that were obtained when the plane at infinity or at least its homographies could be identified a priori (e.g. through pure rotation [55] or pure translation [100, 2]).

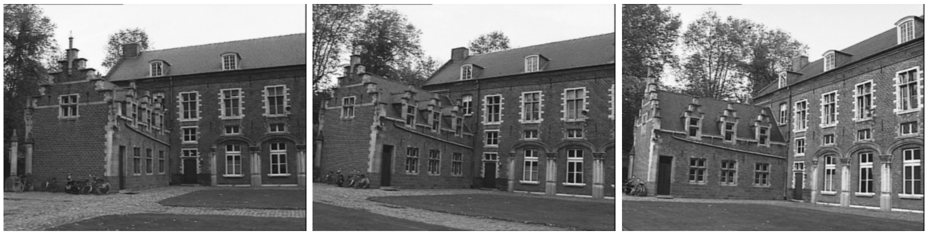

Figure 5.4: *Images of the Arenberg castle which were used to generate the 3D model shown in* Figures 5.5, 5.6, 5.7. *More images of this sequence are shown in Figure 7.2*

5.3.5 Experiments

The stratified approach to the classical self-calibration problem that was presented here has been validated on several real video sequences. Here the results obtained on two sequences of the Arenberg castle are shown. The calibration can be evaluated qualitatively by looking at the reconstruction. A more quantitative evaluation is also performed by comparing the angle between parallel and orthogonal lines. Different parts of the Arenberg castle in Leuven were filmed. These were recorded with a video camera.

Castle sequence

Some images of the first sequence are shown in Figure 5.4. The approach which was followed to generate the 3D model can briefly be summarized as follows. First the corner matching and the projective camera matrices were obtained following the method described in [9]. These camera matrices were upgraded to metric using the self-calibration method described in this text and then a 3D model was generated using these cameras and a dense correspondence map obtained as in [72]. This approach is described more in detail in Chapter 7. Remember that this was obtained without any calibration information. The position of the camera for the different views was also unknown and no knowledge about the scene was given to the algorithm. In Figure 5.5 one can see 3 orthographic views of the reconstructed scene. Parallelism and orthogonality relations clearly have been retrieved. Look for example at the right angles in the top view or at the rectangular windows. Figure 5.6 and Figure 5.7 contain some perspective views of the reconstruction. Because at this point a dense correspondence map was only calculated for two images there are some inaccuracies left in the reconstruction. This also explains the fact that only a partial model is given here. A quantitative assessment of these properties can be made by explicitly measuring angles between lines on the object surface. For this experiment lines were manually selected that are aligned with windows and other prominent surface features. Several lines were identified for three orthogonal directions, see Figure 5.8. The lines along the same direction should be parallel to each other (angle between them should be 0 degrees), while the lines corresponding to different directions should be perpendicular to each other (angle between them should be 90 degrees). The measurement on the

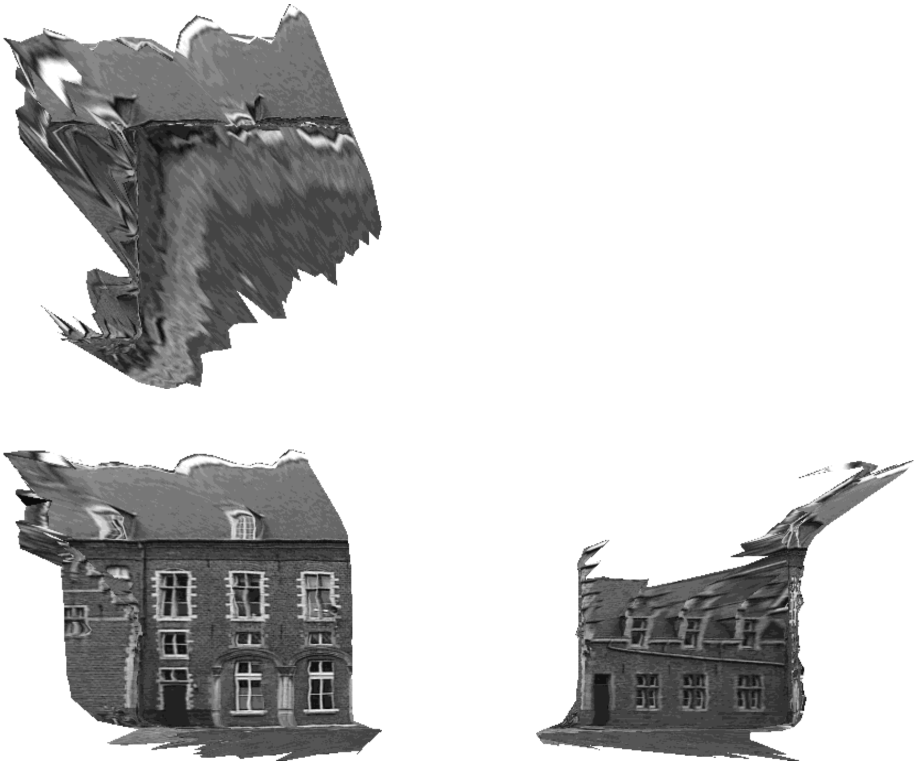

Figure 5.5: *Orthographic views of the reconstruction. Notice parallelism and orthogonality.*

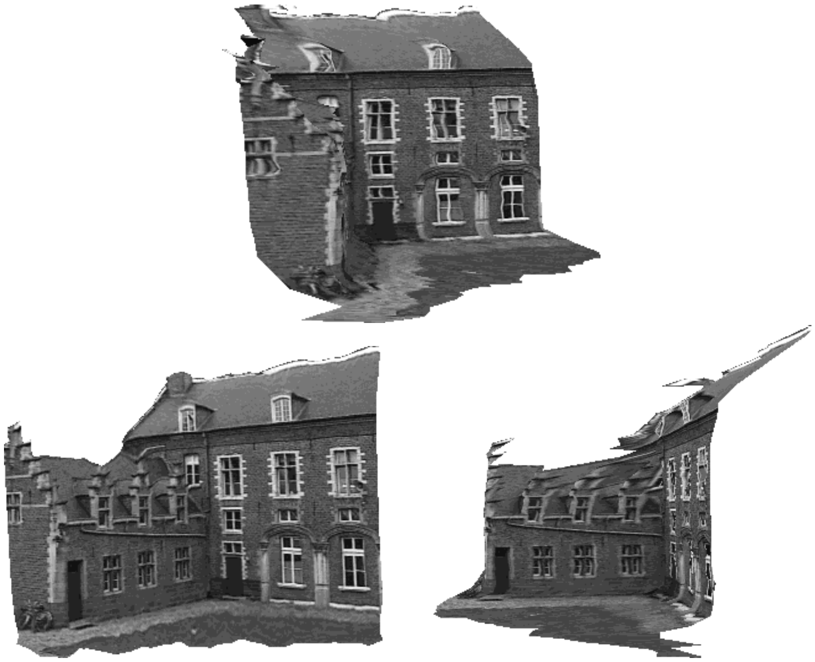

Figure 5.6: *Perspective views of the 3D reconstruction obtained from the sequence seen in* Figure 5.4.

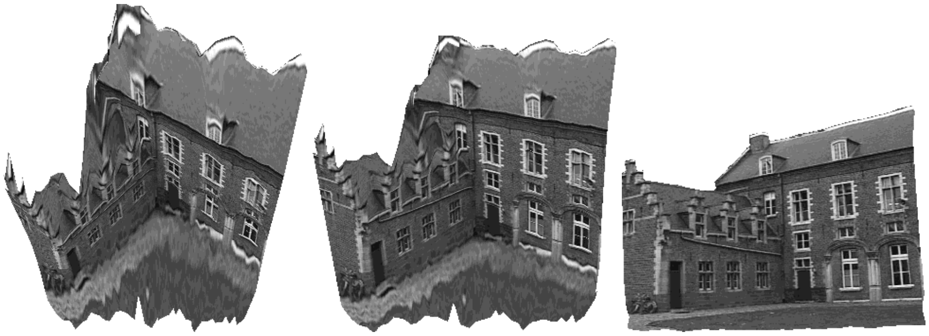

Figure 5.7: *More perspective views of the 3D reconstruction. Notice that once a 3D reconstruction is available it is easy to generate views which where not present in the original sequence.*

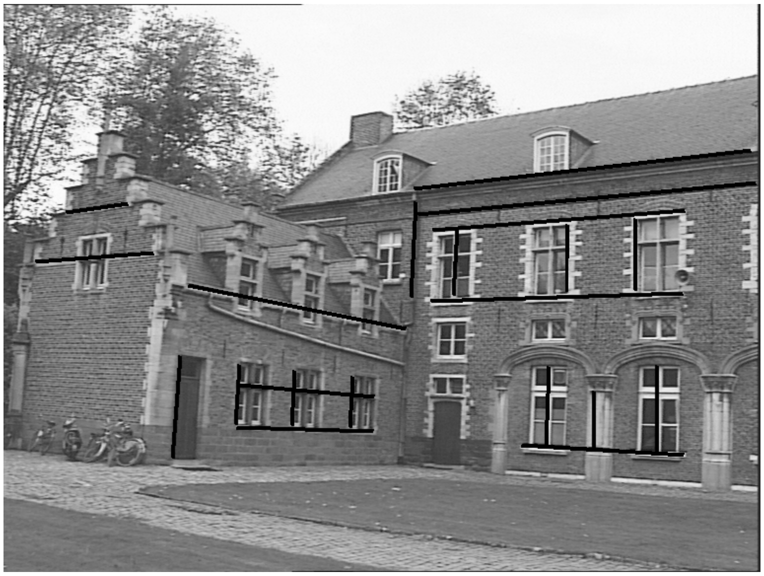

Figure 5.8: *Lines used to verify the parallelism and orthogonality in the reconstructed scene.*

	angle (±std.dev.)
parallel lines	1.8 ± 1.1 degrees
orthogonal lines	89.7 ± 1.4 degrees

Table 5.1: *Results of metric measurements on the reconstruction*

object surface shows that this is indeed close to the expected values (see Table 5.1). Note that the bias for the parallel line comes from the fact that all measured angles are positive.

Some additional experiments were carried out to verify the dependency of the approach on the initialization procedure. As described in Section 5.3.1, the algorithm is initialized with the results obtained from a triplet of views.

For triplets of this sequence the continuation method typically yielded around 5 possible solutions which correspond to real intrinsic parameters. In about half the cases the most probable result (see Section 5.3.1) was already very good in itself (see Table 5.2). In almost all the cases the results obtained from one specific triplet were good enough to succesfully initialize the minimization over all the constraints. For the initializations obtained from 100 arbitrary triplets the algorithm converged to the global minimum in 97 cases (see Table 5.2). Note that these results are good if one takes into account that this sequence is close to critical (see Figure 7.9).

$\tilde{\mathbf{K}} = \begin{bmatrix} 954 & 4 & 356 \\ & 1013 & 329 \\ & & 1 \end{bmatrix}$	$\tilde{\mathbf{K}}_{1,2,3} = \begin{bmatrix} 933 & 10 & 352 \\ & 989 & 244 \\ & & 1 \end{bmatrix}$

Table 5.2: *Estimated intrinsic camera parameters over the whole sequence (left) and for triplet 1,2,3 (right)*

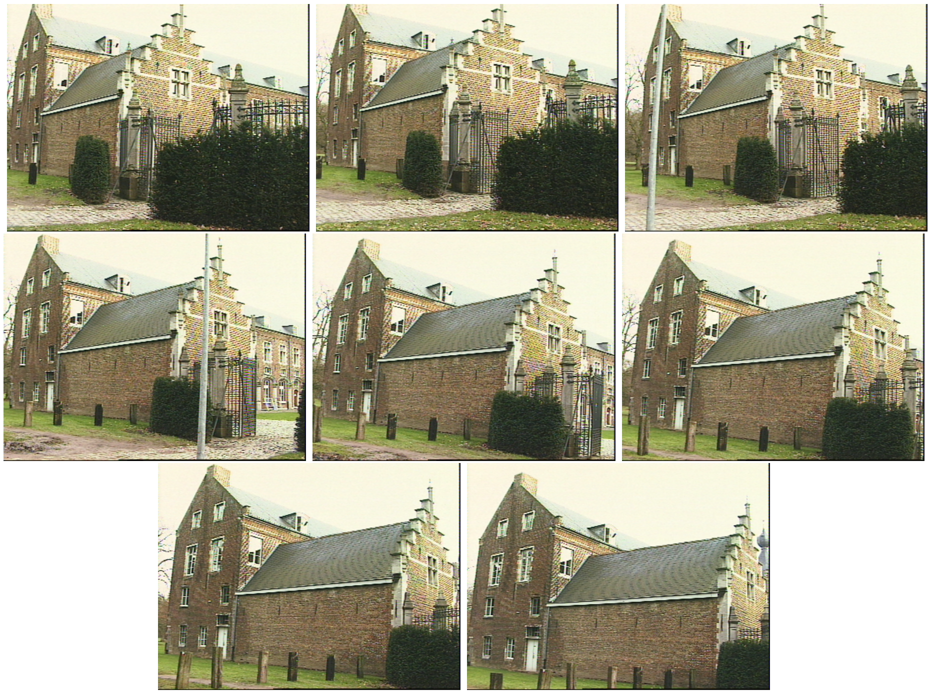

Figure 5.9: *Images of another part of the Arenberg castle. Every second image of the sequence is shown. Views from the 3D model generated from this sequence can be seen in* Figure 5.10.

Castle sequence 2

In Figure 5.9 a part of another sequence is shown. This was filmed at the back of the castle. Here also the method proposed in this text was able to extract a metric reconstruction of the scene. In Figure 5.10 some perspective views of the reconstruction are shown. To illustrate the metric quality of the reconstruction orthographic views of the reconstruction were computed. These are shown in Figure 5.11. Note that these views are extreme views compared to the viewpoints of the sequence. Since dense reconstruction is not the main issue in this chapter (i.e. it is only used to illustrate the

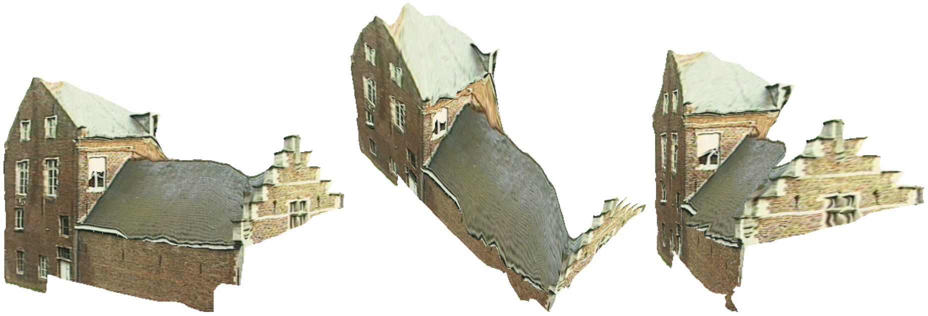

Figure 5.10: *Perspective views of the 3D reconstruction.*

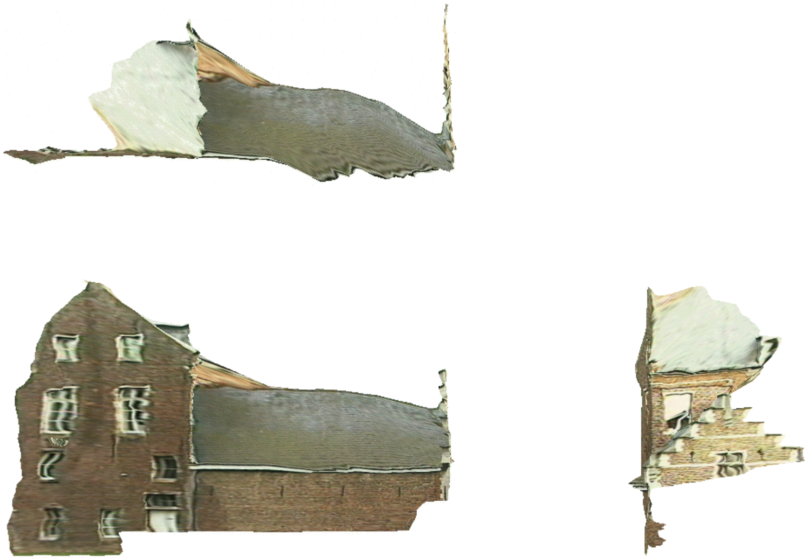

Figure 5.11: *Orthographic views of the 3D reconstruction.*

metric calibration results), only a simple dense correspondence algorithm was used. The dent at the lower part of the roof is due to the alignment of epipolar lines with the gutter which causes an ambiguity for the matcher. Since the left part of the building is only visible in the distance the accuracy of that part of the facade is low. Although some artifacts of the dense reconstruction become clearly visible here, the overall metric qualities are clearly recovered by our reconstruction scheme. Parallelism and orthogonality can be observed from the different orthographic views.

5.4 Special cases

The modulus constraint turns out to be particularly suited for some more specific self-calibration problems. Here two of these cases are discussed.

5.4.1 Two images and two vanishing points

For a pair of images only one modulus constraint exists which is not enough to locate uniquely the plane at infinity. On the other hand, two vanishing points are also insufficient to obtain the plane at infinity. Combining both constraints yields enough information to obtain the affine calibration.

Vanishing points

Some scenes contain parallel lines. These result in vanishing points in the images. Techniques have been proposed to automatically detect such points [90, 172]. To identify the plane at infinity three vanishing points are necessary. By using the modulus constraint this number can be reduced by one. This reduction can be crucial in practice. For example in the castle sequence (see figure 5.4) two vanishing points could be extracted automatically in all frames using the algorithm described in [172]. The third point could only be found in some images at the end of the sequence. This is typical for a lot of scenes where one vanishing point is extracted for horizontal lines and one for vertical lines. Even when three candidate vanishing points are identified, the modulus constraint can still be very useful by providing a means to check the hypothesis.

When a vanishing point $\mathtt{v}$ is identified in the two images it can be used as follows to constrain the homography of the plane at infinity:

$$\mathtt{v}_2 \sim \left[\mathbf{H}_{12}^{\mathtt{REF}} - \mathtt{e}_{12}\pi_\infty^\top\right]\mathtt{v}_1 \tag{5.6}$$

This results in one linear equation for the coefficients of π_∞ (from the three equations only two are independent due to the epipolar correspondence of $\mathtt{v}_1$ and $\mathtt{v}_2$ and one is needed to eliminate the unknown scale factor).

With two known vanishing points we are thus left with a one parameter family of solutions for $\mathbf{H}_{12}^\infty$:

$$\mathbf{H}_{12}^\infty = \mathbf{H}_{12}^{\mathtt{REF}} - \mathbf{e}_{12}(\lambda\pi_a^\top + \pi_b^\top) \tag{5.7}$$

with π_a and π_b being vectors describing the nullspace of the linear equation.

Using the modulus constraint

Applying the modulus constraint is much easier than in the general case. The coefficients l_3, l_2, l_1, l_0 (see equation (5.2)) can be evaluated for both π_a and π_b. The modulus constraint in the two view case then takes on the following form:

$$\begin{aligned}&(\lambda l_3(\pi_a) + l_3(\pi_b))(\lambda l_1(\pi_a) + l_1(\pi_b))^3 \\ &= (\lambda l_2(\pi_a) + l_2(\pi_b))^3(\lambda l_0(\pi_a) + l_0(\pi_b)) \ .\end{aligned} \tag{5.8}$$

This results in a polynomial of degree 4 in only one variable λ (not degree 6 as Sturm anticipated [151]). Therefore at most 4 solutions are possible. Because equation (5.8) is only a necessary condition for $\mathbf{H}_{12}^{\infty}$ to be conjugated with a scaled rotation matrix, this property should be checked out. This can eliminate several solutions. If different solutions persist at this stage some can still be eliminated in the metric calibration part.

Metric calibration

Once the affine calibration is known equation (4.5) can be used as described in 4.4.1. This will however not yield a unique solution. If $\mathtt{V}_R$ is the intersection point of the plane at infinity with the rotation axis to go from one viewpoint to the other, then not only the absolute conic, but also the degenerate conic $\mathtt{V}_R\mathtt{V}_R^\top$ is a fixed conic of the plane at infinity and thus also every linear combination of these. This results in a one parameter family of solutions for the dual absolute conic ω_∞^*.

Additional constraints like some known aspect ratio, orthogonality of the image axes or scene orientations can be used to restrict ω_∞^* to one unique solution. If more than one affine calibration was still under consideration, these constraints can also help out. Also the fact that ω_∞^* should be positive definite and that the principal point should be more or less in the center of the image can be used to find the true affine, and thus also metric, calibration.

The application of these constraints is not so hard. Here the case of orthogonal orientations in the scene will be discussed. This can for example be applied when it is assumed that the extracted vanishing points correspond to orthogonal orientations.

The points $\mathtt{v}_1$ and $\mathtt{v}_1'$ are two vanishing points in the first image. The corresponding scene points can be obtained from the following equation:

$$\mathtt{v}_1 \sim \mathbf{K}[\mathbf{I}|0]\begin{bmatrix}\mathtt{v}\\0\end{bmatrix} \tag{5.9}$$

Thus $\mathtt{v} = \mathbf{K}^{-1}\mathtt{v}_1$ represents the associated direction. Orthogonality now means that $vv' = 0$ or

$$\mathtt{v}_1^\top\mathbf{K}^{-\top}\mathbf{K}^{-1}\mathtt{v}_1' = \mathtt{v}_1^\top\omega\mathtt{v}_1' = 0 \ . \tag{5.10}$$

Therefore it is more appropriate to use the dual equation of (4.5):

$$\omega \sim \mathbf{H}_{12}^{\infty\,\top}\omega\mathbf{H}_{12}^{\infty} \tag{5.11}$$

which of course also yields a one parameter family of solutions for ω. Adding equation (5.10) resolves this ambiguity.

From $\omega = \mathbf{K}^{-\top}\mathbf{K}^{-1}$ first $\mathbf{K}^{-1}$ is extracted by Cholesky factorization and subsequently inverted to obtain $\mathbf{K}$ which contains the camera intrinsic parameters.

Simulation

Some simulations were performed for the 2 view case with 2 known vanishing points. The same type of simulated data was used as in the general case (see section 5.3.4). A noise level of 1 pixel was chosen for the correspondences. The localization of the

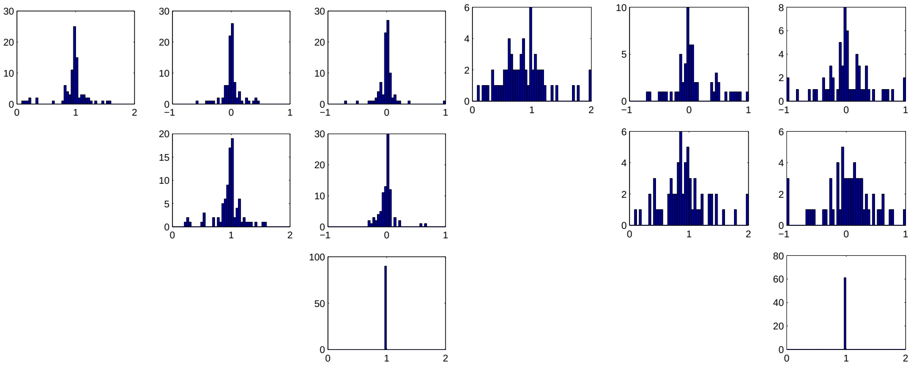

Figure 5.12: *Computed intrinsic camera parameters for low noise level (left) and high noise level (right).*

vanishing points was also perturbed by some noise. First the points at infinity (the 3D points corresponding to vanishing points in the images) were chosen as follows:

$$\mathtt{V}_X = \begin{bmatrix} 1 & 0 & 0 & \nu_X \end{bmatrix}^\top \text{ and } \mathtt{V}_Y = \begin{bmatrix} 0 & 1 & 0 & \nu_Y \end{bmatrix}^\top \tag{5.12}$$

were $\mathtt{V}_X$ and $\mathtt{V}_Y$ are the points at infinity corresponding to the X and Y direction, ν_X and ν_Y is a noise term with a mean of zero and some specific standard deviation σ_1. This means that instead of having a point which is exactly at infinity, mostly a point is obtained of the order of $\frac{1}{\sigma_1}$. The vanishing points used for this simulation are then obtained as follows:

$$\mathtt{v}_{Xi} = \mathbf{P}_i \mathtt{V}_X \text{ and } \mathtt{v}_{Yi} = \mathbf{P}_i \mathtt{V}_Y \quad . \tag{5.13}$$

These coordinates were additionally perturbed by some noise with standard deviation σ_2 pixels.

The modulus constraint and the two vanishing points were used to obtain the homography of the plane at infinity. This homography determines the absolute conic up to one parameter. This parameter was retrieved by imposing orthogonality between the directions corresponding to the vanishing points.

A first set of 100 simulations was carried out with a small amount of noise (i.e. $\sigma_1 = 0.01, \sigma_2 = 10$), a second set of 100 simulations was carried out with a high amount of noise (i.e. $\sigma_1 = 0.1, \sigma_2 = 100$). The results of these simulations can be seen in Figure 5.12. The left part shows the results which were obtained with only a small amount of noise. The layout is similar to Figure 5.3.

It can be seen that the results are good for a small amount of noise (left part of Figure 5.12). With a lot of noise the results seriously deteriorate (see right part of Figure 5.12). Although most solutions are situated around the correct values, a much bigger spread exists. In addition the algorithm fails in 30% of the cases.

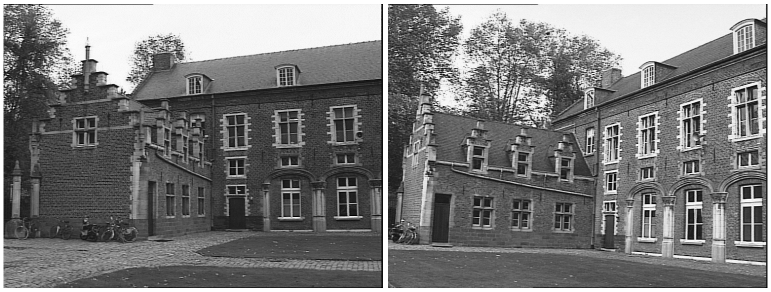

Figure 5.13: *Images 2 and 24 of the castle sequence*

Experiment

It was seen in section 5.4.1 that two images could be sufficient for self-calibration when two vanishing points corresponding to orthogonal directions can be identified in these images.

The first step is to obtain the weak calibration for the two images. This corresponds to identifying the fundamental matrix. From this a valid instance of the projective camera matrices can be chosen. The second step consists of identifying at least two vanishing points in both images. This was done using a cascaded Hough approach. A first Hough transform is used to identify lines in dual space, a second Hough transform is then used to identify coincidence points. In scenes containing regular structures (e.g. buildings) these points are typically vanishing points. More details of this approach can be found in [172].

Combining these results allows us to follow the steps described in section 5.4.1:

- Affine calibration: (1) writing down the modulus constraint for the two images, (2) filling in the one parameter family of solutions for the infinity homography $\mathbf{H}_\infty$ defined by the two vanishing points and the projective calibration, (3) computing the solutions and extrema of the obtained equation, (4) selecting the best one according to equal moduli of eigenvalues.

- Metric calibration: (1) determining the one parameter family for the absolute conic corresponding with the infinity homography, (2) determining the solution for which the vanishing points correspond to orthogonal directions.

Two images of the castle sequence were used in this experiment. These can be seen in Figure 5.13. In addition to the projective calibration two vanishing points were automatically retrieved from these images. The coordinates of the computed vanishing points can be found in Table 5.3. Using the described method the following intrinsic camera parameters were obtained (see Table 5.4). These parameters are relatively close to the parameters obtained with the general method on the whole sequence. Except for the skew which is relatively large (around 6 degrees). These are typical results although for some image pairs less accurate results or no results were obtained. This could be expected from the simulations (see Section 5.4.1) which

im 2 (H)	im 2 (V)	im 24 (H)	im 24 (V)
-2490	-1366	-209	1275
37	-41954	437	-10582

Table 5.3: *Image coordinates of the extracted vanishing points*

$$\mathbf{K} = \begin{bmatrix} 875 & 98 & 381 \\ & 880 & 197 \\ & & 1 \end{bmatrix}$$

Table 5.4: *Result obtained from image 2 and 24*

showed a strong noise sensitivity for this method. For more accurate results a bundle adjustment should be applied on the projective reconstruction and a more accurate localization method should be used for the vanishing points (the main goal of the method that was used [172] was detection and not accurate localization).

In some cases this method wasn't able to obtain any solution. These observations can be explained by the absence of redundancy when one tries to extract the metric calibration from a scene from two images only. The self-calibration problem is known to be a hard problem and therefore as much redundancy as possible should be used. Sometimes even this is not enough since many critical motion sequences exist (see Section 4.4.3).

5.4.2 Varying focal length

The modulus constraint can also be used for other purposes than affine calibration (see [121] for more details). The constraint depends on two conditions: the affine calibration *and* constant intrinsic camera parameters. For each view except the first one we get a valid constraint. This means that instead of "spending" the constraint on solving for affine calibration, one can in the traditional scheme –where affine calibration amounts from translation between the first two views [100]– use the constraint to retrieve one changing parameter for each supplementary view. An alternative approach would be to leave the camera parameters unchanged until the affine calibration is retrieved and only then start varying the camera parameters. The most practical application is to allow the focal length to vary. This allows to cope with auto-focusing and zooming in the image sequence. A similar approach was proposed for stereo rigs [125], this case will be discussed in 5.4.3.

Modeling the change in focal length

The first step is to model the effect of changes in focal length. These changes are relatively well described by scaling the image around a fixed focus of expansion $\mathtt{c}$.

This can be expressed as follows:

$$\mathtt{m}^{f}_{li} = \mathbf{K}_f\, \mathtt{m}_{li} \text{ with } \mathbf{K}_f = \begin{bmatrix} 1 & 0 & (f^{-1}-1)c_x \\ 0 & 1 & (f^{-1}-1)c_y \\ 0 & 0 & f^{-1} \end{bmatrix} \tag{5.14}$$

where $\mathtt{m}_{li}$ are the points that one would have seen if no change in focal length had occurred, and with $\mathtt{m}^{f}{}_{li}$ the image points for some *relative* focal length f. Note that this equation also makes it possible to cancel the effect of the zoom by *de*zooming a projection matrix using $\mathbf{K}_f^{-1}$.

The first thing to do is to retrieve the focus of expansion $\mathtt{c}$. Fortunately, this is easy for a camera with variable focal length, $\mathtt{c}$ being the only finite fixed point when varying the focal length without moving the camera [80, 83]. The affine camera calibration can then be retrieved from two views with a different focal length and a pure translation between the two views, using the method described in [121].

Using the modulus constraint

The modulus constraint is only valid for an affinely calibrated camera with constant intrinsic camera parameters or after that the effect of the change in focal length has been taken away. Stated differently, the modulus constraint must be valid for a camera matrix $\tilde{\mathbf{P}}_A = \mathbf{K}_f^{-1}\mathbf{P}_A$. Writing down the characteristic equation, we get an equation similar to equation (5.2):

$$\det(\mathbf{K}_f^{-1}\mathbf{H}_{1i}^{\infty} - \lambda\mathbf{I}) = l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = 0 \tag{5.15}$$

The expressions for l_1, l_2, l_3, l_4 are worked out in Appendix A.3. Substituting the obtained coefficients in equation (5.3) we obtain a 4^{th} degree polynomial equation in f:

$$\alpha_4 f^4 + \alpha_3 f^3 + \alpha_2 f^2 + \alpha_1 f + \alpha_0 = 0 \tag{5.16}$$

This equation has 4 possible solutions. It can be proven that if f is a real solution, then $-f$ must also be a solution (see appendix A.3). Imposing this to equation (5.16) yields the following result:

$$f = \sqrt{\frac{\alpha_1}{\alpha_3}}\ . \tag{5.17}$$

Now that f has been retrieved, $\mathbf{K}_f^{-1}$ can be used to get normalized images and cameras. These affine camera projection matrices can then be upgraded to metric as described in section 5.3.2.

Simulation

Some simulations were also carried out for this case. Here again, the same type of simulated data was used as in the general case (see section 5.3.4). Different noise levels were used for the simulations. For every experiment four views were generated.

The first two only differ in focal length ($f_1 = 1, f_2 = 2$) which allows us to estimate the focus of expansion. The focal length for the other views is chosen at

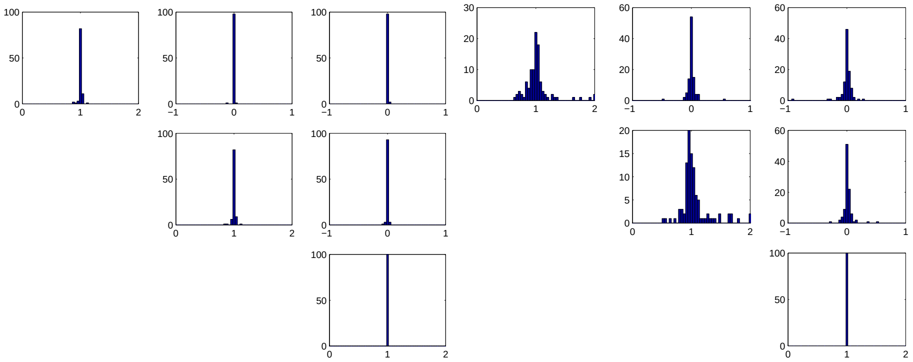

Figure 5.14: *Computed intrinsic camera parameters for low noise level (left) and high noise level (right).*

random between f_1 and f_2. For the third view a pure translation was carried out. From this the affine reconstruction was obtained, in spite of an unknown change in focal length. For the last image a combination of a rotation and a translation was used.

For every noise level hundred experiments were carried out. The results for 0.2 and 2 pixels of noise are shown in Figure 5.14. This corresponds to a low and a high level of noise on the correspondences. The layout is similar to Figure 5.3.

It can be seen that the results are very good for a small amount of noise (left part of Figure 5.12). With more noise the results are still good. The other computed parameters (i.e. the focus of expansion and the relative focal lengths for the different views) are not shown. In general the focus of expansion and the relative focal length for views 2 and 3 are very accurately obtained. This is due to the high redundancy of the equations in these cases. The quality of the estimate for the relative focal length for view 4 is of the same order as the absolute estimate of the focal length (i.e. f_x).

Experiment

In this case an indoor scene was used for our experiment. First the focus of expansion was identified by zooming. Then a pure translation was carried out which allowed to retrieve the affine structure. Finally an additional motion was used to get the metric structure of the scene.

Here some results obtained from a scene consisting of two boxes and a cup are given. The images that were used can be seen in Figure 5.15. The scene was chosen to allow a good qualitative evaluation of the metric reconstruction. The boxes have right angles and the cup is cylindrical. These characteristics must be preserved by a *metric* reconstruction, but will in general not be preserved by an *affine* or *projective* reconstruction.

First a corner matcher was used to extract point correspondences between the three images. From these the method allowing varying focal lengths was used to obtain a metric calibration of the cameras. Subsequently, the algorithm of [134] was used to

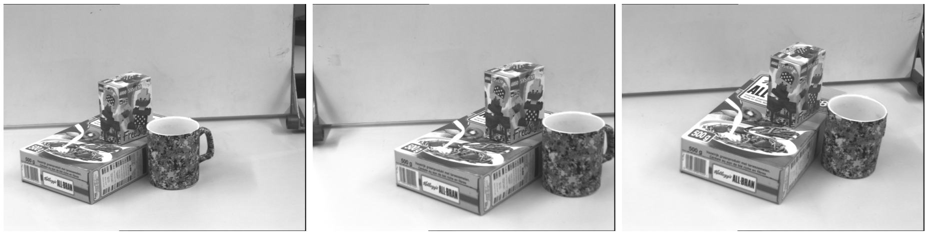

Figure 5.15: *The 3 images that were used to build a Euclidean reconstruction. The camera was translated between the first two views (the zoom was used to keep the size more or less constant). For the third image the camera was also rotated.*

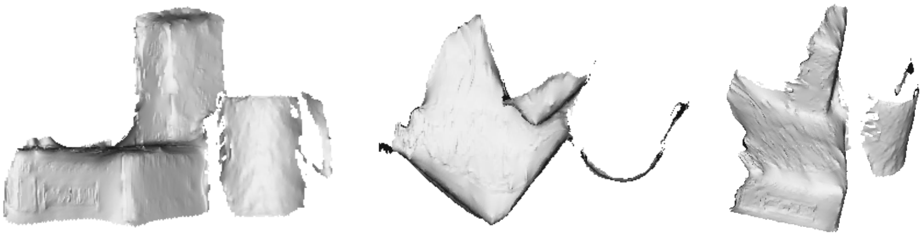

Figure 5.16: *Different views of the 3D reconstruction.*

compute dense point correspondences. These were used to build the final 3D reconstruction using the previously recovered calibration.

Figure 5.16 shows three views of the reconstructed scene. The left image is a front view, the middle image a top view, while the right image is a side view. Note especially from the top view, that 90^o angles are preserved and that the cup keeps its cylindrical form which is an indication of the quality of the metric reconstruction.

5.4.3 Stereo rig with varying focal length

The auto-calibration techniques proposed by Zisserman et al. [188] and Devernay and Faugeras [28] (or more recently by Horaud and Csurka [65]) for two views taken with a rotating fixed stereo rig can also be generalized to allow changes in focal lengths for both cameras independently and purely translational motions. In fact the method is easier than for a single camera. For a fixed stereo rig the epipoles are fixed as long as one doesn't change the focal length. In this case the movement of the epipole in one camera is in direct relation with the change of its focal length. This is illustrated in figure 5.17. Knowing the relative change in focal length and the principal points allows to remove the effect of this change from the images. From then on the standard techniques for fixed stereo rigs can be used [188, 28, 65].

The same idea was recently extended by Shashua [143] to allow all intrinsic camera parameters and orientations of an omnirig (rig with 5 cameras in which the centers of projection are fixed) to vary. With common points between two views of this omnirig it is possible to self-calibrate it. Without common points it is still possible to obtain a metric frame if the rig had been calibrated before, even if all parameters changed in the meanwhile. The 5 projection centers are used to transform the projective reconstruction to a metric one.

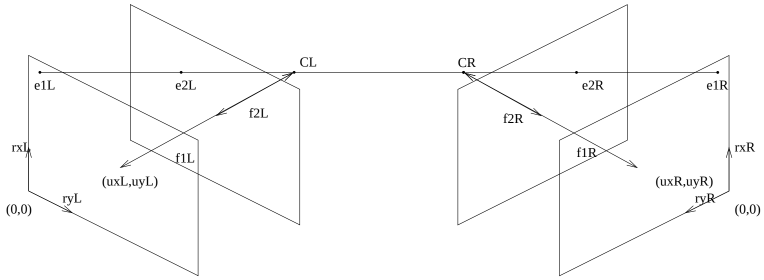

Figure 5.17: *this figure illustrates how the epipoles will move in function of a change in focal length.*

By first extracting the principal points -i.e. mildly calibrating the camera - one can then also get a Euclidean reconstruction even for a translating stereo rig, which was not possible with earlier methods [188, 28]. Between any pair of cameras i and j we have the following constraints:

$$\lambda_{ij}\omega_j^* = \mathbf{H}_{ij}^{\infty}\omega_i^*\mathbf{H}_{ij}^{\infty\top} \tag{5.18}$$

For two views with a fixed stereo rig there are 3 different constraints of the type of equation (5.18): for the left camera (between view 1 and 2), for the right camera (between view 1 and 2) and between the left and the right camera. For a translation $\mathbf{H}_{12\infty} = \mathbf{I}$ which means that the two first constraints become trivial. The constraint between the left and the right camera in general gives 5 independent equations[2]. This is not enough to solve for $f_x^L, f_y^L, s^L, c_x^L, c_y^L, f_x^R, f_y^R, s^R, c_x^R, c_y^R$. Assuming that the principal point coincides with the focus of expansion and assuming perpendicular image axes [188] restricts the number of unknowns to 4. This problem can then be solved through a system of linear equations (see [126]). In practical cases this can be important because with earlier techniques any motion of the stereo rig which was close to translation gave unstable results.

It is also useful to note that in the case of a translational motion of the rig, the epipolar geometry can be obtained with as few as 3 points seen in all 4 views. After cancelling the change in focal length and superimposing the translated views, it is as if one observes two translated copies of the points. Choosing two of the three points, one obtains four coplanar points from the two copies (coplanarity derives from the fact that the rig translates). Together with projections of the third point, this suffices to apply the algorithm propounded in [13]. Needing as few as 3 points clearly is advantageous to detect e.g. independent motions using RANSAC strategies [162].

---

[2] the cameras of the stereo rig should not have the same orientation.

Experiments

The algorithm described in the previous section, was applied to synthetic images as well as real images. From tests with synthetic data one can conclude that restricting the motion to translation gives more stable results. For a report on these results, we refer to [131].

Some results obtained from a real scene are presented here. The scene consists of a box and a cylindrical object on a textured background. Images were acquired with a translating stereo rig. They can be seen in Figure 5.18. Figure 5.19 shows the reconstruction results. Notice that angles are well preserved (e.g. the top and the front view differ by 90^o, the box and the floor have right angles in the reconstruction). The inaccuracies in the reconstruction (like the dent in the cylindrical object) are mainly due to the rendering process which uses triangulation between matched points and are not related to the accuracy of the calibration.

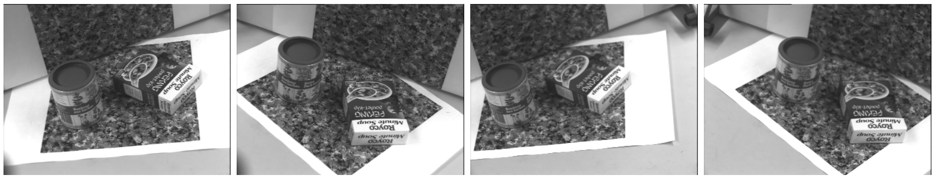

Figure 5.18: *Two pairs of images of a scene taken with a translating stereo rig.*

5.5 Conclusion

In this chapter the modulus constraint and its application to self-calibration problems was discussed. It has been shown that this versatile constraint can be used to obtain the self-calibration of cameras in different circumstances. First the classical self-calibration problem was solved. The modulus constraint allows to follow a stratified approach. The projective calibration is upgraded to affine by identifying the plane

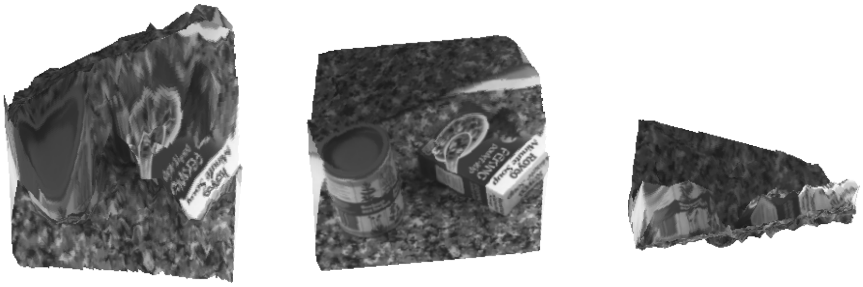

Figure 5.19: *Different views of the 3D reconstruction of the scene.*

at infinity, then the absolute conic is located and a metric calibration is obtained. A non-linear minimization step allows to refine this calibration.

The experiments on synthetic data show that a stratified approach can be more successful in obtaining a satisfying calibration than an approach based on an a priori guess for the intrinsic camera parameters. These experiments confirm the importance of a good localization of the plane at infinity for a successful calibration. Besides experiments on synthetic data some real image sequences were used to illustrate the feasibility of the method.

Some other applications of the modulus constraint were also proposed. In many circumstances two or more vanishing points are known or can be found in the images. In this case the modulus constraint can even result in self-calibration from two images only. Interesting results have also been obtained for a varying focal length. Once the affine calibration has been obtained (i.e. from a pure translation) the modulus constraint can be used to retrieve the focal length through a closed form equation. A similar approach for stereo rigs was also briefly discussed.

Chapter 6

Flexible self-calibration

6.1 Introduction

In recent years, researchers have been studying self-calibration methods for cameras. Mostly completely unknown but constant intrinsic camera parameters were assumed. Several of these techniques were discussed in the two previous chapters. This has the disadvantage that the zoom can not be used and even focusing is prohibited. On the other hand, the proposed perspective model is often too general compared to the range of existing cameras. Mostly the image axes can be assumed orthogonal and often the aspect ratio is known. Therefore a tradeoff can be made and by assuming these parameters to be known, one can allow (some of) the other parameters to vary throughout the image sequence.

Since it became clear that projective reconstructions could be obtained from image sequences alone [36, 51], researchers tried to find ways to upgrade these reconstructions to metric (i.e. Euclidean up to unknown scale). Many methods were developed which assumed constant intrinsic camera parameters [37, 87, 183, 60, 166, 124, 123, 53]. Most of these methods are based on the absolute conic which is the only conic which stays fixed under all Euclidean transformations [142]. This conic lays in the plane at infinity and its image is directly related to the intrinsic camera parameters, hence the advantage for self-calibration.

So far not much work had been done on varying intrinsic camera parameters. In the previous chapter (see also [121]) a stratified approach for the case of a varying focal length was proposed, but this method required a pure translation as initialization, along the lines of Armstrong et al.'s [2] earlier account for fixed intrinsic camera parameters. Recently Heyden and Åström [61] proved that self-calibration was possible when the pixels could be assumed to be squares. The self-calibration method proposed in their paper is based on bundle adjustment which requires non-linear minimization over all reconstructed points and cameras simultaneously. No method was proposed to obtain a suitable initialization.

In this chapter this proof is extended. It will be shown that the knowledge that the pixels are rectangular is sufficient to allow self-calibration. Through geometric reasoning this result will even be extended to *any* intrinsic camera parameter. A versatile self-calibration method is proposed which can deal with varying types of constraints. This will then be specialized towards the case where the focal length varies, possibly also the principal point.

Section 6.2 discusses some theoretical aspects of self-calibration for varying camera parameters. Section 6.4 discusses the problem of critical motion sequences in the case of varying intrinsic parameters. Section 6.5 discusses some aspects of the automatic selection of constraints for self-calibration. The method is then validated through the experiments of Section 6.6. Section 6.7 concludes this chapter.

6.2 Some theory

Before developing a practical self-calibration algorithm for varying intrinsic parameters some theoretical aspects of the problem are studied in this section. First a counting argument is given which states the minimal sequence length that allows self-calibration from a specific set of constraints. Then a theorem is given which states that self-calibration is possible for the minimal case where the only available constraint is the absence of skew. Based on the geometric interpretation of some calibration constraints, this is generalized to any known intrinsic parameter.

6.2.1 A counting argument

To restrict the projective ambiguity (15 degrees of freedom) to a metric one (3 degrees of freedom for rotation, 3 for translation and 1 for scale), at least 8 constraints are needed. This thus determines the minimum length of a sequence from which self-calibration can be obtained, depending on the type of constraints which are available for each view. Knowing an intrinsic camera parameter for n views gives n constraints, fixing one yields only $n-1$ constraints.

$$n \times (\#known) + (n-1) \times (\#fixed) \geq 8$$

Of course this counting argument is only valid if all the constraints are independent. In this context critical motion sequences are of special importance (see Section 6.4 and esp. 6.4.2).

Therefore the absence of skew (1 constraint per view) should in general be enough to allow self-calibration on a sequence of 8 or more images. In Section 6.2.3 it will be shown that this simple constraint is not bound to be degenerate. If in addition the aspect ratio is known (e.g. $f_x = f_y$) then 4 views should be sufficient. When also the principal point is known, a pair of images is enough. A few more examples are given in Table 6.2.1.

6.2.2 A geometric interpretation of self-calibration constraints

In this section a geometric interpretation of a camera projection matrix is given. It is seen that constraints on the internal camera parameters can easily be given a geometric

constraints	known	fixed	min #images
no skew	s		8
fixed aspect ratio and no skew	s	$\frac{f_y}{f_x}$	5
known aspect ratio and no skew	s, $\frac{f_y}{f_x}$		4
only focal length is unknown	s, $\frac{f_y}{f_x}$, c_x, c_y		2
standard self-calibration problem		f_x, f_y, c_x, c_y, s	3

Table 6.1: *A few examples of minimum sequence length required to allow self-calibration*

interpretation in space. This will then be used in the next section to generalize the theorem given in [120] (and Appendix B).

A camera projection plane defines a set of three planes. The first one is parallel to the image and goes through the center of projection. This plane can be obtained by back-projecting the line at infinity of the image (i.e. $[001]^\top$). The two others respectively correspond to the back-projection of the image x- and y-axis (i.e. $[010]^\top$ and $[100]^\top$ resp.). A line can be back-projected through equation (3.9):

$$\Pi \sim \mathbf{P}^\top \mathtt{l} \sim \begin{bmatrix} \mathbf{R} \\ \text{-}\mathtt{t}^\top \mathbf{R} \end{bmatrix} \mathbf{K}^\top \mathtt{l} \tag{6.1}$$

Let us look at the relative orientation of these planes. Therefore the rotation and translation can be left out without loss of generality (i.e. a camera centered representation is used). Let us then define the vectors $\mathtt{v}_x$, $\mathtt{v}_y$ and $\mathtt{v}_i$ as the three first coefficients of these planes. This then yields the following three vectors:

$$\mathtt{v}_x = \begin{bmatrix} 0 \\ f_y \\ c_y \end{bmatrix}, \mathtt{v}_y = \begin{bmatrix} f_x \\ s \\ c_x \end{bmatrix}, \mathtt{v}_i = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{6.2}$$

The vectors coinciding with the direction of the x and the y axis can be obtained by intersections of these planes:

$$\mathtt{l}_x = \mathtt{v}_x \times \mathtt{v}_i = \begin{bmatrix} f_y \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathtt{l}_y = \mathtt{v}_y \times \mathtt{v}_i = \begin{bmatrix} s \\ -f_x \\ 0 \end{bmatrix} . \tag{6.3}$$

The following dot products can now be taken:

$$\mathtt{l}_x.\mathtt{l}_y = s f_y \text{ , } \mathtt{v}_x.\mathtt{v}_i = c_y \text{ and } \mathtt{v}_y.\mathtt{v}_i = c_x \tag{6.4}$$

Equation (6.4) proves that the constraint for rectangular pixels (i.e. $s = 0$), and zero coordinates for the principal point (i.e. $c_x = 0$ and $c_y = 0$) can all be expressed in terms of orthogonality between vectors in space. Note further that it is possible to pre-warp the image so that a known skew[1] or known principal point parameters coincide with zero. Similarly a known focal length or aspect ratio can be scaled to one.

---

[1] In this case the skew should be given as an angle in the image plane. If the aspect ratio is also known, this corresponds to an angle in the retinal plane (e.g. CCD-array).

It is also possible to give a geometric interpretation to a known focal length or aspect ratio. Put a plane parallel with the image at distance d from the center of projection (i.e. $Z = d$ in camera centered coordinates). In this case a horizontal motion in the image of f_x pixels will move the intersection point of the line of sight over a distance d. In other words a known focal length is equivalent to knowing that the length of two (typically orthogonal) vectors are equal. If the aspect ratio is defined as the ratio between the horizontal and vertical sides of a pixel (which makes it independent of s), a similar interpretation is possible.

6.2.3 Minimal constraints for self-calibration

In this paragraph it is stated that the knowledge that the pixels are rectangular can be sufficient to yield a metric reconstruction (the actual proof is given in Appendix B). This is an extension of the theorem proposed by Heyden and Åström [61] which besides orthogonality also requires the aspect ratio to be known (i.e. square pixels). In the meanwhile Heyden and Åström [64] independently proposed a generalization of their theorem.

Theorem 6.1 *The class of transformations which preserves rectangular pixels is the group of similarity transformations.*

The original proof is given in appendix B. Here we give a geometric proof based on the insights of the previous section. These insights allow us to generalize Theorem 6.1 for any known intrinsic parameter. The proof for rectangular pixels (no skew) or known coordinates of the principal point is given in Figure 6.1. While the proof for known focal length or known aspect ratio is given in Figure 6.2. In these figures a pair of lines represents a specific camera pose. Depending on the figure orthogonality or equal length constraints should be satisfied for every pair of lines. If a construction can be obtained which fixes all parameters of projective distortion (i.e. restrict the ambiguity to a similarity) we dispose of an example of a non-critical motion sequence. This shows that the constraint is not bound to be degenerate and thus that in general the constraints are independent.

If a sequence is general enough (in its motion) it follows from Theorem 6.1 that only a projective representation of the cameras which can be related to the original ones through a similarity transformation (possibly including a mirroring) would satisfy the orthogonality of rows and columns for all views. Using oriented projective geometry [78] the mirroring ambiguity can easily be eliminated. The geometric proofs given in Figure 6.1 and Figure 6.2 generalize this towards any known intrinsic camera parameter. Therefore self-calibration and metric reconstruction are possible from minimal constraints supposing the motion is general enough. Of course more restricted camera models yield more accurate results and diminish the probability of encountering critical motion sequences (see Section 6.4).

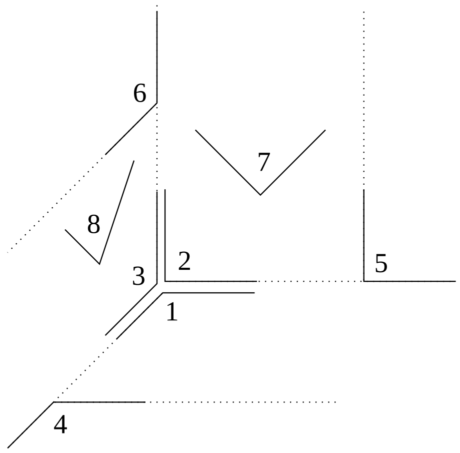

Figure 6.1: *geometric proof for skew or principal point: Every pair of lines represent an orthogonality constraint. Constraints 1,2 and 3 define a set of mutually orthogonal lines. The pairs of constraints 1-4, 2-5 and 3-6 define the points at infinity of these lines and thus fix the plane at infinity. At this point only a scaling along one of the three axes is possible. This is made impossible through the constraints 7 and 8. Only a global scale and a rigid transformation of the whole construction are still possible.* QED.

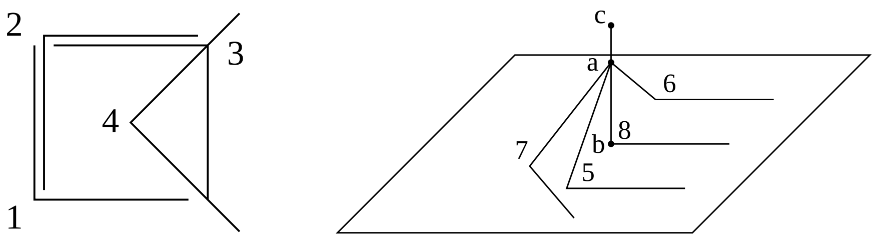

Figure 6.2: *Geometric proof for focal length or aspect ratio: In this case every pair of lines represent an equal length constraint. Starting in a plane, constraints 1,2 and 3 give us a rhombus. Constraint 4 imposes equal diagonals, which thus results in a square. At this stage the ambiguity on the plane is restricted to a 2D similarity. Putting up the construction 5-6-7 we obtain a first fixed point out of the plane. Constraint 8 is positioned so that the line* `bc` *passes through* `a` *and is perpendicular to the plane (the basepoint can easily be computed from the positions of 5,6 and 7). The known scale factor* $\frac{|\mathtt{ab}|}{|\mathtt{cb}|}$ *allows to obtain the point at infinity of the line. Together with the line at infinity of the plane this fully determines the affine structure of the scene. Since* `bc` *was constructed perpendicular to the plane (and the relative scale to the plane is also fixed) the ambiguity is restricted to a similarity.* QED.

6.3 Self-calibration of a camera with varying intrinsic parameters

It is a well-known result that from image correspondences alone the camera projection matrices and the reconstruction of the scene points can be retrieved up to a projective transformation without assuming any knowledge about the intrinsic camera parameters [36, 51]. A practical method is described in Section 7. Note that without additional constraints nothing more can be achieved. This was shown in Section 4.2.

In general, however, some additional constraints are available. Some intrinsic parameters are known or can be assumed constant. This yields constraints which should be verified when $\mathbf{P}$ is factorized as in equation (3.7).

It was shown in the previous section that when no skew is present (or when another intrinsic camera parameter is known), the ambiguity of the reconstruction can be restricted to metric. Although this is theoretically sufficient, under practical circumstances often much more constraints are available and should be used.

In Euclidean space two entities are invariant –setwise, not pointwise– under rigid transformations. The first one is the plane at infinity Π_∞ which allows to obtain affine measurements. The second entity is the absolute conic Ω which is embedded in the plane at infinity. If besides the plane at infinity Π_∞ the absolute conic Ω has also been localized, metric measurements are possible.

When looking at a static scene from different viewpoints the relative position of the camera towards Π_∞ and Ω is invariant. If the motion is general enough, only one conic in one specific plane will satisfy this condition. The absolute conic can therefore be used as a virtual calibration pattern which is always present in the scene.

A practical way to encode both the absolute conic Ω and the plane at infinity Π_∞ is through the use of the absolute dual quadric Ω^* [142] (introduced in computer vision by Triggs [166]). It was introduced in Section 2.3.3 and its application in the context of self-calibration was discussed in Section 4.4.1. Its projection in the images is given by equation (4.3):

$$\omega_i^* = \mathbf{K}_i\mathbf{K}_i^\top \sim \mathbf{P}_i\Omega^*\mathbf{P}_i^\top \tag{6.5}$$

This equation can also be seen as describing the back-projection of the dual image absolute conic $\omega_i^* = \mathbf{K}_i\mathbf{K}_i^\top$ into 3D space. Constraints on the intrinsic camera parameters in $\mathbf{K}_i$ can therefore be translated to constraints on the absolute dual quadric. If enough constraints are at hand only one quadric will satisfy them all, i.e. the absolute dual quadric. At that point the scene can be transformed to the metric frame (which brings Ω^* to its canonical form). Some of these concepts are illustrated in Figure 6.3.

6.3.1 Non-linear approach

Equation (4.3) can be used to obtain the metric calibration from the projective one. The dual image absolute conics ω_i^* should be parameterized in such a way that they enforce the constraints on the calibration parameters. For the absolute dual quadric Ω^* a minimum parameterization (8 parameters) should be used. This can be done by

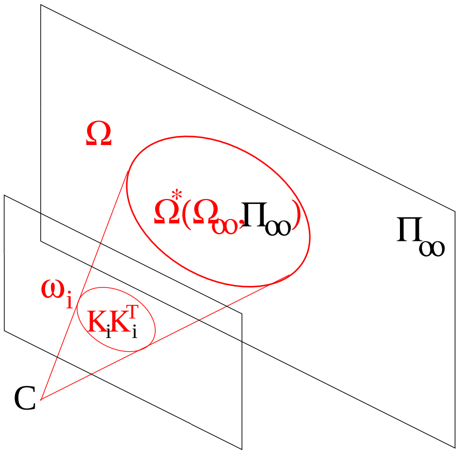

Figure 6.3: *The absolute dual quadric* Ω^* *which encodes both the plane at infinity* Π_∞ *(affine reference entity) and the absolute conic* $\mathbf{\Omega}$ *(metric reference entity), projects to the dual image of the absolute conic* $\omega_i = \mathbf{K}_i\mathbf{K}_i^\top$. *The projection equation allows to translate constraints on the intrinsic parameters to constraints on* Ω^*.

putting $\Omega^*_{33} = 1$ and by calculating Ω^*_{44} from the rank 3 constraint. The following parameterization satisfies these requirements:

$$\Omega^* = \left[\begin{array}{cc} \mathbf{K}\mathbf{K}^\top & -\mathbf{K}\mathbf{K}^\top\pi_\infty \\ -\pi_\infty^\top\mathbf{K}\mathbf{K}^\top & \pi_\infty^\top\mathbf{K}\mathbf{K}^\top\pi_\infty \end{array}\right] . \tag{6.6}$$

Here π_∞ defines the position of the plane at infinity $\Pi_\infty = [\pi_\infty^\top \ 1]^\top$. In this case the transformation from projective to metric is particularly simple:

$$\mathbf{T}_{PM} = \left[\begin{array}{cc} \mathbf{K}^{-1} & 0_3 \\ \pi_\infty^\top & 1 \end{array}\right] \tag{6.7}$$

An approximate solution to these equations can be obtained through non-linear least squares. The following criterion should be minimized (with $\mathbf{F}(\mathbf{A}) \equiv \frac{\mathbf{A}}{\|\mathbf{A}\|_F}$):

$$\mathcal{C}_F(\mathbf{K}_i, \Omega^*) = \mathcal{C}_F(\mathbf{K}_i, \mathbf{K}, \pi_\infty) = \sum_{i=1}^{n} \left\| \mathbf{F}(\mathbf{K}_i\mathbf{K}_i^\top) - \mathbf{F}(\mathbf{P}_i\Omega^*\mathbf{P}_i^\top) \right\|_F \quad . \tag{6.8}$$

If one chooses $\mathbf{P}_1 = [\mathbf{I}|\mathbf{0}]$, equation (4.3) can be rewritten as follows:

$$\mathbf{K}_i\mathbf{K}_i^\top \sim \mathbf{P}_i\Omega_1^*\mathbf{P}_i^\top \text{ with } \Omega_1^* \equiv \left[\begin{array}{cc} \mathbf{K}_1\mathbf{K}_1^\top & -\mathbf{K}_1\mathbf{K}_1^\top\pi_\infty^\top \\ -\pi_\infty\mathbf{K}_1\mathbf{K}_1^\top & \pi_\infty\mathbf{K}_1\mathbf{K}_1^\top\pi_\infty^\top \end{array}\right] \tag{6.9}$$

and the following criterion is obtained:

$$\mathcal{C}'_F(\mathbf{K}_i, \pi_\infty) = \sum_{i=\mathbf{2}}^{n} \left\| \mathbf{F}(\mathbf{K}_i\mathbf{K}_i^\top) - \mathbf{F}(\mathbf{P}_i\Omega_1^*\mathbf{P}_i^\top) \right\|_F \quad . \tag{6.10}$$

In this way 5 of the 8 parameters of the absolute conic are eliminated at once, which simplifies convergence issues. On the other hand this formulation implies a bias towards the first view since using this parameterization the equations for the first view are perfectly satisfied, whereas the noise has to be spread over the equations for the other views. In the experiments it will be seen that this is not suitable for longer sequences where in this case the present redundancy can not be used optimally. Therefore it is proposed to first use the simplified criterion of equation (6.10) and then to refine the results with the unbiased criterion of equation (6.8).

To apply this self-calibration method to standard zooming/focusing cameras, some assumptions should be made. Often it can be assumed that there is no skew and that the aspect ratio is tuned to one. If necessary (e.g. when only a short image sequence is at hand, when the projective calibration is not accurate enough or when the motion sequence is close to critical without additional constraints), it can also be used that the principal point is close to the center of the image. This leads to the following parameterizations for $\mathbf{K}_i$ (transform the images to have $(0,0)$ in the middle):

$$\mathbf{K}_i = \begin{bmatrix} f_i & 0 & c_{xi} \\ & f_i & c_{yi} \\ & & 1 \end{bmatrix} \text{ or } \mathbf{K}_i = \begin{bmatrix} f_i & 0 & 0 \\ & f_i & 0 \\ & & 1 \end{bmatrix} . \qquad (6.11)$$

These parameterizations can be used in (6.8). It will be seen in the experiments of Section 6.6 that this method gives good results on synthetic data as well as on real data.

6.3.2 Linear approach

The previous nonlinear approach needs an initialization. Therefore in this section a non-iterative approach is developed based on linear constraints. Even if some of these constraints are only roughly satisfied this can be sufficient to provide a suitable initialization for the nonlinear method.

To obtain a linear algorithm based on equation (4.3) linear constraints on the dual image absolute conic are needed. The problem is that the constraints are given in terms of intrinsic camera parameters which are squared in ω^*:

$$\omega^* \sim \mathbf{K}\mathbf{K}^\top \sim \begin{bmatrix} f_x^2 + s^2 + c_x^2 & sf_y + c_x c_y & c_x \\ sf_y + c_x c_y & f_y^2 + c_y^2 & c_y \\ c_x & c_y & 1 \end{bmatrix} \qquad (6.12)$$

However, a known principal point yields two independent linear constraints on ω^*. If the principal point is known then a known skew also results in a linear constraint (i.e. by transforming the image so that the skew vanishes). If these three constraints are available, then a known focal length or aspect ratio also results in a linear constraint. The different useful cases are summarized in Table 6.2.

Here case B is worked out, but the approach is similar for the two other cases. The

	unknown parameters	constraints	ω^*
case A	everything known but free/unknown focal length free/unknown aspect ratio	$c_x = c_y = s = 0$	$\begin{bmatrix} f_x^2 & 0 & 0 \\ 0 & f_y^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
case B	everything known but free/unknown focal length	$f_x = f_y$ $c_x = c_y = s = 0$	$\begin{bmatrix} f^2 & 0 & 0 \\ 0 & f^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
case C	everything known	$f_x = f_y = 1$ $c_x = c_y = s = 0$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Table 6.2: *Useful cases which can be solved with the linear algorithm*

constraints we impose simplify equation (4.3) as follows:

$$\lambda \begin{bmatrix} f_i^2 & 0 & 0 \\ 0 & f_i^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{P}_i \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \\ c_2 & c_5 & c_6 & c_7 \\ c_3 & c_6 & c_8 & c_9 \\ c_4 & c_7 & c_9 & c_{10} \end{bmatrix} \mathbf{P}_i^\top \tag{6.13}$$

with λ an explicit scale factor. From the left-hand side of equation (6.13) it can be seen that the following equations have to be satisfied:

$$\begin{aligned} \omega_i^{*(11)} &= \omega_i^{*(22)}, && (6.14)\\ \omega_i^{*(12)} &= \omega_i^{*(13)} = \omega_i^{*(23)} = 0 && (6.15)\\ \omega_i^{*(21)} &= \omega_i^{*(31)} = \omega_i^{*(32)} = 0 \ . && (6.16) \end{aligned}$$

with $\omega_i^{*(kl)}$ representing the element on row k and column l of ω_i^*. Note that due to symmetry (6.15) and (6.16) result in identical equations. These constraints can thus be imposed on the right-hand side, yielding 4 independent linear constraints in $c_i, i = 1 \ldots 10$ for every image:

$$\begin{aligned} P_i^{(1)} \Omega^* P_i^{(1)\top} &= P_i^{(2)} \Omega^* P_i^{(2)\top} \\ 2P_i^{(1)} \Omega^* P_i^{(2)\top} &= 0 \\ 2P_i^{(1)} \Omega^* P_i^{(3)\top} &= 0 \\ 2P_i^{(2)} \Omega^* P_i^{(3)\top} &= 0 \end{aligned}$$

with $P_i^{(k)}$ representing row k of $\mathbf{P}_i$ and Ω^* parameterized as in (6.13). The rank 3 constraint can be imposed by taking the closest rank 3 approximation (using SVD for example). This approach holds for sequences of 3 or more images.

The special case of 2 images can also be dealt with, but with a slightly different approach. When only two views are available the solution is only determined up to a one parameter family of solutions $\Omega^*_a + \gamma\Omega^*_b$. Imposing the rank 3 constraint in this case should be done through the determinant:

$$\det\left(\Omega^*_a + \gamma\Omega^*_b\right) = 0 \ . \tag{6.17}$$

This results in up to 4 possible solutions. The constraint that the squared parameters should be positive can be used to eliminate some of these solutions. If more than one solution persists additional constraints should be used. These can come from knowledge about the camera (e.g. constant focal length) or about the scene (e.g. known angle).

6.3.3 Maximum Likelihood approach

The calibration results could be refined even more through a Maximum Likelihood approach. Traditionally several assumptions are made in this case. It is assumed that the error is only due to mislocalization of the image features. Additionally, this error should be uniformly and normally distributed[2]. This means that the proposed camera model is supposed to be perfectly satisfied. In these circumstances the maximum likelihood estimation corresponds to the solution of a least-squares problem. In this case a criterion of the type of equation (4.2) should be minimized:

$$\mathcal{C}_{ML}(\mathtt{M}_l, \mathbf{K}_i, \mathbf{R}_i, \mathtt{t}_i) = \sum_{i=1}^{n} \sum_{l \in I_i} \left((x_{li} - \frac{\mathtt{P}_{i1}\mathtt{M}_l}{\mathtt{P}_{i3}\mathtt{M}_l})^2 + (y_{li} - \frac{\mathtt{P}_{i2}\mathtt{M}_l}{\mathtt{P}_{i3}\mathtt{M}_l})^2 \right) \tag{6.18}$$

where I_i is the set of indices corresponding to the points seen in view i and $\mathbf{P}_i \equiv \left[\mathtt{P}_{i1}^\top \mathtt{P}_{i2}^\top \mathtt{P}_{i3}^\top\right]^\top = \mathbf{K}_i[\mathbf{R}_i^\top|\text{-}\mathbf{R}_i^\top \mathtt{t}_i]$. In this equation $\mathbf{K}_i$ should be parameterized so that the self-calibration constraints are satisfied. The model could also be extended with parameters for radial distortion.

An interesting extension of this approach would be to introduce some uncertainty on the applied camera model and self-calibration constraints. Instead of having hard constraints on the intrinsic camera parameters imposed through the parameterization, one could impose soft constraints on these parameters through a trade-off during the minimization process. This would yield a criterion of the following form:

$$\begin{aligned} \mathcal{C}'_{ML}(\mathtt{M}_l, \mathbf{K}_i, \mathbf{R}_i, \mathtt{t}_i) \quad = \quad & \textstyle\sum_{i=1}^{n} \sum_{l \in I_i} \left((x_{li} - \frac{\mathtt{P}_{i1}\mathtt{M}_l}{\mathtt{P}_{i3}\mathtt{M}_l})^2 + (y_{li} - \frac{\mathtt{P}_{i2}\mathtt{M}_l}{\mathtt{P}_{i3}\mathtt{M}_l})^2 \right) \\ & + \textstyle\sum_{i=1}^{n} \sum_{k=1}^{m} \lambda_k C_{ki}(\mathbf{K}_i)^2 \end{aligned} \tag{6.19}$$

with λ_k a regularization factor and $C_{ki}(\mathbf{K}_i)$ representing the constraints on the intrinsic camera parameters, e.g. $C_{1i} = f_{x_i} - f_{y_i}$ (known aspect ratio), $C_{2i} = u_{xi}$ (known principal point) or $f_{x_i} - f_x$ (constant focal length). The values of the factors λ_k depend on how strongly the constraints should be enforced.

---

[2] This is a realistic assumption since outliers should have been removed at this stage of the processing.

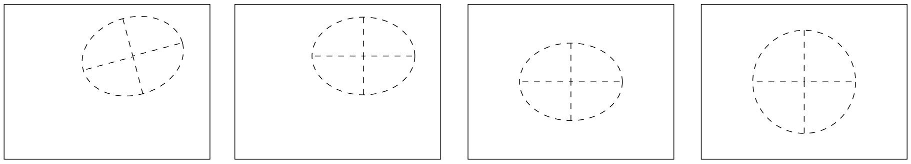

Figure 6.4: *Potential images of the absolute conic for different constraints. From left to right more constraints are added. No constraints (far left), known skew (middle left), known principal point (middle right) and known aspect ratio (far right). A known focal length would result in a known radius for the image of the absolute conic. Note that the projection of the true absolute conic in the retinal plane is always* **I***, what is observed is, however, the image plane.*

6.4 Critical motion sequences

In this chapter a self-calibration method was proposed that could deal with different sets of constraints on the intrinsic camera parameters. In this case critical motion sequences can, of course, also occur. The type of critical motion sequences which can occur, however, depend on the constraints which are imposed for self-calibration. The extremes being all parameters known, in which case almost no degeneracies exist; and, no constraints at all, in which case all motion sequences are critical.

The work of Sturm on critical motion sequences [152, 153, 150] can be used as a starting point. This analysis supposed that all intrinsic camera parameters were constant. The algorithm proposed in this chapter can however deal with known, constant or varying camera parameters.

Known intrinsic parameters are relatively easy to deal with since they reduce the possibilities for critical motion sequences. Only the potential absolute conics which satisfy the constraints have to be dealt with. Since Sturm's analysis was done for a constant image of the absolute conic, all the conics of a specific constant parameter critical motion sequence will satisfy the constraint or none will. This means that the class of motion involved will be the same when some parameters are known and the others are constant (assuming this type is still critical with the added constraints). The ambiguity, however, will be reduced since potential absolute conics have to satisfy the imposed constraints. Figure 6.4 illustrates the effect of the constraints on the image absolute conic. Note for example that in the case where only the focal length is not known the projection cone of the image absolute conic is bound to be a circular cone.

Varying intrinsic parameters are more difficult to deal with. An analysis can be made along the same lines as for constant intrinsic camera parameters: *Given a specific type of conic, which combinations of position and orientation satisfy the constraints?* From this classes of critical motion sequences can be derived.

6.4.1 Critical motion sequences for varying focal length

One of the most interesting cases consists of a camera with all intrinsic parameters known, except the focal length which can vary. To analyze CMS it is simpler to use the projection of the absolute conic in the retinal plane instead of in the image. For a Euclidean representation the true image in the retinal plane is obviously **I**. Due to an unknown focal length, self-calibration algorithms have no way to discard alternative solutions which satisfy

$$\omega_i \sim \begin{bmatrix} \lambda_i^{-2} & 0 & 0 \\ 0 & \lambda_i^{-2} & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{6.20}$$

A $\lambda_i \neq 1$ corresponds to a wrong estimate $\tilde{f}_i$ of the focal length (i.e. $\tilde{f}_i = \lambda_i f_i$ instead of f_i). One can also deduce from equation (6.20) that ω_∞ must be of the form $\omega_\infty \sim diag(a, a, 1)$. Here again the cases for potential absolute conics on and off the plane at infinity should be considered.

Potential absolute conics in the plane at infinity In this case only double or triple eigenvalues are possible for the absolute conic. Since a triple eigenvalue corresponds to the real absolute conic, only the case of a double eigenvalue should be considered. The projection of the absolute conic on its image for a camera with orientation $\mathbf{R}_i$ is described as $\omega_i \sim \mathbf{R}_i \omega_\infty \mathbf{R}_i^\top$. Due to the form of ω_i and ω_∞ the possible orientations are:

$$\mathbf{R} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ or } \mathbf{R} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ \sin\alpha & -\cos\alpha & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{6.21}$$

CMS-Class F1 This means that critical motion sequences of this class consist of motions for which the rotations are about the optical axis by an arbitrary angle or by 180 degrees about an axis perpendicular to the optical axis.

Potential absolute conic out of the plane at infinity Since we have to deal with a *circular cone* the classes of Section 4.4.3 based on elliptic cones can be discarded immediately. The locus is however more general since for almost every λ another position can be found. Note that the projection cone corresponding to the true absolute conic is always an absolute one and that therefore cylindrical cones correspond to wrong estimates of the focal length. Therefore in this context relative focal length means the estimated focal length over real focal length.

CMS-Class F2 An *ellipse* as potential absolute conic is obtained when the motion is restricted to a hyperbola and an associated ellipse in an orthogonal plane (this specific case was worked out by Sturm [155]). Since a unique potential absolute conic corresponds to a hyperbola and the same is true for the ellipse, restricting the motion to one of them does not increase the criticality of the sequence. This case is illustrated in Figure 6.5 based on the principle we have derived in Appendix C. The optical axis of the camera must be tangent to the hyperbola resp. ellipse; arbitrary rotation about

Class	Description	$\#\mathtt{t}$	$\#\mathbf{R}$	$\#\Omega$
F1	translations and rotations about opt. axis	∞^3	$2 \times \infty$	∞
F2	hyperbolic and/or elliptic motion	$2 \times \infty$	$2 \times \infty$	2
F3		∞	$2 \times \infty$ or ∞^3	2
F3.1	forward motion	∞	$2 \times \infty$	∞^2
F3.2		2	∞^3	2
F3.2.1	pure rotations	1	∞^3	∞^3

Table 6.3: *Classes of critical motion sequences for varying focal length. $\#\mathtt{t}$ and $\#\mathbf{R}$ represent respectively the number of different positions and the different orientations at each position of which a critical motion sequence of the specific class can consist. $\#\omega$ indicates the minimum level of ambiguity on the absolute conic for this class of critical motion sequences.*

the optical axis, and, of 180 degrees about an axis perpendicular to it, are thus allowed. Note that, to every position corresponds a specific λ (see figure).

CMS-Class F3 A *circle* as potential absolute conic is obtained when the motion is restricted to a line. The orientation should be so that the optical axis is aligned with this line, except for two positions where an arbitrary orientation is allowed (for these two points the projection cone is an absolute cone). There are several important subclasses for this class.

CMS-Class F3.1 All optical axes aligned with the line containing all the centers of projection. In this case all circles in planes perpendicular to the line (including the plane at infinity) and centered around the lines intersection with this plane are potential absolute conics.

CMS-Class F3.2 All centers of projection restricted to two positions. An arbitrary orientations is allowed for the cameras. In this case the projection cones are an absolute cones. Note that this case is still degenerate when the focal length is known!

CMS-Class F3.2.1 If only one of the two positions is used then the position of the plane at infinity is completely free. This corresponds to pure rotation.

The different classes of critical motion sequences for varying focal length are summarized in Table 6.3.

Specific degeneracies of the linear algorithms

In the case of a linear algorithm some specific degeneracies can occur. These are due to the fact that linear algorithms only impose the rank 3 constraint of Ω^* a posteriori. Note that this is not specific to the algorithm that was proposed in this chapter. Triggs also reported this problem for his linear algorithm for constant intrinsic camera parameters [166]. Sturm also discussed specific algorithm dependent degeneracies [150]. The problem of the linear algorithms and of the Kruppa equations consist in the fact that the planarity of the absolute conic is not enforced.

Specific degeneracies occur when the camera fixes a real point. In the case of this algorithm the degeneracies only occur when the fixed point corresponds with

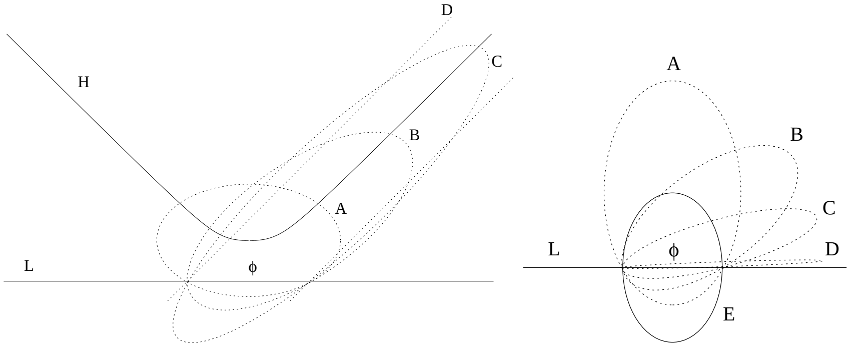

Figure 6.5: *Critical motion sequence for varying focal length, Class F2. Motion along a hyperbola (*$\mathbf{H}$*, left) and/or an ellipse (*$\mathbf{E}$*, right). The potential absolute conic* ϕ *is the intersection of the circular cones* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *(represented by ellipsoids as explained in Appendix C) with the plane* Π*. It has the same excentricity as the ellipsoid* $\mathbf{A}$ *(in both cases). The plane containing* $\mathbf{H}$*, the plane containing* $\mathbf{E}$ *and the plane* Π *are mutually orthogonal. Note that for the ellipsoids associated with the hyperbola (left) the smallest eigenvalue is double while on for the ones associated with the ellipse (right) it is the largest one that is double. The factor* λ *is given by the ratio of the single eigenvalue (associated with the optical axis) and the double eigenvalue (associated with the image plane). This gives the ratio between the true focal length and the one obtained by considering* ϕ *as the absolute conic.*

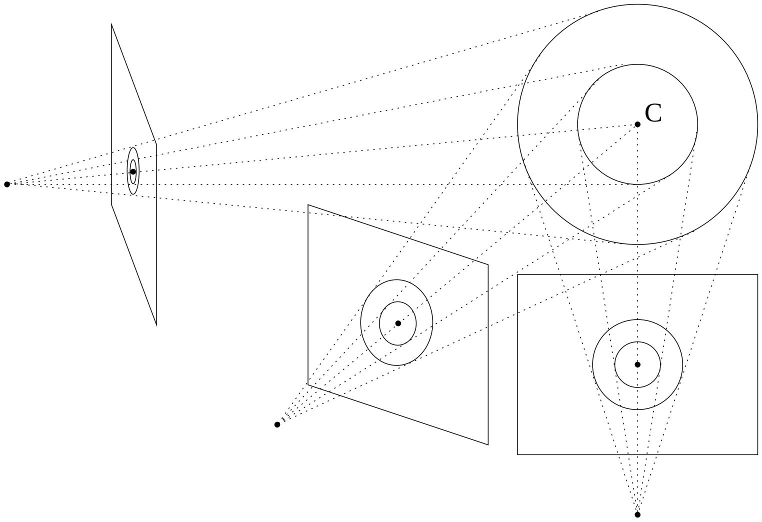

Figure 6.6: *Specific degeneracies of the linear algorithm. If point* `C` *is fixated at the center of the image, all real and imaginary spheres with their center in* `C` *will satisfy the constraints of the linear algorithm (case A and B). Note that the distance between the camera and the point* `C` *can vary freely.*

the principal point. Since the fixation point is typically kept close to the middle of the image and thus close to the principal point, this constraint is not very helpful in practice. Since the focal length is free to vary, varying the distance to the fixated point does not restrict the ambiguity.

In this case all spheres –real or imaginary– with their center at the fixed point are solutions of the linear algorithm since it does not enforce rank 3. Suppose for example that the fixed point is $\mathtt{C} = [0001]^\top$, then the following quadrics will all satisfy the self-calibration constraints:

$$\Phi^* = \Omega^* - R^{-2}\mathtt{C}\mathtt{C}^\top = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -R^{-2} \end{bmatrix} \tag{6.22}$$

This is illustrated in Figure 6.6. The only constraint that can still differentiate the absolute quadric from all the others is that it has exactly rank 3. Note that the double point quadric $\mathtt{C}\mathtt{C}^\top$ has only rank 1. A nonlinear algorithm enforcing that the determinant of the quadric is zero would thus probably end up with Ω^* (given a good initialization), but could also yield $\mathtt{C}\mathtt{C}^\top$. A linear algorithm will just give any solution. Thus, an additional class of critical motion sequences exists which is specific to the linear algorithm:

CMS-Class FL In this case the position is unrestricted ($\#\mathtt{t} = \infty^3$). The orientation is so that a specific point is always projected on the principal point ($\#\mathbf{R} = 2\times\infty$). The absolute conic is only determined up to one parameter ($\#\Omega = \infty$).

Note that for most classes of Table 6.3 the dimensionality of the solution space (i.e. $\#\Omega$) will be higher for the linear algorithm.

Note also that if the focal length is known (see Table 6.2, case C) then almost no degeneracies can occur (i.e. only CMS-Class F4 and F4.1, see Table 6.3). In some critical cases it can thus be useful to use an approximate focal length to initialize the nonlinear algorithm. How this can be done automatically is explained in Section 6.5.

6.4.2 Detecting critical motion sequences

A complete description of all types of critical motion sequences for all different possible sets of constraints would take us too far. For the specific case where all intrinsic camera parameters are fixed, such an analysis was carried out by Sturm [152].

Here a more practical approach is taken. Given an image sequence, a method is given to analyze if that particular sequence is (close to) critical or not. The method can deal with all different combinations of constraints. It is based on a sensitivity analysis towards the constraints. An important advantage of the technique is that it also indicates quasi-critical motion sequences. It can be used on a synthetic motion sequence as well as on a real image sequence from which the –potentially ambiguous– rigid motion sequence was obtained through self-calibration.

Without loss of generality the calibration matrix can be chosen to be $\mathbf{K} = \mathbf{I}$ (and thus $\omega_i^* = \mathbf{I}$). In the case of real image sequences this implies that the images should first be transformed with $\mathbf{K}^{-1}$. In this case it can be verified that $df_x = \frac{1}{2}d(\omega_i^*)_{11}$,

$df_y = \frac{1}{2}d(\omega_i^*)_{22}$, $du = d(\omega_i^*)_{13} = d(\omega_i^*)_{31}$, $dv = d(\omega_i^*)_{23} = d(\omega_i^*)_{32}$ and $ds = d(\omega_i^*)_{12} = d(\omega_i^*)_{21}$. Now the typical constraints which are used for self-calibration can all be formulated as linear equations in the coefficients of $\mathbf{K}$. As an example of such a system of equations, consider the case $s = 0$, $\frac{f_y}{f_x} = 1$ and f_x=constant. By linearizing around $\omega^* = \mathbf{I}$ this yields $d(\omega_i^*)_{12} = 0$, $d(\omega_i^*)_{11} = d(\omega_i^*)_{22}$, $d(\omega_i^*)_{11} = d{\omega^*_1}_{11}$. Which can be rewritten as

$$\begin{bmatrix} 0 & 0 & 1 & \cdots & 0 & \cdots \\ 1 & 0 & 0 & \cdots & -1 & \cdots \\ 1 & -1 & 0 & \cdots & 0 & \cdots \end{bmatrix} \begin{bmatrix} d{\omega^*_1}_{11} \\ d{\omega^*_1}_{22} \\ d{\omega^*_1}_{12} \\ \vdots \\ d{\omega^*_2}_{11} \\ \vdots \end{bmatrix} = 0 \ . \tag{6.23}$$

More in general the linearized self-calibration equations can be written as follows:

$$\mathbf{C}d\omega^* = 0 \tag{6.24}$$

with $d\omega^*$ a column vector containing the differentials of the coefficients of the dual image absolute conic ω_i^* for all views. The matrix $\mathbf{C}$ encodes the imposed set of constraints. Since these equations are satisfied for the exact solution, this solution will be an isolated solution of this system of equations if and only if any arbitrary small change to the solution violates at least one of the conditions of equation (6.24). Using equation (4.3) a small change can be modeled as follows:

$$\mathbf{C}d\omega^* = \mathbf{C}\left[\frac{d\omega^*}{d\Omega^*}\right]d\Omega^* = \mathbf{C}'d\Omega^* \tag{6.25}$$

with $\Omega^* = [\Omega^*_{11}\Omega^*_{22}\Omega^*_{12}\Omega^*_{31}\Omega^*_{32}\Omega^*_{14}\Omega^*_{24}\Omega^*_{34}]^\top$ and the Jacobian $\left[\frac{d\omega^*}{d\Omega^*}\right]$ evaluated at the solution. To have the expression of equation (6.25) different from zero for every possible $d\Omega^*$, means that the matrix $\mathbf{C}'$ should be of rank 8 ($\mathbf{C}'$ should have a right null space of dimension 0). In practice this means that all singular values of $\mathbf{C}'$ should significantly differ from zero, else a small change of the absolute quadric proportional to right singular vectors associated with small singular values will almost not violate the self-calibration constraints.

To use this method on results calculated from a real sequence the camera matrices $\mathbf{P}$ should first be adjusted to have the calculated solution become an exact solution of the self-calibration equations.

6.5 Constraint selection

In this chapter a method was proposed that could deal with different types of constraints on the intrinsic camera parameters. These parameters can be known, constant or free to vary. This technique therefore offers a lot of flexibility. On the other hand one must now determine which constraints should be used to determine the absolute quadric.

It would be nice if the algorithm could determine from the data which set of constraints is most appropriate for determining the metric structure of the scene. At first one could think of applying model selection based on some kind of Akaike's information criterion [1]. Kanatani [70] and Torr et al. [164] showed us that this could give good results to determine the best motion model. The principle is the following: *Select the model which has the smallest expected residual*. The expected residual consists of the actual residual plus a term penalizing for the complexity of the model.

The problem is, however, that, in the case of self-calibration, this can not be used. This can be understood as follows. Assume some projection matrices which have a very small reprojection error. All these matrices satisfy more or less the self-calibration constraints, but not exactly. It is then very probable that the best model (in the sense of Akaike) is the projective one.

This is very well, but we would have preferred an approximate calibration over no calibration at all. In fact what Akaike tells us is to keep the projective camera matrices after self-calibration (i.e. to use $\tilde{\mathbf{P}}_i\mathbf{T}^{-1}$ and not $\tilde{\mathbf{K}}_i[\tilde{\mathbf{R}}_i^\top | \text{-}\tilde{\mathbf{R}}_i^\top \tilde{\mathtt{t}}_i]$ with $\tilde{\mathbf{K}}_i$ exactly satisfying the self-calibration constraints).

The approach we propose is thus not to minimize the expected residue on the reconstruction, but the expected variance on the absolute quadric. The problem is however that no theory has been worked out for this yet. Intuitively one can understand that when not enough constraints are used the variance is high and that the use of non-consistent constraints will cause the variance to get larger.

Linear algorithm To deal with the degeneracies of the linear algorithm a simplified version of the above idea was implemented. For linear equations the ratio of the two smallest singular values gives an indication of the variance on the solution.

This value is compared for cases B (free focal length) and C (known focal length) of the linear algorithm. For case C a very crude estimate of the focal length can be taken. For the cases degenerate for B and not for C, the solution for B will have two comparably small singular values ($\sigma_{n-1} \approx \sigma_n \approx 0$) while the solution for C will only have one ($\sigma_{n-1} \gg \sigma_n \approx 0$).

In non near-degenerate cases a unique solution will clearly stand out for case B. This means that the smallest singular value will be much smaller than all the others ($\sigma_{n-1} \gg \sigma_n$). Enforcing a wrong value for the focal length will cause a large residue in this case. Since this corresponds to the square of the smallest singular value the ratio under consideration will be small ($\sigma_{n-1} \approx \sigma_n$).

Since most motions causing case B of the linear algorithm to fail can be solved correctly by the non-linear approach, the focal length can be released in the refinement step.

6.6 Experiments

In this section some experiments are described. First synthetic image sequences were used to assess the quality of the algorithm under simulated circumstances. Both the amount of noise and the length of the sequences were varied. Then results are given for two outdoor video sequences. Both sequences were taken with a standard semi-professional camcorder that was moved freely around the objects. The first sequence is the Castle sequence which was already used in Section 5.3.5. This sequence was filmed with constant camera parameters –like most algorithms require. The new algorithm –which doesn't impose this– could therefore be tested on this. A second sequence was recorded with varying intrinsic parameters. A zoom factor ($2\times$) was applied while filming.

6.6.1 Simulations

The simulations were carried out on sequences of views of a synthetic scene. The scene consisted of 50 points uniformly distributed in a unit sphere with its center at the origin. The intrinsic camera parameters were chosen as follows. The focal length was different for each view, randomly chosen with an expected value of 2.0 and a standard deviation of 0.5. The principal point had an expected value of $(0, 0)$ and a standard deviation of $0.1\sqrt{2}$. In addition the synthetic camera had an aspect ratio of one and no skew. The views were taken from all around the sphere and were all more or less pointing towards the origin. An example of such a sequence can be seen in Figure 6.7.

The scene points were projected into the images. Gaussian white noise with a known standard deviation was added to these projections. Finally, the self-calibration method proposed in this paper was carried out on the sequence. For the different algorithms the metric error was computed. This is the mean deviation between the scene points and their reconstruction after alignment. The scene and its reconstruction are aligned by applying the metric transformation which minimizes the difference between both. For comparison the same error was also calculated after alignment with a projective transformation. By default the noise had an equivalent standard deviation of 1.0 pixel for a 500×500 image. To obtain significant results every experiment was carried out 10 times and the mean was calculated.

To analyze the influence of noise on the algorithms, noise values of 0, 0.1, 0.2, 0.5, 1, 1.5 and 2 pixels were used on sequences of 6 views. The results can be seen in Figure 6.8. It can be seen that for small amounts of noise the more complex model (i.e. varying focal length and principal point) should be preferred. If more noise is added, the simple model (i.e. only focal length varying) gives the best results. This is due to the low redundancy of the system of equations for the models which, beside the focal length, also try to estimate the position of the principal point.

Another experiment was carried out to evaluate the performance of the algorithm for different sequence lengths. Sequences ranging from 4 to 40 views were used. A noise level of one pixel was used. The results are shown in Figure 6.9. For short image sequences the results are better when the principal point is assumed in the middle

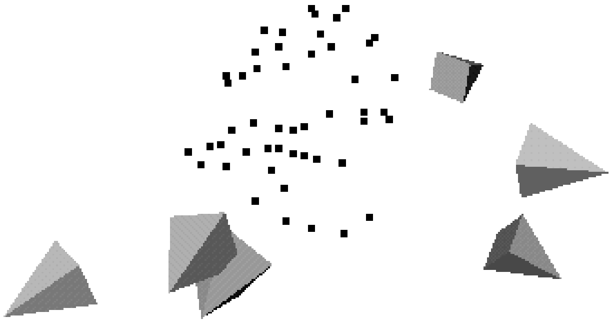

Figure 6.7: *Example of sequence used for simulations (the views are represented by the optical axis and the image axes of the camera in the different positions.)*

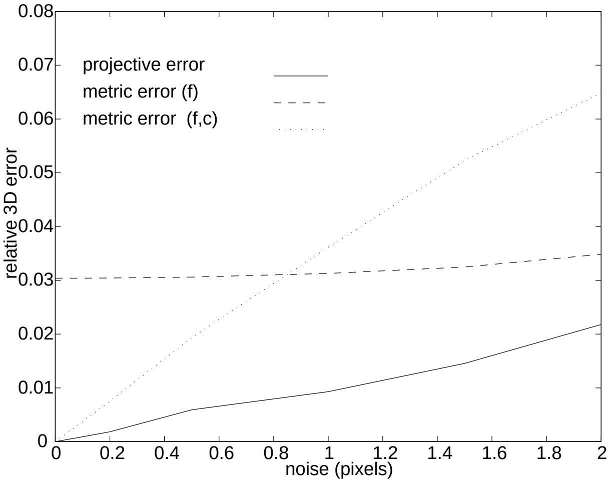

Figure 6.8: *Relative 3D error in function of noise*

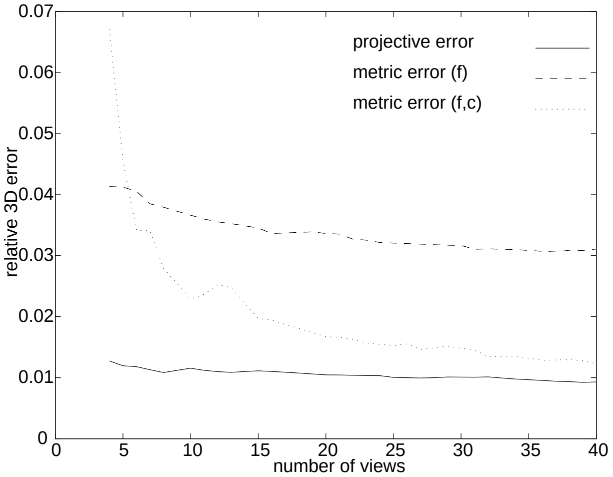

Figure 6.9: *Relative 3D error for sequences of different lengths*

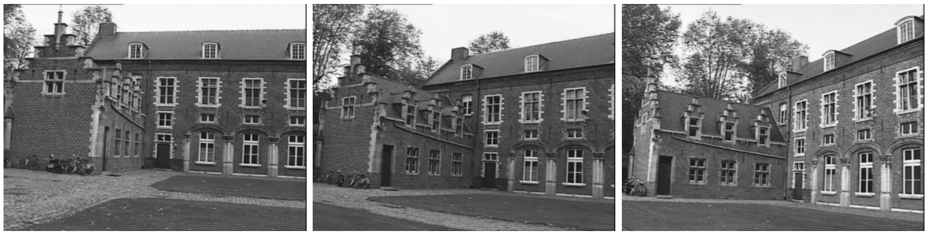

Figure 6.10: *Some of the Images of the Arenberg castle which were used for the reconstruction*

of the image, even though this is not exactly true. For longer image sequences the constraints on the aspect ratio and the image skew are sufficient to allow an accurate estimation of the metric structure of the scene. In this case fixing the principal point will degrade the results by introducing a bias.

6.6.2 Castle sequence

The first sequence showing part of the Arenberg castle was recorded with a fixed zoom/focus. It is therefore a good test for the algorithms presented in this chapter to check if they indeed return constant intrinsic parameters for this sequence. In Figure 6.10 some of the images of the sequence are shown, more of these can be seen in Figure 7.2.

The linear algorithm provides a good result, even though the motion sequence is close to critical with respect to this algorithm[3]. This solution is used as the initial-

---

[3] The reconstructed motion sequence (see Figure 6.13) is close to the **CMS class FL** described in Sec-

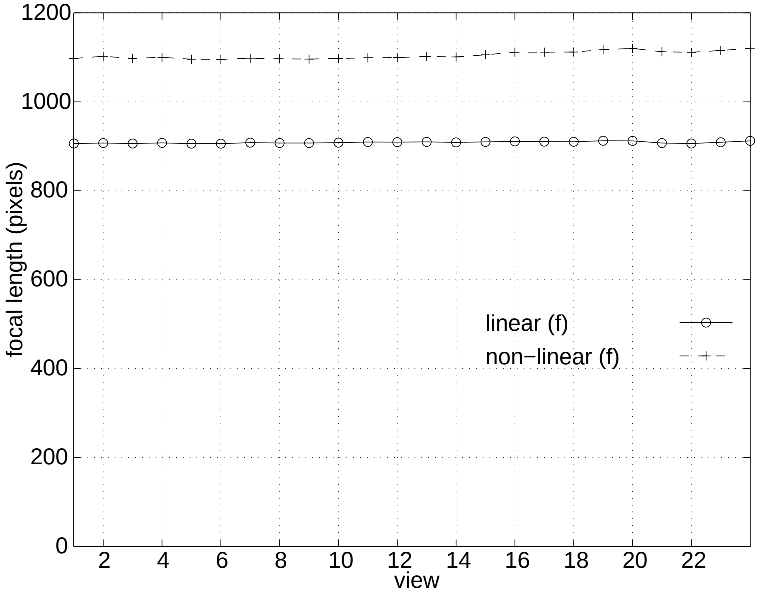

Figure 6.11: *focal length (in pixels) versus views for the different algorithms*

ization for the nonlinear algorithm. This algorithm converges without problems for the case of a varying focal length. After 6 iterations the norm of the residue vector is reduced from 0.57 to 0.11. When also the principal point is allowed to vary, the algorithm convergences to a wrong solution. This is most probably due to the restricted motion of the camera during the acquisition of the sequence.

In Figure 6.11 the computed focal lengths are plotted for every view, both for the linear algorithm and for the nonlinear algorithm. For both algorithms the retrieved focal length is almost constant, as it should be. This constant value however differs for both algorithms. The reason for this is explained based on the analysis of the jacobian of the self-calibration equations.

On the left side of Figure 6.12 the singular values of this jacobian are given. Note that the last one is much smaller than the others. The associated singular vector is shown on the right of this figure. It indicates that it is not possible to accurately determine the absolute value of the focal length from this specific image sequence[4]. This is not so bad since Bougnoux indicated that an inaccurate estimate of the focal length only had a relatively small effect on the reconstruction [14].

---

tion 6.4.1.

[4] Note that an orbital motion is not critical in this case. The problem is that the angle of rotation between the extreme views of the castle sequence is too small to allow for an accurate self-calibration. This was verified with synthetic data. Two sequences of orbital motion were generated. One of 60 degrees (similar to the castle sequence) and one of 360 degrees. The first sequences also yielded a jacobian with a small singular value. The jacobian of the second sequence did not have any small singular value. In fact, in terms of self-calibration, the castle sequence is relatively close to pure translation.

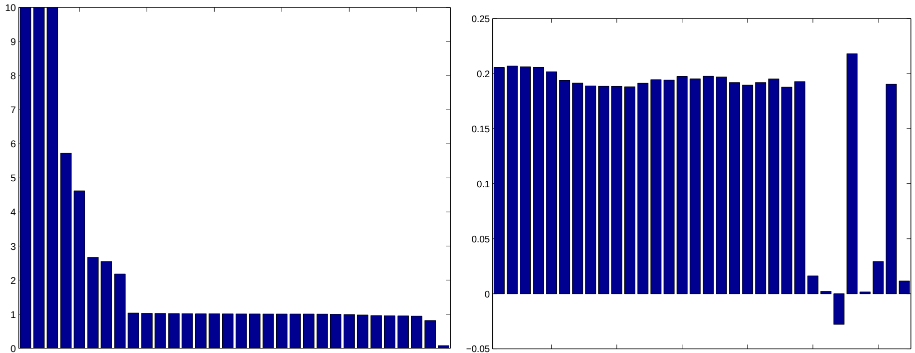

Figure 6.12: *Structure of the jacobian. Singular values (left) and right singular vector associated with the smallest singular value (right). The first 3 singular values were clipped, the values are 42, 39 and 31. The first 24 unknowns are associated with the focal lengths, the next 3 with the position of the plane at infinity and the last 5 with the absolute conic.*

Figure 6.13 shows the reconstruction together with the estimated viewpoints of the camera. In Figure 6.14 some orthographic views of the reconstructed scene are given. The resulting reconstruction is visually convincing and preserves the metric properties of the original scenes (i.e. parallelism, orthogonality, . . .).

A quantitative assessment of these properties can be made by explicitly measuring angles directly on the object surface. For this experiment 6 lines were placed along prominent surface features, three on each object plane, aligned with the windows. The three lines inside of each object plane should be parallel to each other (angle between them should be 0 degrees), while the lines of different object planes should be perpendicular to each other (angle between them should be 90 degrees). The measurement on the object surface shows that this is indeed close to the expected values (see Table 6.4).

6.6.3 Pillar sequence

This sequence shows a stone pillar with curved surfaces. While filming and moving away the zoom was changed to keep the image size of the object constant. The focal length was not changed between the two first images, then it was changed more or less linearly. From the second image to the last image the focal length has been doubled (if the markings on the camera can be trusted). In Figure 6.15 the 8 images of the

	angle (±std.dev.)
parallelism	1.0 ± 0.6 degrees
orthogonality	92.5 ± 0.4 degrees

Table 6.4: *Results of metric measurements on the reconstruction*

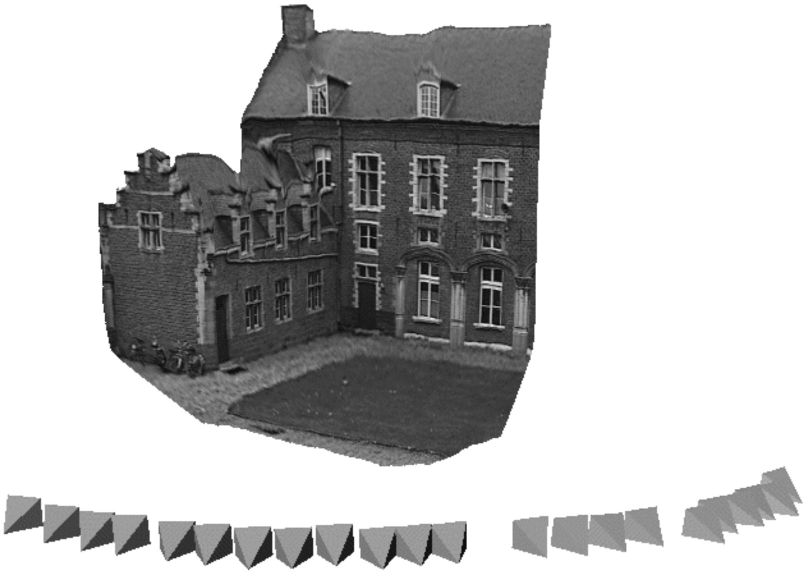

Figure 6.13: *Perspective view of the reconstruction together with the estimated position of the camera for the different views of the sequence*

sequence can be seen. Notice that the perspective distortion is most visible in the first images (wide angle) and diminishes towards the end of the sequence (longer focal length).

The linear self-calibration algorithm yields a good result. This solution is used as the initialization for the nonlinear algorithm. For a varying focal length the algorithm yields a solution after 2 iterations (the norm of the residue vector is reduced from 0.027 to 0.019). For varying focal length and principal point the algorithm needs 11 iterations (the norm of the residue vector is reduced from 0.027 to 0.001).

In Figure 6.16 the focal length for every view is plotted for the different algorithms. It can be seen that the calculated values of the focal length correspond to what could be expected. When the principal point was estimated independently for every view, it moved around up to 50 pixels. It is probable that the accuracy of the projective reconstruction is not sufficient to allow for the accurate estimation of a varying principal point when only a short sequence with not too much motion is available.

Similarly to the castle sequence, the changes in focal length seem to be more accurately retrieved than the absolute value of the focal length. This variation can again be explained by an analysis of the jacobian. On the left side of Figure 6.17 the singular values of this jacobian are given. Here also the last one is much smaller than the others (the ratio of the last two singular values is $\frac{0.029}{0.216}$). The associated singular vector is shown on the right of this figure. A change of the estimated calibration parameters proportional to this vector will only slightly affect the residue. Note that the longer focal lengths will be more affected by these variations (even proportionally)

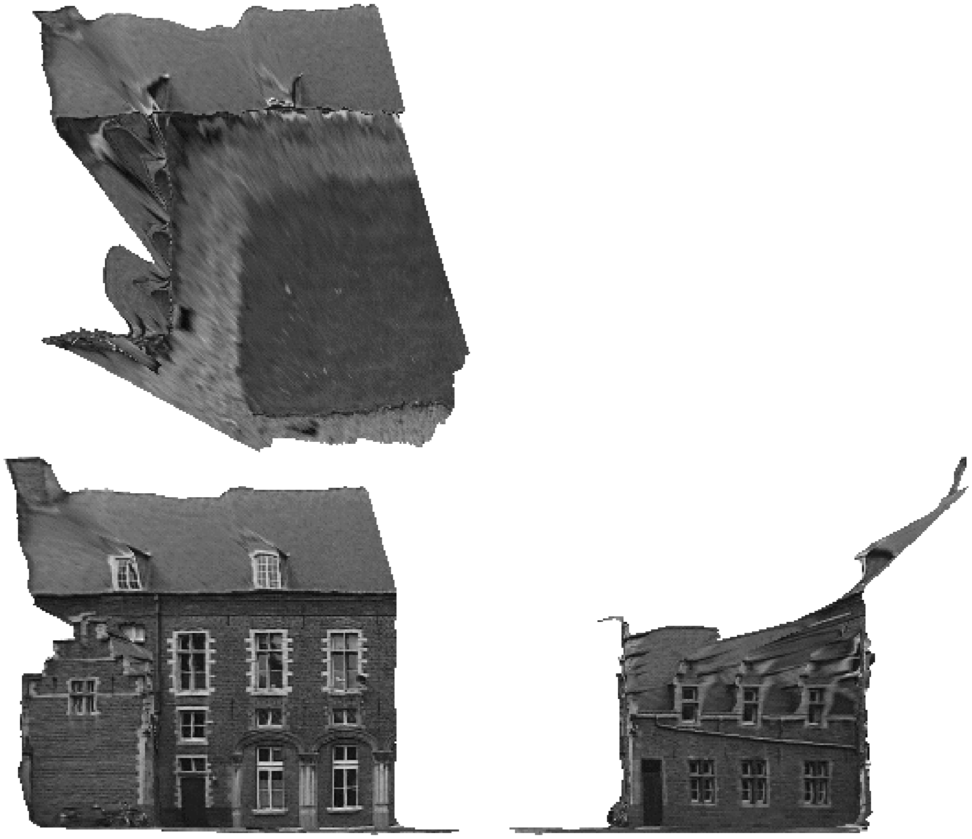

Figure 6.14: *Top, front ans side views of the Arenberg castle reconstruction. Orthographic projection was used to generate these views.*

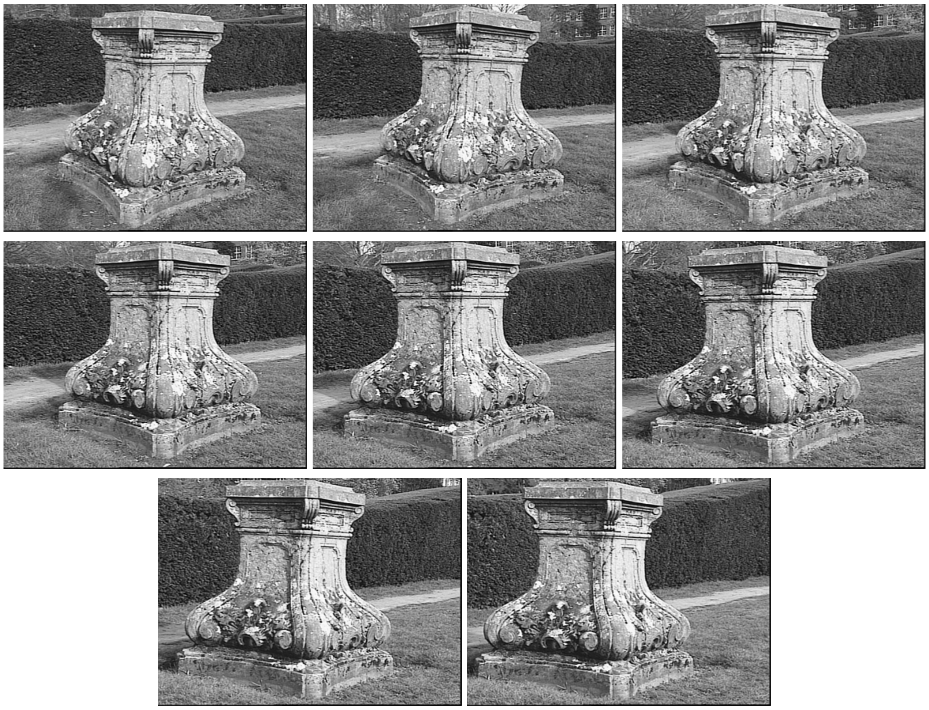

Figure 6.15: *Images of the pillar sequence. Note the short focal length/wide angle in the first image and the long focal length in the last image (close to orthographic)*

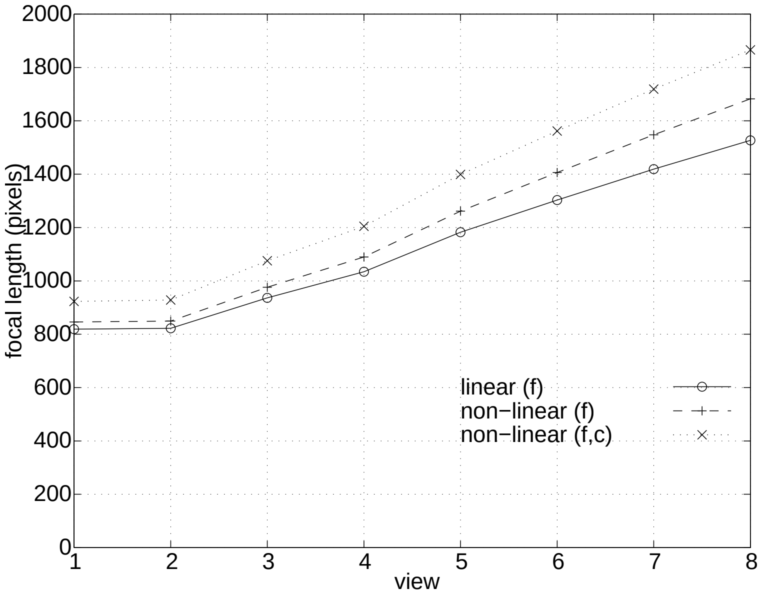

Figure 6.16: *focal length (in pixels) versus views for the different algorithms*

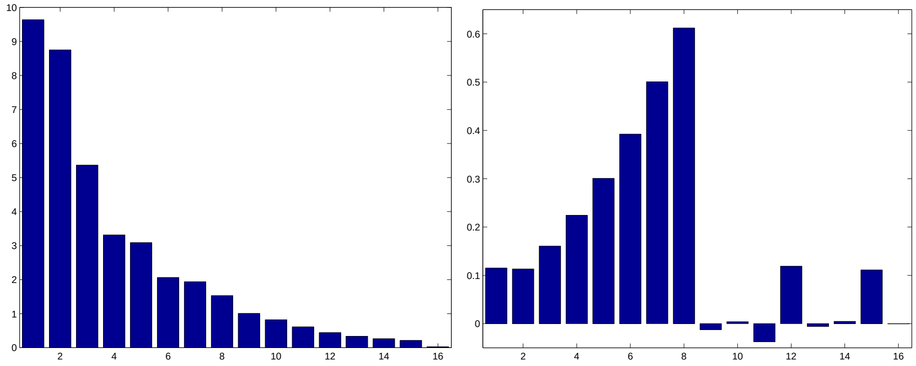

Figure 6.17: *Structure of the jacobian. Singular values (left) and right singular vector associated with the smallest singular value (right). The first 8 unknowns are associated with the focal lengths, the next 3 with the position of the plane at infinity and the last 5 with the absolute conic.*

than the shorter ones.

Figure 6.18 shows a top view of the reconstructed pillar together with the estimated camera viewpoints. These viewpoints are illustrated with small pyramids. Their height is proportional to the estimated focal length. In Figure 6.19 perspective views of the reconstruction are given. The view on top is rendered both shaded and

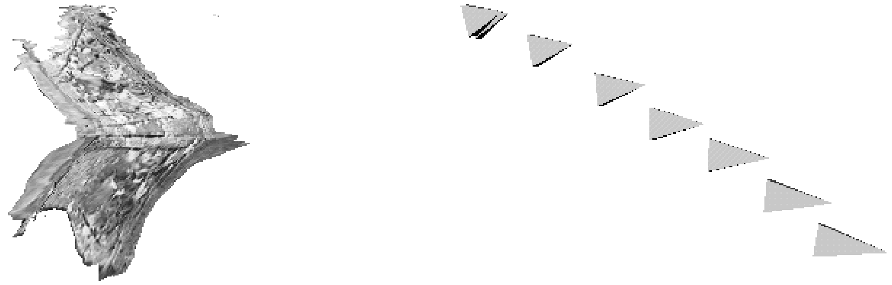

Figure 6.18: *Top view of the reconstructed pillar together with the different viewpoints of the camera. Note the change in focal length (height of pyramids).*

	ratio (±std.dev.)
all points	40.25 ± 2.2
interior points	± 0.9

Table 6.5: *Results of metric measurements on the reconstruction*

with surface texture mapped. The shaded view shows that even most of the small details of the object are retrieved. The bottom part shows a left and a right side view of the reconstructed object. Although there is some distortion at the outer boundary of the object, a highly realistic impression of the object is created. Note the free-form surface that has been reconstructed.

A quantitative assessment of the metric properties for the pillar is not so easy because of the curved surfaces. It is, however, possible to measure some distances on the real object as reference lengths and compare them with the reconstructed model. In this case it is possible to obtain a measure for the absolute scale and verify the consistency of the reconstructed lengths within the model. For this comparison a network of reference lines was placed on the original object and 27 manually measured object distances were compared with the reconstructed distances on the model surface, as seen in Figure 6.20. From each comparison the absolute object scale factor was computed. Due to the increased reconstruction uncertainty at the outer object silhouette some distances show a larger error than the interior points. This accounts for the outliers. The results are found in Table 6.5. Averaging all 27 measured distances gave a consistent scale factor of 40.25 with a standard deviation of 5.4% overall. For the interior distances, the reconstruction error dropped to 2.3%. These results demonstrate the metric quality of the reconstruction even for complicated surface shapes and varying focal length.

6.7 Conclusions

In this chapter we focussed on self-calibration in the presence of varying and unknown intrinsic camera parameters. The calibration models used in previous research are on

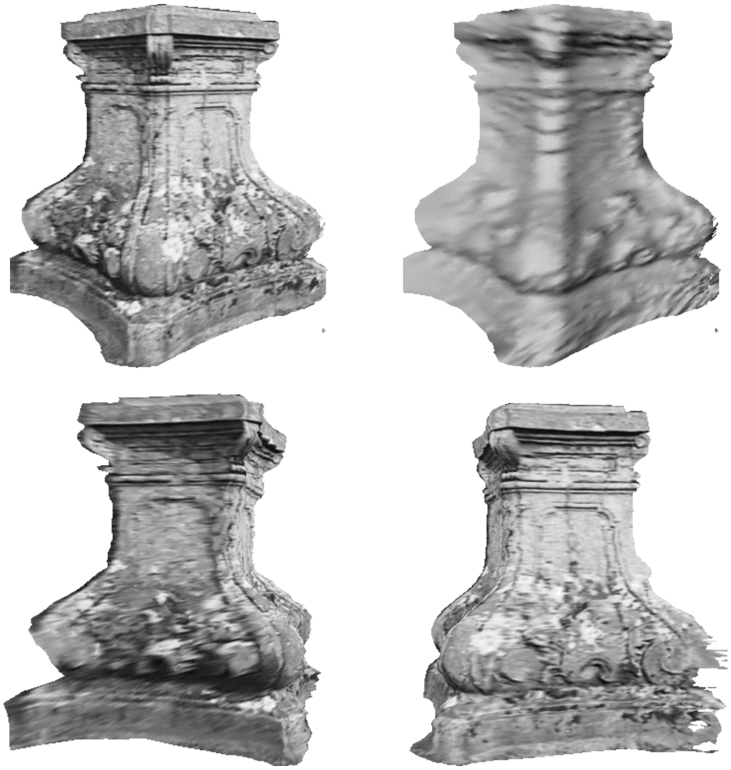

Figure 6.19: *Perspective views of the reconstruction (with texture and with shading)*

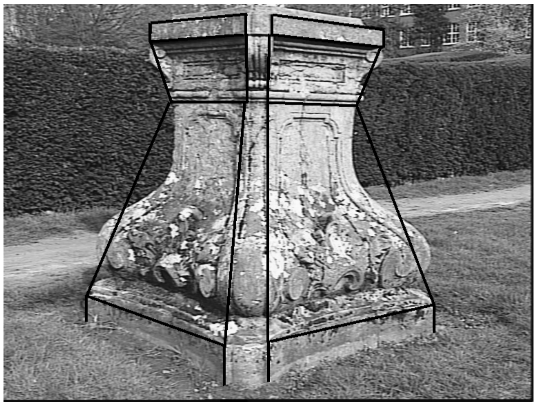

Figure 6.20: *To allow for a quantitative comparison between the real pillar and its reconstruction, some distances, superimposed in black, were measured.*

the one hand too restrictive in real imaging situations (constant parameters) and on the other hand too general (all parameters unknown). The more pragmatic approach which is followed here results in more flexibility for the image acquisition.

A counting argument was derived which gives the minimum number of views needed for self-calibration depending on which constraints are used. We proved that self-calibration is possible using only the most general constraint (i.e. that image rows and columns are orthogonal). Through geometric reasoning it was shown that in fact any known intrinsic camera parameter could be sufficient for this purpose. Of course more constraints will in general yield more accurate results.

A versatile self-calibration method which can work with different types of constraints (some of the intrinsic camera parameters constant or known) was derived. This method was then specialized towards the practically important case of a zooming/focusing camera (without skew and an aspect ratio $\frac{f_y}{f_x} = 1$). Both known and unknown principal points were considered. It is proposed to always start with the principal point in the center of the image and to first use the linear algorithm. The non-linear minimization is then used to refine the results, possibly –for longer sequences– allowing the principal point to be different for each image. This can however degrade the results if the projective calibration was not accurate enough, the sequence not long enough, or the motion sequence critical towards the set of constraints.

As for all self-calibration algorithms it is important to deal with critical motion sequences. In this chapter the problem of critical motion sequences for varying camera parameters was discussed. Some specific problems of the linear algorithm were analyzed. We also proposed a general method to detect critical and quasi-critical motion sequences. Some ideas are given on how constraints could be selected automatically.

The self-calibration algorithms are validated by experiments which are carried out on synthetic as well as real image sequences. The former are used to analyze noise sensitivity and influence of the length of the sequence. The latter show the practical feasibility of the approach.

Chapter 7

Metric 3D Reconstruction

7.1 Introduction

Obtaining 3D models from objects is an ongoing research topic in computer vision. A few years ago the main applications were robot guidance and visual inspection. Nowadays however the emphasis is shifting. There is more and more demand for 3D models in computer graphics, virtual reality and communication. This results in a change in emphasis for the requirements. The visual quality becomes one of the main points of attention.

The acquisition conditions and the technical expertise of the users in these new application domains can often not be matched with the requirements of existing systems. These require intricate calibration procedures every time the system is used. There is an important demand for flexibility in acquisition. Calibration procedures should be absent or restricted to a minimum.

Additionally, the existing systems are often built around specialized hardware (e.g. laser range scanners or stereo rigs) resulting in a high cost for these systems. Many new applications however require robust low cost acquisition systems. This stimulates the use of consumer photo- or video cameras.

In this chapter we present a system which retrieves a 3D surface model from a sequence of images taken with off-the-shelf consumer cameras. The user acquires the images by freely moving the camera around the object. Neither the camera motion nor the camera settings have to be known. The obtained 3D model is a scaled version of the original object (i.e. a *metric* reconstruction), and the surface texture is obtained from the image sequence as well.

Other researchers have presented systems for extracting 3D shape and texture from image sequences acquired with a freely moving camera. The approach of Tomasi and Kanade [159] used an affine factorization method to extract 3D from image sequences. An important restriction of this system is the assumption of orthographic projection.

Debevec, Taylor and Malik [26, 158, 27] proposed a system that starts from an approximate 3D model and camera poses and refines the model based on images. View dependent texturing is used to enhance realism. The advantage is that only a restricted number of images are required. On the other hand a preliminary model must be available and the geometry should not be too complex.

Our system uses full perspective cameras and does not require prior models nor calibration. It combines state-of-the-art algorithms of different domains: *projective reconstruction*, *self-calibration* and *dense depth estimation*.

The rest of the chapter is organized as follows: In section 7.2 a general overview of the system is given. In the subsequent sections the different steps are explained in more detail: projective reconstruction (section 7.3), self-calibration (section 7.4), dense matching (section 7.5) and model generation (section 7.6). Section 7.7 gives an idea of the required computation times while Section 7.8 proposes some possible improvements. Section 7.9 concludes the chapter.

7.2 Overview of the method

The presented system gradually retrieves more information about the scene and the camera setup. The first step is to relate the different images. This is done pairwise by retrieving the epipolar geometry. An initial reconstruction is then made for the first two images of the sequence. For the subsequent images the camera pose is estimated in the projective frame defined by the first two cameras. For every additional image that is processed at this stage, the interest points corresponding to points in previous images are reconstructed, refined or corrected. It is not necessary that the initial points stay visible throughout the entire sequence. For sequences where points can disappear and reappear (i.e. the images should sometimes also be matched to other images than the previous one), the algorithm can be adapted to efficiently deal with this. The result of this step is a reconstruction of typically a few hundred to a few thousand interest points and the (projective) pose of the camera. The reconstruction is only determined up to a projective transformation.

The next step is to restrict the ambiguity of the reconstruction to a metric one. In a projective reconstruction not only the scene, but also the camera is distorted. Since the algorithm deals with unknown scenes, it has no way of identifying this distortion in the reconstruction. Although the camera is also assumed to be unknown, some constraints on the intrinsic camera parameters (e.g. rectangular or square pixels, constant aspect ratio, principal point in the middle of the image, ...) can often still be assumed. A distortion on the camera mostly results in the violation of one or more of these constraints. A metric reconstruction/calibration is obtained by transforming the projective reconstruction until all the constraints on the cameras intrinsic parameters are satisfied.

At this point the system effectively disposes of a calibrated image sequence. The relative position and orientation of the camera is known for all the viewpoints. This calibration facilitates the search for corresponding points and allows us to use a stereo algorithm that was developed for a calibrated system and which allows to find correspondences for most of the pixels in the images.

From these correspondences the distance from the points to the camera center can be obtained through triangulation. These results are refined and completed by combining the correspondences from multiple images.

A dense metric 3D surface model is obtained by approximating the depth map with a triangular wireframe. The texture is obtained from the images and mapped onto the surface.

In figure 7.1 an overview of the system is given. It consists of independent modules which pass on the necessary information to the next modules. The first module computes the projective calibration of the sequence together with a sparse reconstruction. In the next module the metric calibration is computed from the projective camera matrices through self-calibration. Then dense correspondence maps are estimated. Finally all results are integrated in a textured 3D surface reconstruction of the scene under consideration.

Throughout the rest of this chapter the different steps of the method will be explained in more detail. The **castle sequence** of the Arenberg castle in Leuven will be used for illustration. The same sequence was used to validate the self-calibration algorithms of the previous chapters (see Section 5.3.5 and 6.6.2). Images of this sequence can be seen in Figure 7.2. The full sequence consists of 24 images recorded with a video camera.

7.3 Projective reconstruction

At first the images are completely unrelated. The only assumption is that the images form a sequence in which consecutive images do not differ too much. Therefore the local neighborhood of image points originating from the same scene point should look similar if images are close in the sequence. This allows for automatic matching algorithms to retrieve correspondences. The approach taken to obtain a projective reconstruction is very similar to the one proposed by Beardsley, Zisserman and Murray [8] (see also Beardsley et al. [9]).

7.3.1 Relating the images

It is not feasible to compare every pixel of one image with every pixel of the next image. It is therefore necessary to reduce the combinatorial complexity. In addition not all points are equally well suited for automatic matching. The local neighborhoods of some points contain a lot of intensity variation and are therefore easy to differentiate from others. An interest point detector is used to select a certain number of such suited points. These points should be well located and indicate salient features that stay visible in consecutive images. Schmid et al. [141] compared several interest point detectors. The Harris corner detector [50] obtained the best repeatability (i.e. invariance to rotation, scale and illumination) and the highest information content (in the local neighborhood around the extracted points). This detector is therefore used to extract the points of interest. Correspondences between these image points need to be established through a matching procedure.

Matches are determined through normalized cross-correlation of the intensity values of the local neighborhood. Since images are supposed not to differ too much,

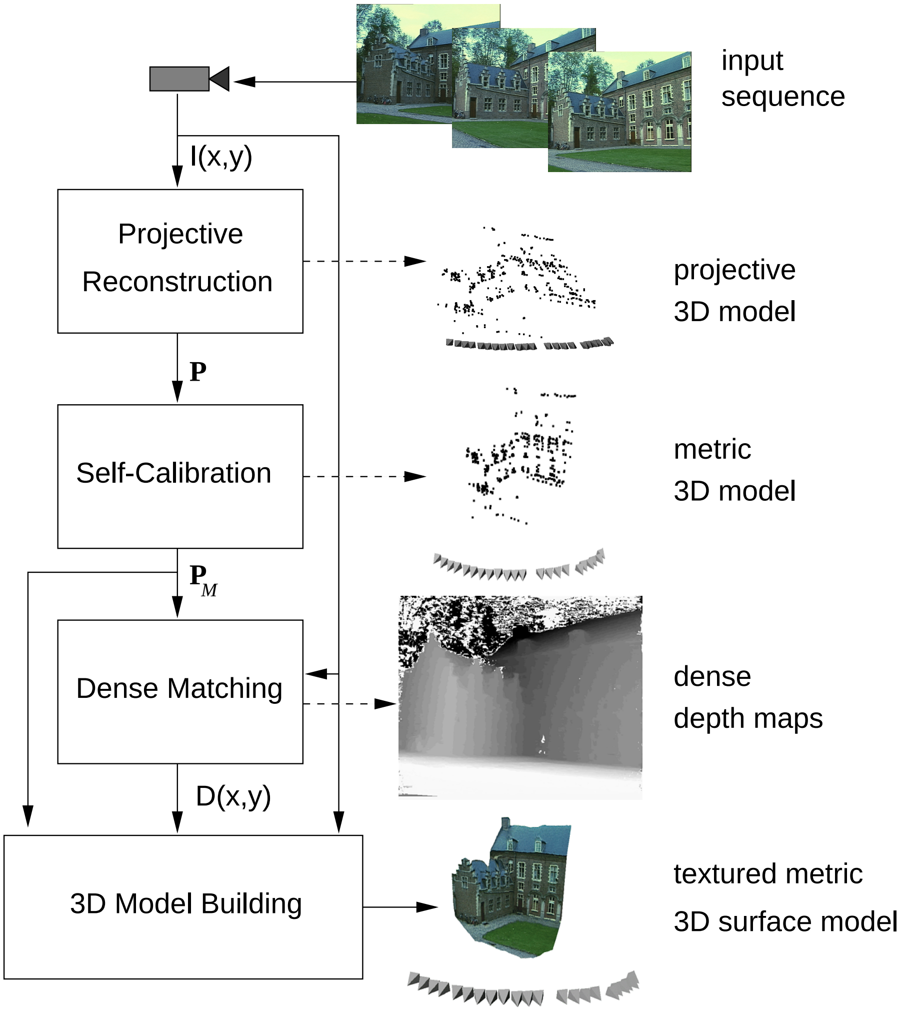

Figure 7.1: *Overview of the system: from the image sequence* $(I(x,y))$ *the projective reconstruction is computed; the projection matrices* $\mathbf{P}$ *are then passed on to the self-calibration module which delivers a metric calibration* $\mathbf{P}_M$*; the next module uses these to compute dense depth maps* $D(x,y)$*; all these results are assembled in the last module to yield a textured 3D surface model. On the right side the results of the different modules are shown: the preliminary reconstructions (both projective and metric) are represented by point clouds, the cameras are represented by little pyramids, the results of the dense matching are accumulated in dense depth maps (light means close and dark means far).*

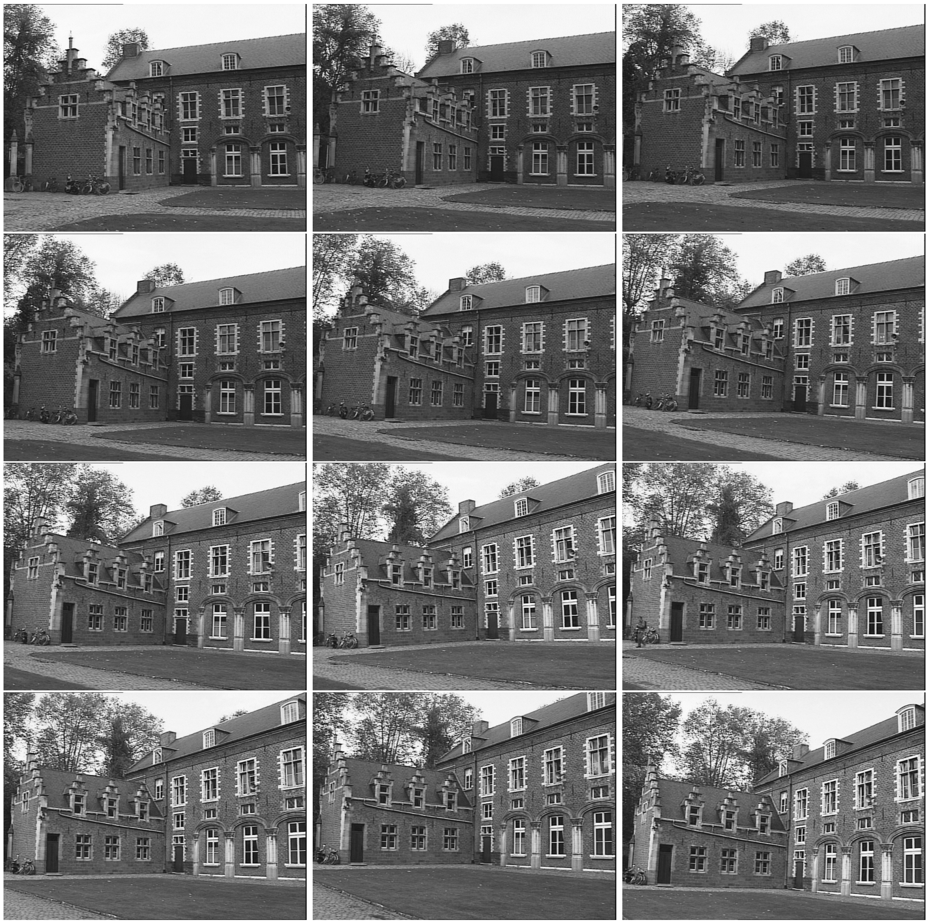

Figure 7.2: *Some images of the Arenberg castle sequence (every second image is shown). This sequence is used throughout this chapter to illustrate the different steps of the reconstruction system.*

```
• repeat
    - take minimal sample (7 matches)
    - compute $\mathbf{F}$
    - estimate $\%inliers$
  until $P_{\mathsf{OK}}(\%inliers, \#trials) > 95\%$
• refine $\mathbf{F}$ (using all inliers)
```

Table 7.1: *Robust estimation of the epipolar geometry from a set of matches containing outliers using RANSAC (P_{OK} indicates the probability that the epipolar geometry has been correctly estimated).*

corresponding points can be expected to be found in the same region of the image. Therefore at first only interest points which have similar positions are considered for matching. When two points are mutual best matches they are considered as potential correspondences.

Since the epipolar geometry describes the complete geometry relating two views, this is what should be retrieved. Computing it from the set of potential matches through least squares does in general not give satisfying results due to its sensitivity to outliers. Therefore a robust approach should be used. Several techniques have been proposed [162, 186] based on robust statistics [136]. Our system incorporates the RANSAC (RANdom SAmpling Consensus) [40] approach used by Torr et al. [162, 163]. Table 7.1 sketches this technique.

Once the epipolar geometry has been retrieved, one can start looking for more matches to refine this geometry. In this case the search region is restricted to a few pixels around the epipolar lines.

The results of the procedure described here are shown in Figure 7.3 and 7.4. The top row shows the corners extracted in the two first images of the Arenberg castle sequence. The bottom row shows the points which were matched using the robust epipolar geometry computation method.

7.3.2 Initial reconstruction

The two first images of the sequence are used to determine a reference frame. The world frame is aligned with the first camera. The second camera is chosen so that the epipolar geometry corresponds to the retrieved $\mathbf{F}_{12}$ (see [98]).

$$\begin{array}{lcl} \mathbf{P}_1 & = & [\ \mathbf{I}_{3\times 3} \mid 0_3\ ] \\ \mathbf{P}_2 & = & [\ [\mathsf{e}_{12}]_\times \mathbf{F}_{12} + \mathsf{e}_{12}\pi^\top \mid \sigma \mathsf{e}_{12}\ ] \end{array} \tag{7.1}$$

where $[\mathsf{e}_{12}]_\times$ indicates the vector product with e_{12}. Equation 7.1 is not completely determined by the epipolar geometry (i.e. $\mathbf{F}_{12}$ and e_{12}), but has 4 more degrees of

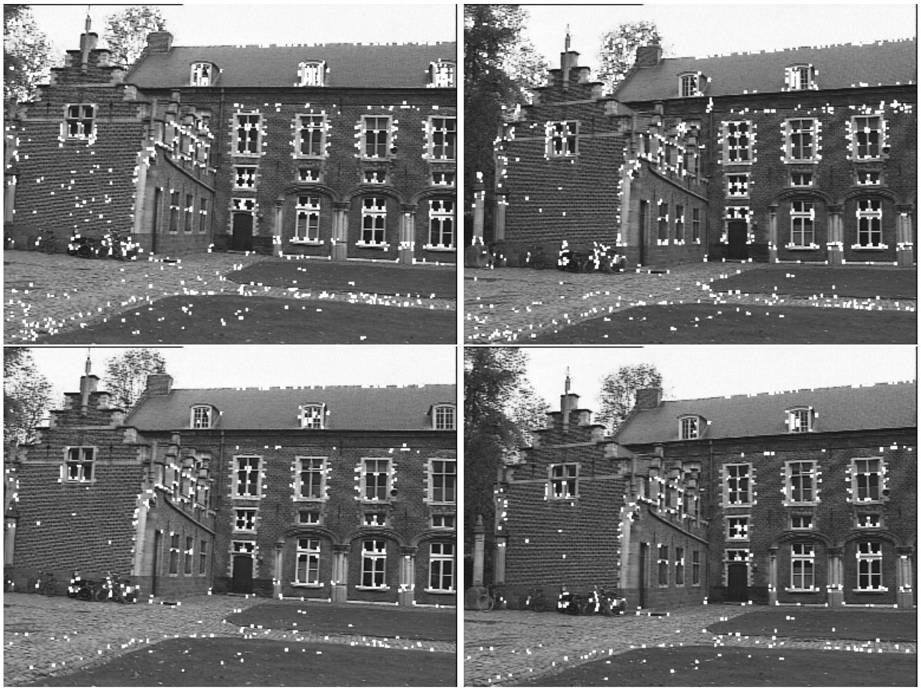

Figure 7.3: *The two images of the top row shows the interest points which are extracted in the first two images of the castle sequence. The bottom row shows the matches which were obtained after carrying out a robust epipolar geometry computation. Note that in the bottom row most points in the trees and in parts which were only visible in one image have disappeared.*

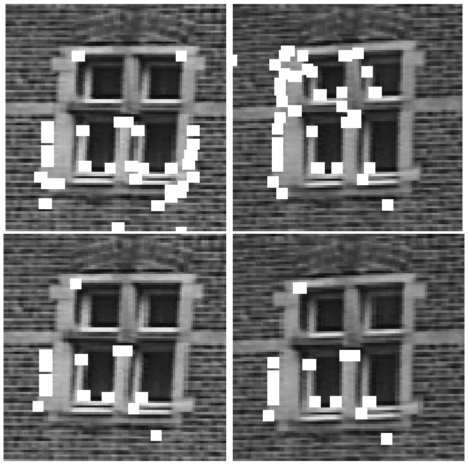

Figure 7.4: *Details extracted from Figure 7.3. The top row gives all the extracted interest points while the bottom row gives the matched ones.*

freedom (i.e. π and σ). π determines the position of the plane at infinity and σ determines the global scale of the reconstruction. To avoid some problems during the reconstruction it is recommended to determine l_X, l_Y, l_Z in such a way that the plane at infinity does not cross the scene. Our implementation uses an approach similar to the quasi-Euclidean approach proposed in [9], but the focal length is chosen so that most of the points are reconstructed in front of the cameras[1]. This approach was inspired by Hartley's cheirality [53] and the oriented projective geometry introduced by Laveau [79]. Since there is no way to determine the global scale from the images, σ can arbitrarily be chosen to $\sigma = 1$.

Once the cameras have been fully determined the matches can be reconstructed through triangulation. The optimal method for this is given in [58]. This gives us a preliminary reconstruction.

7.3.3 Adding a view

For every additional view the pose towards the pre-existing reconstruction is determined, then the reconstruction is updated. This is illustrated in Figure 7.5. The first step consists of finding the epipolar geometry as described in Section 7.3.1. Then the matches which correspond to already reconstructed points are used to compute the projection matrix $\mathbf{P}_k$. This is done using a robust procedure similar to the one laid

---

[1] The quasi-Euclidean approach computes the plane at infinity based on an approximate calibration. Although this can be assumed for most intrinsic parameters, this is not the case for the focal length. Several values of the focal length are tried out and for each of them the algorithm computes the ratio of reconstructed points that are in front of the camera. If the computed plane at infinity –based on a wrong estimate of the focal length– passes through the object, then many points will end up behind the cameras. This procedure allows us to obtain a rough estimate of the focal length for the initial views.

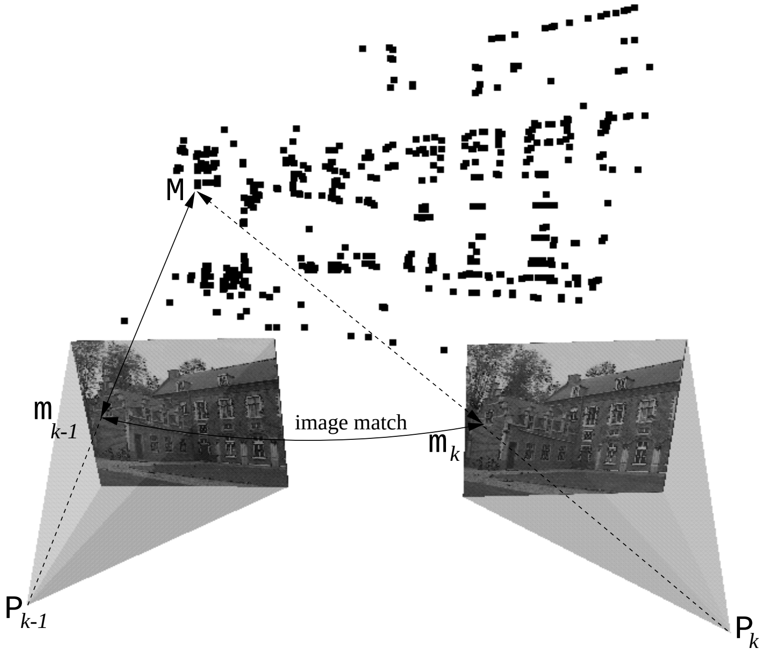

Figure 7.5: *Image matches ($\mathtt{m}_{k-1}, \mathtt{m}_k$) are found as described before. Since the image points, $\mathtt{m}_{k-1}$, relate to object points, $\mathtt{M}_k$, the pose for view k can be computed from the inferred matches ($\mathtt{M}, \mathtt{m}_k$).*

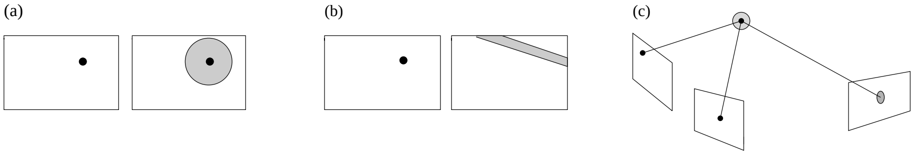

Figure 7.6: *(a) a priori search range, (b) search range along the epipolar line and (c) search range around the predicted position of the point.*

out in Table 7.1. In this case a minimal sample of 6 matches is needed to compute $\mathbf{P}_k$. Once $\mathbf{P}_k$ has been determined the projection of already reconstructed points can be predicted. This allows to find some additional matches to refine the estimation of $\mathbf{P}_k$. This means that the search space is gradually reduced from the full image to the epipolar line to the predicted projection of the point. This is illustrated in Figure 7.6.

Once the camera projection matrix has been determined the reconstruction is updated. This consists of refining, correcting or deleting already reconstructed points and initializing new points for new matches.

After this procedure has been repeated for all the images, one disposes of camera poses for all the views and the reconstruction of the interest points. In the further modules mainly the camera calibration is used. The reconstruction itself is used to obtain an estimate of the disparity range for the dense stereo matching.

7.3.4 Relating to other views

The procedure to add a view described in the previous section only relates the image to the previous image. In fact it is implicitly assumed that once a point gets out of sight, it will not come back. Although this is true for many sequences, this assumptions does not always hold. Assume that a specific 3D point got out of sight, but that it is visible again in the last two views. In this case a new 3D point will be instantiated. This will not immediately cause problems, but since for the system these two 3D points are unrelated nothing enforces their position to correspond.

This is especially crucial for longer image sequences where the errors accumulate. It results in a degraded calibration or even causes the failure of the algorithm after a certain number of views.

A possible solution consists of relating every new view with all previous views using the procedure of Section 7.3.1. It is clear that this would require a considerable computational effort. We propose a more pragmatic approach. This approach worked well on the cases we encountered (see Section 8.4).

Let $\tilde{\mathbf{P}}_i$ be the initial estimate of the camera pose obtained as described in the previous section. A criterion is then used to define which views are close to the actual view. All these close views are matched with the actual view (as described in Section 7.3.1). For every close view a set of potential 2D-3D correspondences is obtained. These sets are merged and the camera projection matrix $\mathbf{P}_i$ is reestimated using the same robust procedure as described in the previous section. Figure 7.7 illustrates this approach.

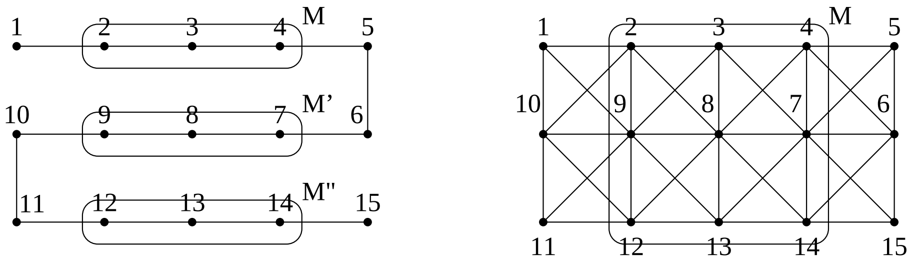

Figure 7.7: *Sequential approach (left) and extended approach (right). In the traditional scheme view 8 would be matched with view 7 and 9 only. A point* `M` *which would be visible in views 2,3,4,7,8,9,12,13 and 14 would therefore result in 3 independently reconstructed points. With the extended approach only one point will be instantiated. It is clear that this results in a higher accuracy for the reconstructed point while it also dramatically reduces the accumulation of calibration errors.*

We applied a very simple criterion to decide if views were close or not. It worked well for the applications we had in mind, but it could easily be refined if needed. The position $\mathtt{t}_i$ of the camera is extracted from $\tilde{\mathbf{P}}_i$ and the distance d_{ij} to all the other camera positions is computed. Close views are selected as views for which $d_{ij} < 1.6 d_{i(i-1)}$. Note that strictly speaking such measure is meaningless in projective space, but since a quasi-Euclidean initialization was carried out and only local qualitative comparisons are made the obtained results are good.

7.4 Upgrading the reconstruction to metric

The reconstruction obtained as described in the previous paragraph is only determined up to an arbitrary projective transformation. This might be sufficient for some robotics or inspection applications, but certainly not for visualization.

The self-calibration methods described in the previous chapters can be used to upgrade the projective reconstruction to metric. In the case of a camera with constant intrinsic parameters it is proposed to use the technique described in Chapter 5. If the parameters can vary during the acquisition of the image sequence –due to the use of the zoom or by refocusing– it is advised to use the method of Chapter 6.

To apply this last method to standard zooming/focusing cameras, some assumptions should be made. Often it can be assumed that pixels are rectangular or even square. If necessary (e.g. when only a short image sequence is at hand, when the projective calibration is not accurate enough or when the motion sequence is close to critical [152] without additional constraints), it can also be used that the principal point is close to the center of the image. These assumptions are especially useful in an initialization stage and some can be relaxed in a refinement step.

The discussion of the previous chapters will not be repeated. Some reconstructions *before* and *after* the self-calibration stage are shown. Figure 7.8 gives the re-

Figure 7.8: Before self-calibration. *These different images illustrate the reconstruction before self-calibration. A top view (top-left), a general view (top right), a front view (bottom left) and a side view (bottom left) are given. Note the important skew on the reconstruction. It is clear that this kind of reconstruction can not be used for rendering new views.*

construction before self-calibration. Therefore it is only determined up to an arbitrary projective transformation and metric properties of the scene can not be observed from this representation. Figure 7.9 shows the result after self-calibration. At this point the reconstruction has been upgraded to metric.

7.5 Dense depth estimation

Only a few scene points are reconstructed from feature tracking. Obtaining a dense reconstruction could be achieved by interpolation, but in practice this does not yield satisfactory results. Small surface details would never be reconstructed in this way. Additionally, some important features are often missed during the corner matching and would therefore not appear in the reconstruction.

These problems can be avoided by using algorithms which estimate correspondences for almost every point in the images. Because the reconstruction was upgraded to metric, algorithms which were developed for calibrated stereo rigs can be used.

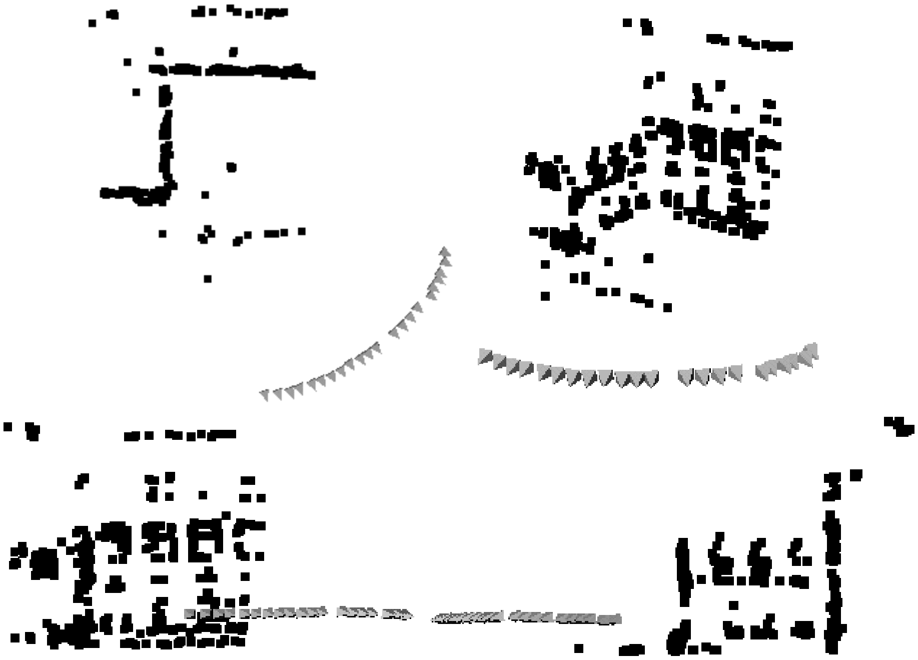

Figure 7.9: After self-calibration. *These images illustrate the reconstruction after the self-calibration step. Here also a top view (top-left), a general view (top right), a front view (bottom left) and a side view (bottom left) are given. In this case the ambiguity on the reconstruction has been restricted to metric. Notice that parallelism, orthogonality and other constraints can now be verified.*

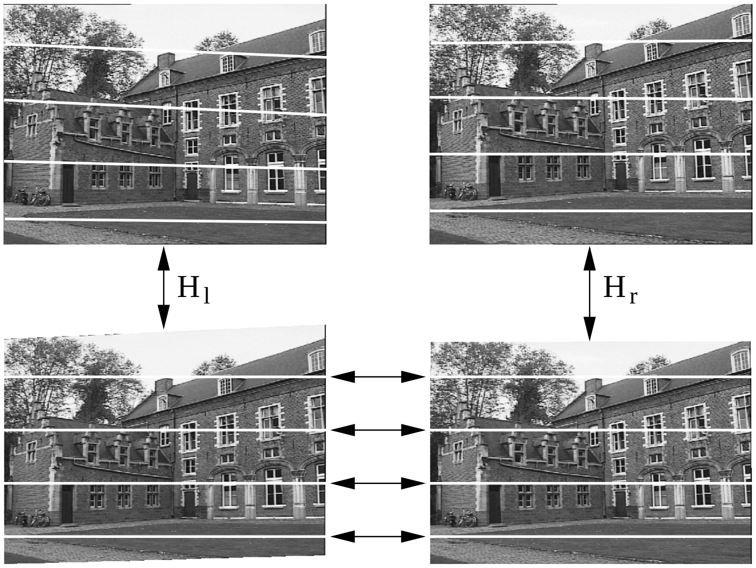

Figure 7.10: *Through the rectification process the image scan lines are brought into epipolar correspondence. This allows important gains in computational efficiency and simplification of the dense stereo matching algorithm.*

7.5.1 Rectification

Since we have computed the calibration between successive image pairs we can exploit the epipolar constraint that restricts the correspondence search to a 1-D search range. It is possible to re-map the image pair to standard geometry with the epipolar lines coinciding with the image scan lines [72]. The correspondence search is then reduced to a matching of the image points along each image scan-line. This results in a dramatic increase of the computational efficiency of the algorithms by enabling several optimizations in the computations. The rectification procedure is illustrated in Figure 7.10. For some motions (i.e. when the epipole is located in the image) standard rectification based on planar homographies is not possible and a more advanced procedure should be used (see Section 7.8.4).

7.5.2 Dense stereo matching

In addition to the epipolar geometry other constraints like preserving the order of neighboring pixels, bidirectional uniqueness of the match, and detection of occlusions can be exploited. These constraints are used to guide the correspondence towards the most probable scan-line match using a dynamic programming scheme [30]. These concepts are illustrated in Figure 7.11.

For dense correspondence matching a disparity estimator based on the dynamic programming scheme of Cox et al. [20], is employed that incorporates the above men-

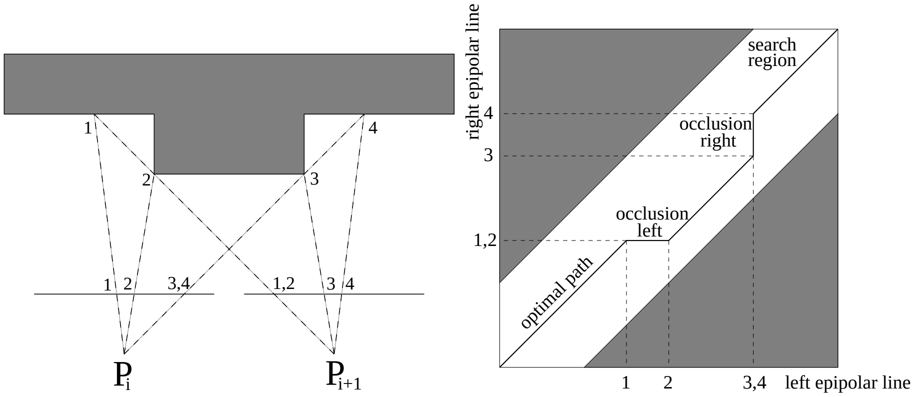

Figure 7.11: *Illustration of the ordering constraint (left), Dense matching as a path search problem (right).*

tioned constraints. It operates on rectified image pairs (I_k, I_l) where the epipolar lines coincide with image scan lines. The matcher searches at each pixel in image I_k for maximum normalized cross correlation in I_l by shifting a small measurement window (kernel size 5x5 to 7x7 pixels) along the corresponding scan line. The selected search step size ΔD (usually 1 pixel) determines the search resolution. Matching ambiguities are resolved by exploiting the ordering constraint in the dynamic programming approach (see Koch [72]). The algorithm was further adapted to employ extended neighborhood relationships and a pyramidal estimation scheme to reliably deal with very large disparity ranges of over 50% of image size (see Falkenhagen [31, 30]). This algorithm that was at first developed for calibrated stereo rigs (See Koch [72]) could easily be used for our purposes since at this stage the necessary calibration information had already been retrieved from the images.

7.5.3 Multi view matching

The pairwise disparity estimation allows to compute image to image correspondence between adjacent rectified image pairs, and independent depth estimates for each camera viewpoint. An optimal joint estimate is achieved by fusing all independent estimates into a common 3D model. The fusion can be performed in an economical way through controlled correspondence linking. The approach utilizes a flexible multi viewpoint scheme which combines the advantages of small baseline and wide baseline stereo (see Koch, Pollefeys and Van Gool [73]).

Assume an image sequence with $i = 1 \rightarrow n$ images. Starting from a reference viewpoint i the correspondences between adjacent images $(i+1, i+2, ..., n)$ and $(i-1, i-2, ..., 1)$ are linked in a chain. The depth for each reference image point m_i is computed from the correspondence linking that delivers two lists of image correspondences relative to the reference, one linking down from $i \rightarrow 1$ and one linking up from $i \rightarrow n$. For each valid corresponding point pair $(\mathtt{m}_i, \mathtt{m}_j)$ we can triangu-

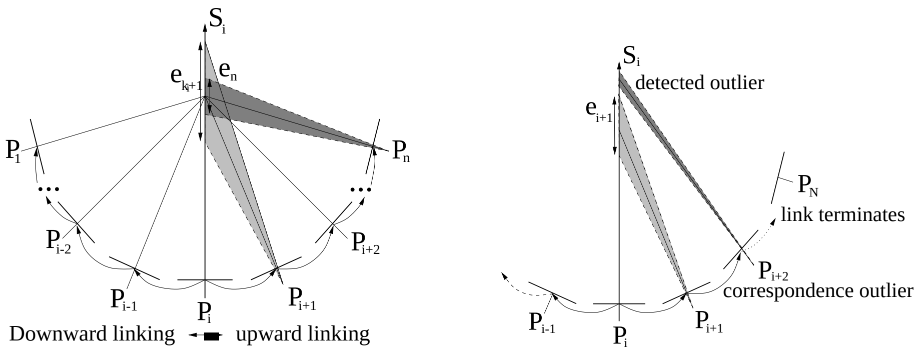

Figure 7.12: *Depth fusion and uncertainty reduction from correspondence linking (left), linking stops when an outlier is encountered (right).*

late a depth estimate $d(x_i, x_j)$ along S_{m_i} with e_j representing the depth uncertainty. Figure 7.12 visualizes the decreasing uncertainty interval during linking.

While the disparity measurement resolution ΔD in the image is kept constant (at 1 pixel), the reprojected depth error e_j decreases with the baseline. Outliers are detected by controlling the statistics of the depth estimate computed from the correspondences. All depth values that fall within the uncertainty interval around the mean depth estimate are treated as inliers. They are fused by a 1-D kalman filter to obtain an optimal mean depth estimate. Outliers are undetected correspondence failures and may be arbitrarily large. As threshold to detect the outliers we utilize the depth uncertainty interval e_j.

The result of this procedure is a very dense depth map. Most occlusion problems are avoided by linking correspondences from up and down the sequence. An example of such a very dense depth map is given in Figure 7.13.

7.6 Building the model

The dense depth maps as computed by the correspondence linking must be approximated by a 3D surface representation suitable for visualization. So far each object point was treated independently. To achieve spatial coherence for a connected surface, the depth map is spatially interpolated using a parametric surface model. The boundaries of the objects to be modeled are computed through depth segmentation. In a first step, an object is defined as a connected region in space. Simple morphological filtering removes spurious and very small regions. Then a bounded thin plate model is employed with a second order spline to smooth the surface and to interpolate small surface gaps in regions that could not be measured.

The spatially smoothed surface is then approximated by a triangular wire-frame mesh to reduce geometric complexity and to tailor the model to the requirements of computer graphics visualization systems. The mesh triangulation currently utilizes the reference view only to build the model. The surface fusion from different view-

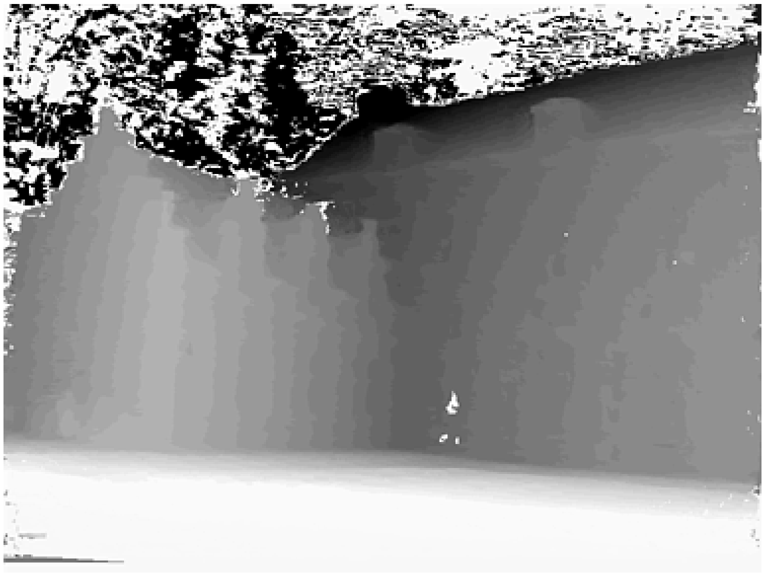

Figure 7.13: *Dense depth map (light means near and dark means far).*

points to completely close the models remains to be implemented. Sometimes it is not possible to obtain a single metric framework for large objects like buildings since one may not be able to record images continuously around it. In that case the different frameworks have to be registered to each other. This could be done using available surface registration schemes [18].

Texture mapping onto the wire-frame model greatly enhances the realism of the models. As texture map one could take the reference image texture alone and map it to the surface model. However, this creates a bias towards the selected image and imaging artifacts like sensor noise, unwanted specular reflections or the shading of the particular image is directly transformed onto the object. A better choice is to fuse the texture from the image sequence in much the same way as depth fusion.

The viewpoint linking builds a controlled chain of correspondences that can be used for texture enhancement as well. The estimation of a robust mean texture will capture the static object only and the artifacts (e.g. specular reflections or pedestrians passing in front of a building) are suppressed [67]. The texture fusion could also be done on a finer grid, yielding a super resolution texture [106, 19].

An example of the resulting model can be seen in Figure 7.14. More results are given in the next chapter.

7.7 Some implementation details

The system was mainly developed in TargetJr [157]. This is a software package consisting of a number of libraries. It contains an important number of computer vision

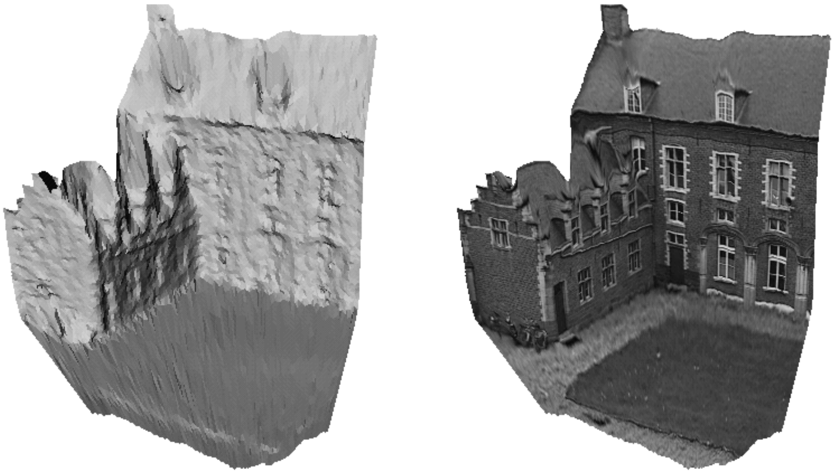

Figure 7.14: *3D surface model obtained automatically from an uncalibrated image sequence, shaded (left), textured (right).*

and image processing algorithms. In addition, TargetJr also offers routines to solve numerical problems, e.g. singular value decomposition or Levenbergh-Marquardt minimization.

It is not so easy to give an idea of the computation times needed to obtain a 3D model from an image sequence. This clearly depends on the number of images, the size of the images and on the available computing power. Some additional aspects also influence the computation times. The disparity range, for example, has an important effect on the time needed to compute the disparity map[2].

For images of 768×512 (i.e. PAL resolution) processed on a Silicon Graphics O2 workstation (R10000@195Mhz) the following figures give an idea of typical computataion times. Around 30 seconds are needed per image for the projective reconstruction. This consists of interest point extraction, epipolar geometry computation and matching, projection matrix computation and adaptation of the projective structure. Upgrading the structure to metric through self-calibration only requires a few seconds for a sequence of 10 images.

The dense depth estimation requires more time since in this case computations have to be carried out on every pixel. The actual implementation of the rectification requires about 30 seconds per pair of views. The dense matching requires around 5 minutes per pair (depending on the disparity range). Computing a dense depth map from multiple views takes around 3 minutes for 10 views (starting from the pairwise disparity maps). Generating a 3D mesh from a dense depth maps takes around 1 minute.

In conclusion, generating a dense 3D model from 10 views takes about 1 hour of computing time.

---

[2] This can be understood from Figure 7.11

It should be noticed, however, that the system was not optimized towards speed. Several possibilities exist to reduce the amount of computations. In addition, most of the computations can be carried out in parallel (e.g. one processor per image).

7.8 Some possible improvements

7.8.1 Interest point matching

The procedure described in Section 7.3.1 only works when the two images under consideration are very similar. This requires some special care during the acquisition of the images. The problem lies with the similarity measure between corners (i.e. the normalized cross-correlation of the neighborhoods) and with the assumption that matching corners should have similar image coordinates. Dissimilarity is not only caused by a wide baseline (i.e. a camera displacement which is big compared to the camera-scene distance), but can also be caused by a different focus of attention or a different orientation of the camera for the two views. In the following paragraphs some specific possibilities for enhancement are proposed.

Global and local homographies Possible extensions to the standard procedure to deal with more different images have recently been proposed by Pritchett and Zisserman [132, 133]. This approach is briefly described here.

Large image differences induced by rotation of the camera can be cancelled through a global homography applied to the image (see Section 4.4.2 which handles pure rotations). It is proposed to use a hierarchical approach based on a Gaussian pyramid of images. An exhaustive search is carried out for the different parameters of the considered image transformations at the lowest resolution of the pyramid. At every resolution the parameters of the transformation are refined. One of the originals is replaced by the warped image which resembles the most the other image.

In the case of wide baseline (i.e. a translation of the order of the camera-scene distance) a global homography is not sufficient. Local homographies are instantiated based on groups of 4 lines which form parallelograms in the images. These are used to generate sets of potential matches (using a RANSAC approach as described in Table 7.1 for homographies instead of for the epipolar geometry). The different sets of matches are then combined to estimate the epipolar geometry.

Similarity measures An alternative to transform the image so that the differences are minimized consists of using more general similarity measures. Often normalized intensity cross-correlation is carried out to obtain potential correspondences. This similarity measure is only invariant to translations in the images and therefore degrades in the presence of other transformations. Therefore recently some more involved measures have been proposed. Schmid and Mohr [139] proposed a measure invariant to 2D Euclidean transformations in the context of recognition. Tuytelaars et al. [171] proposed a scheme based on affine invariants, with the crucial extension that it comes with affine invariant neighborhoods around the interest points.

Repeated structures Repeated structures in the scene can lead to matching ambiguities. This problem occurs typically in man-made environments (e.g. a building with identical windows). In this case it can happen that the robust matcher gets confused. Often different repetitions of the structure can not be differentiated. It is even possible that due to foreshortening the matcher is biased towards the wrong hypothesis. Since any hypothesis will have a lot of support due to the rest of the repeated structure, the robust corner matcher could very well not recover from it.

One possibility consists of allowing user interaction to correct for these mistakes. A better alternative could be based on the recent work of Schaffalitzky and Zisserman [138]. Their work is aimed at automatically detecting repeated structures in images. Once the repetition is discovered it can be dealt with at a higher level to disambiguate the matching. More in general, higher level matching could be used to bootstrap the matching.

7.8.2 Projective reconstruction

The actual method to obtain the projective reconstruction is not optimal. Although the quasi-euclidean strategy [8] works fine in most of the cases, it can happen that this approach fails. This happens when the algorithm is not able to obtain a suitable initialization from the first views. In this case it can not start at all or it fails after a number of images due to the proximity of the plane at infinity.

Recently, other strategies for projective reconstruction have been proposed. The hierarchical approach of Fitzgibbon and Zisserman [41], for example, does not depend on the choice of some specific frames for the initialization. In addition this method only relies on purely projective concepts (in 3D space). This results in a more robust approach to obtain the projective reconstruction.

The robustness of the 3D reconstruction system would be increased by incorporating a similar approach.

7.8.3 Surviving planar regions

An important problem of projective reconstruction is that the projection matrices need 3D structure to be uniquely determined. When at some point in an image sequence a dominant plane is encountered projective reconstruction will therefore fail. This is an important problem for which no satisfying solution has yet been proposed.

However, some important results have already been achieved to deal with this problem. Recently Torr, Fitzgibbon and Zisserman [164] proposed an interesting approach which detects these degenerate cases through robust model selection. The principle of model selection is based on the work of Kanatani [70]. This approach differentiates between three models: general (3D), planar and pure rotation. Since the correct model is used, the points can still be tracked reliably and the system is aware that the projection matrices can not be determined uniquely. A homography determines 8 degrees of freedom, while a projection matrix has 11 degrees of freedom, the problem being that the center of projection can not be determined.

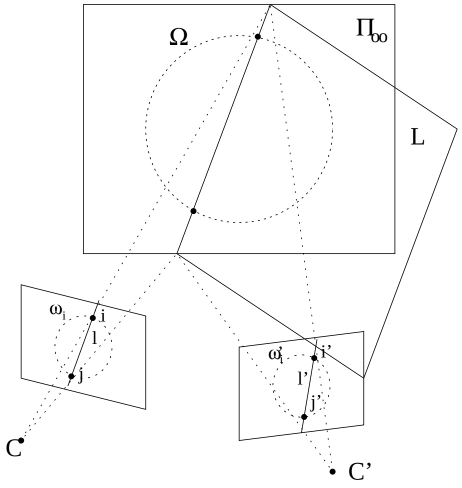

Figure 7.15: *The circular points of plane* Π *must always project on the image of the absolute conic. This constraint can be used to retrieve the absolute conic and the circular points. A minimum of 5 views is needed (*$5 + 2 \times 2 < 2n$*).*

On the other hand, when the camera is calibrated it is possible to determine uniquely the pose from the image of a plane. Therefore self-calibration could maybe yield a solution to this problem. If the part of the sequence before the occurrence of the degeneracy is long enough (and the motion is not critical) and the intrinsic camera parameters are constant, then the problem can easily be solved. This will not always be the case. In addition, it would be better to have a method which uses all available information.

Recently Triggs proposed an approach for calibration from planar scenes [167]. The idea consists of looking at the intersection of the plane with the absolute conic. The image of these two circular points should always be located on the image of the absolute conic. One can therefore look for two points which are systematically transferred on a certain conic. This conic is then the image of the absolute conic. This idea is illustrated in Figure 7.15

In our case however we are not interested in pure planar scenes. The goal is to uniquely relate the 3D structure obtained before and after the planar region. Combining Triggs' constraints for the planar region with the traditional constraints when 3D structure is seen, will give us a maximum of information on the metric calibration of the scene. Using this calibration the ambiguity on the reconstruction can be eliminated and both scenes can be integrated in the same metric basis.

A further refinement could be achieved by using guided matching based on the recovered structure. If two parts of the reconstruction obtained before and after the planar region end up (partially) superimposed, registration techniques could be used to refine the alignement of both parts.

7.8.4 Generalized rectification

In some cases the rectification technique based on homographies (see Section 7.5.1) is not possible. For an image pair with the epipoles in the images this rectification procedure can not be used since it transforms the epipoles at infinity. Recently Roy, Meunier and Cox [137] presented an alternative rectification technique. This method does not rectify the images to a plane but to a cylinder, thereby avoiding problems when the epipole is in the image. A second advantage of their technique is that pixel loss can be avoided without requiring huge images, which is difficult to ensure with the standard technique.

The method is however relatively complex. This is mainly due to the fact that the rectification process is carried out in 3D space. Epipolar lines are moved from the image plane to the cylindrical surface. This operation consists of a rotation, a change of coordinate system and a projection onto the cylinder. Although the authors claim that the method is readily applicable to uncalibrated systems it is not since the orientation of the epipolar lines is not verified. The result could be that the right half of one rectified image matches the left side of the other and vice versa. This would cause most stereo algorithms –which enforce ordering– to fail.

In this work we have developed a technique which avoids the above mentioned problems of both techniques. Although the technique is much simpler than the one described in [137] it can achieve even smaller images without any pixel loss. Our technique is purely image based and does not require more than an oriented fundamental matrix.

The idea is to use polar coordinates with the epipole as origin. Since the ambiguity for the location of matching points in a pair of images is restricted to half epipolar lines [78] we only have to use positive values for the longitudinal coordinate. Since (half-)epipolar line transfer is fully described by an oriented fundamental matrix, this is all our method needs. The necessary information is also easily extracted from (oriented) camera projection matrices. The angle between two consecutive half-epipolar lines is computed to have the worst case pixels preserve their area. This is done independently for every half-epipolar line. This therefore results in a minimal image size. In [137] the worst case pixel is computed for the whole image, which results in larger images. The details of our technique are given in appendix D.

As an example a rectified image pair from the castle is shown for both the standard technique and our new generalized technique. Figure 7.16 shows the original image pair and Figure 7.17 shows the rectified image pair for both methods.

A second example shows that the methods works properly when the epipole is in the image. Figure 7.18 shows the two original images while Figure 7.19 shows the two rectified images. In this case the standard rectification procedure can not deliver rectified images.

A stereo matching algorithm was used on this image pair to compute the disparities. The interpolated disparity map can also be seen in Figure 7.19. Figure 7.20 shows the depth map that was obtained and two views of the resulting 3D model. Note from these images that there is an important depth uncertainty around the epipole. In fact the epipole forms a singularity for the depth estimation. In the depth map an artefact

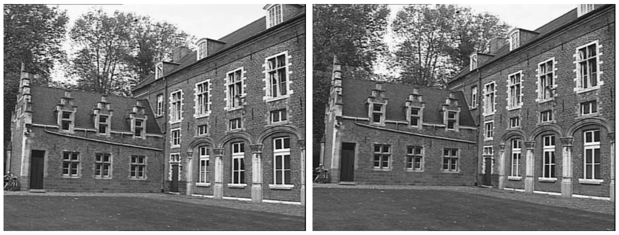

Figure 7.16: *Image pair from the castle sequence*

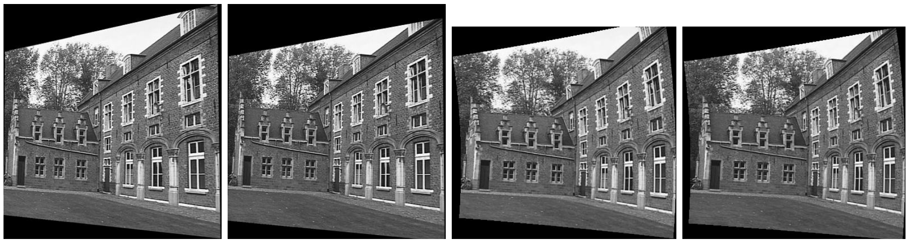

Figure 7.17: *Rectified image pair for both methods: standard homography based method (left), new method (right).*

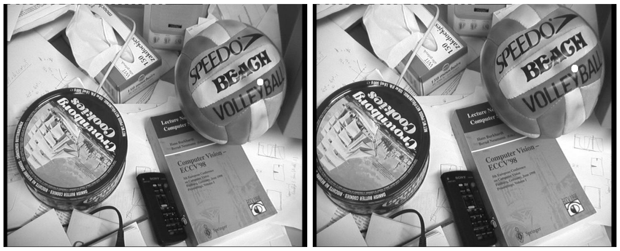

Figure 7.18: *Image pair of a desk a few days before a deadline. The epipole is indicated by a white dot (top-right of 'Y' in 'VOLLEYBALL').*

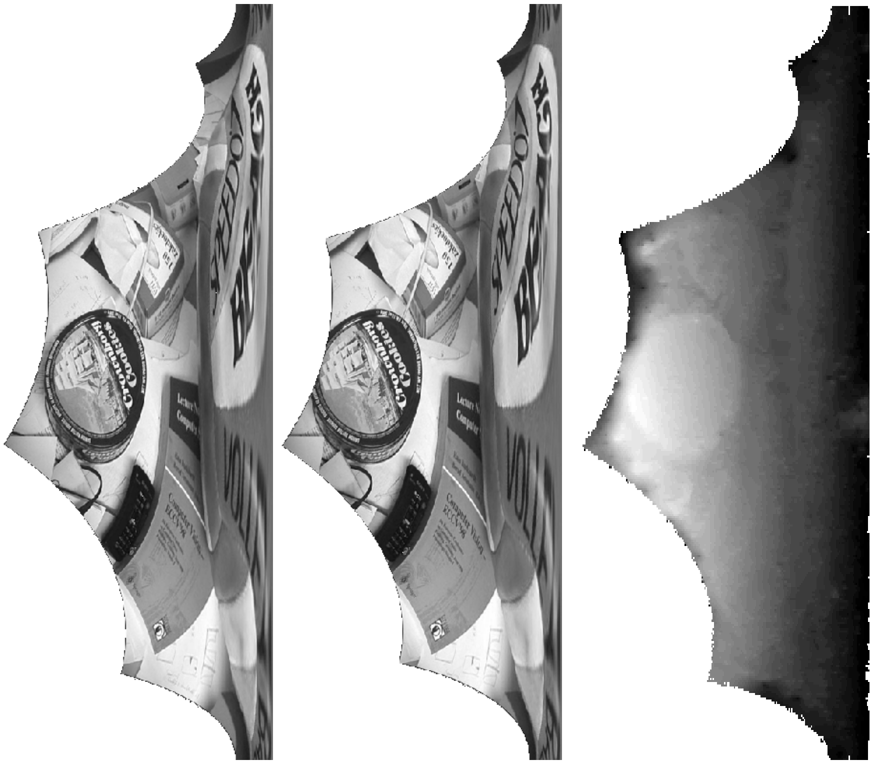

Figure 7.19: *Rectified pair of images of the desk (left and middle) and interpolated disparity map (right). It can be verified visually that corresponding points are located on corresponding image rows. The right side of the images corresponds to the epipole.*

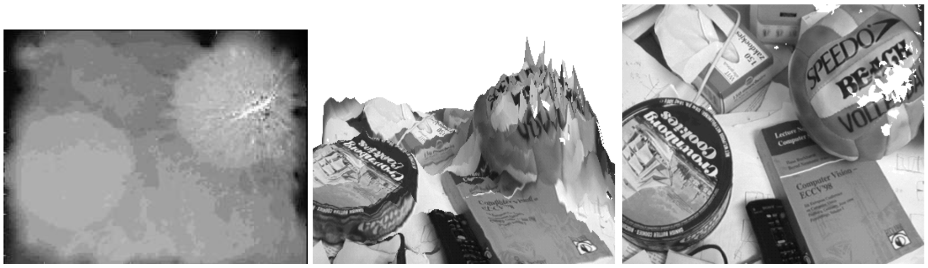

Figure 7.20: *Depth map for the far image of the desk image pair (left) and two views of the reconstruction obtained from the desk image pair (middle and right). The inaccuracy at the top of the volleyball is due to the depth ambiguity at the epipole.*

can be seen around the position of the epipole. The extent is much longer in one specific direction due to the matching ambiguity in this direction (see the original image or the middle-right part of the rectified image).

7.9 Conclusion

An automatic 3D scene modeling technique was discussed that is capable of building models from uncalibrated image sequences. The technique is able to extract metric 3D models without any prior knowledge about the scene or the camera. The calibration is obtained by assuming a rigid scene and some constraints on the intrinsic camera parameters (e.g. square pixels). Some possible improvements have been proposed concerning interest point matching, problems with dominant planar structures and rectification.

Work remains to be done to get more complete models by fusing the partial 3D reconstructions. This will also increase the accuracy of the models and eliminate artifacts at the occluding boundaries. For this we can rely on work already done for calibrated systems.

Chapter 8

Results and applications

8.1 Introduction

In this chapter we will focus on the results obtained by the system described in the previous chapter. First some more results on 3D reconstruction from photographs are given. Then the flexibility of our approach is shown by reconstructing an amphitheater from old film footage. Next –using the extension of section 7.3.4– the application to the acquisition of plenoptic models is described. Finally, several applications in archaeology are discussed. The application of our system to the construction of a virtual copy of the archaeological site of Sagalassos (Turkey) –a *virtualized* Sagalassos– is described. Some more specific applications in the field of archaeology are also discussed.

8.2 Acquisition of 3D models from photographs

The main application for our system is the generation of 3D models from images. One of the simplest methods to obtain a 3D model of a scene is therefore to use a photo camera and to shoot a few pictures of the scene from different viewpoints. Realistic 3D models can already be obtained with a restricted number of images. This is illustrated in this section with a detailed model of a part of a Jain temple in India.

A Jain Temple in Ranakpur

These images were taken during a tourist trip after ICCV'98 in India. A sequence of images was taken of a highly decorated part of one of the smaller Jain temples at Ranakpur, India. These images were taken with a standard Nikon F50 photo camera and then scanned. All the images which were used for the reconstruction can be seen in Figure 8.1. Figure 8.2 shows the reconstructed interest points together with the estimated pose and calibration of the camera for the different viewpoints. Note that only 5 images were used and that the global change in viewpoint between these

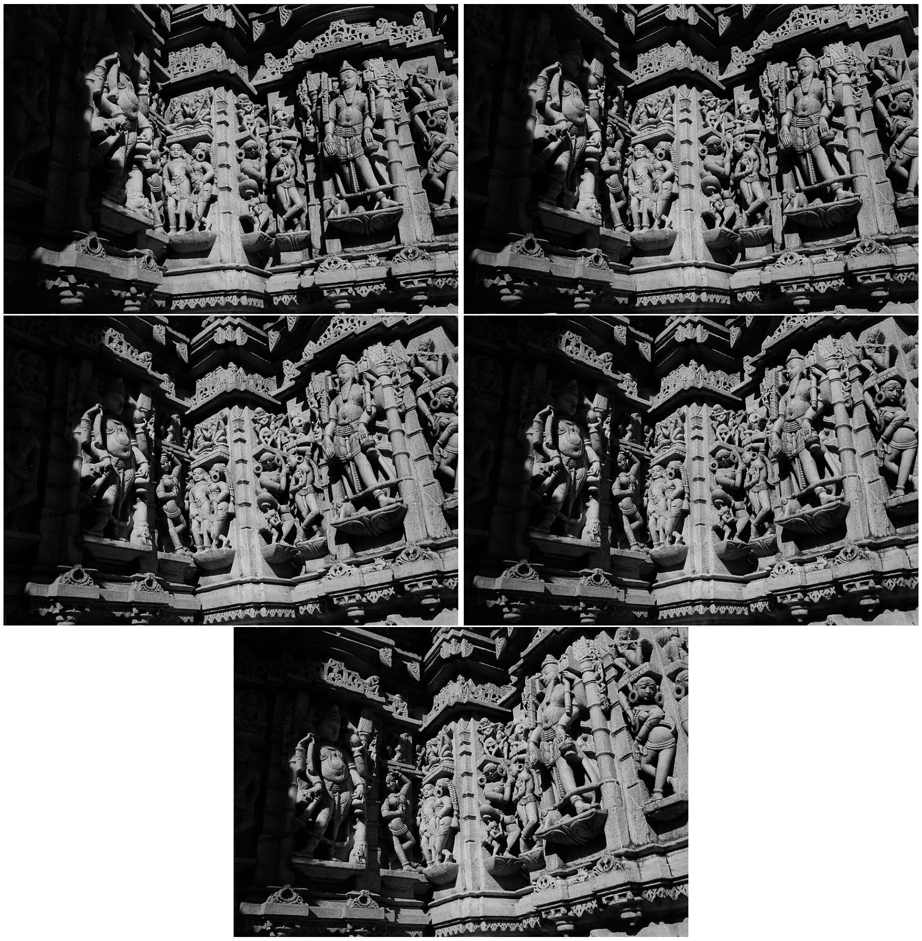

Figure 8.1: *Photographs which were used to generate a 3D model of a detail of a Jain temple of Ranakpur.*

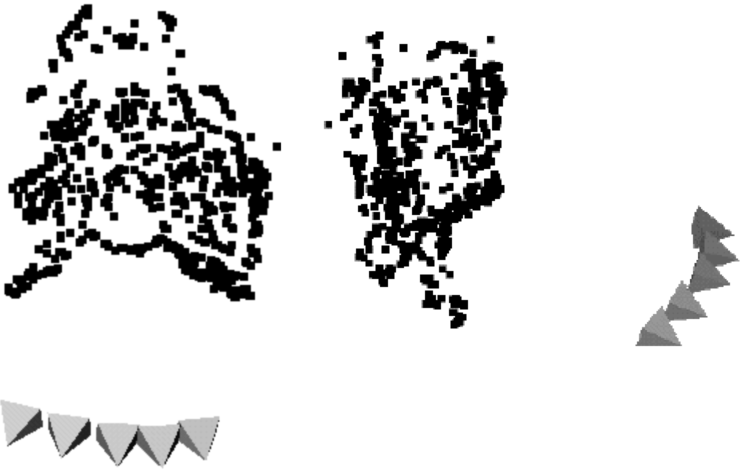

Figure 8.2: *Reconstruction of interest points and cameras. The system could automatically reconstruct a realistic 3D model of this complex scene without any additional information.*

different images is relatively small. In Figure 8.3 a global view of the reconstruction is given. In the lower part of the image the texture has been left out so that the recovered geometry is visible. Note the recovered shape of the statues and details of the temple wall. In Figure 8.4 two detail views from very different angles are given. The visual quality of these images is still very high. This shows that the recovered models allow to extrapolate viewpoints to some extent. Since it is difficult to give an impression of 3D shape through images we have put three views of the same part –but slightly rotated each time– in Figure 8.5. This reconstruction shows that the proposed approach is able to recover realistic 3D models of complex shapes. To achieve this no calibration nor prior knowledge about the scene was required.

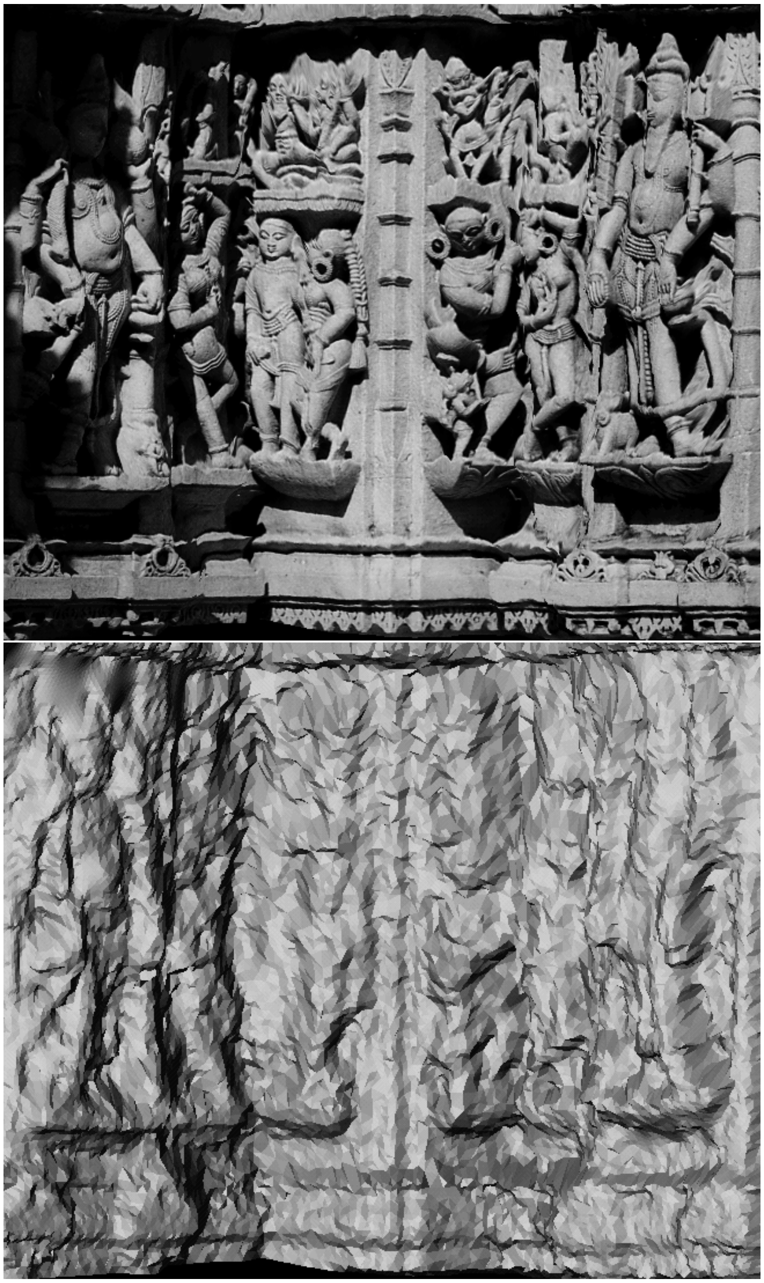

Figure 8.3: *Reconstruction of a part of a Jain temple in Ranakpur (India). Both textured (top) and shaded (bottom) views are given to give an impression of the visual quality and the details of the recovered shape.*

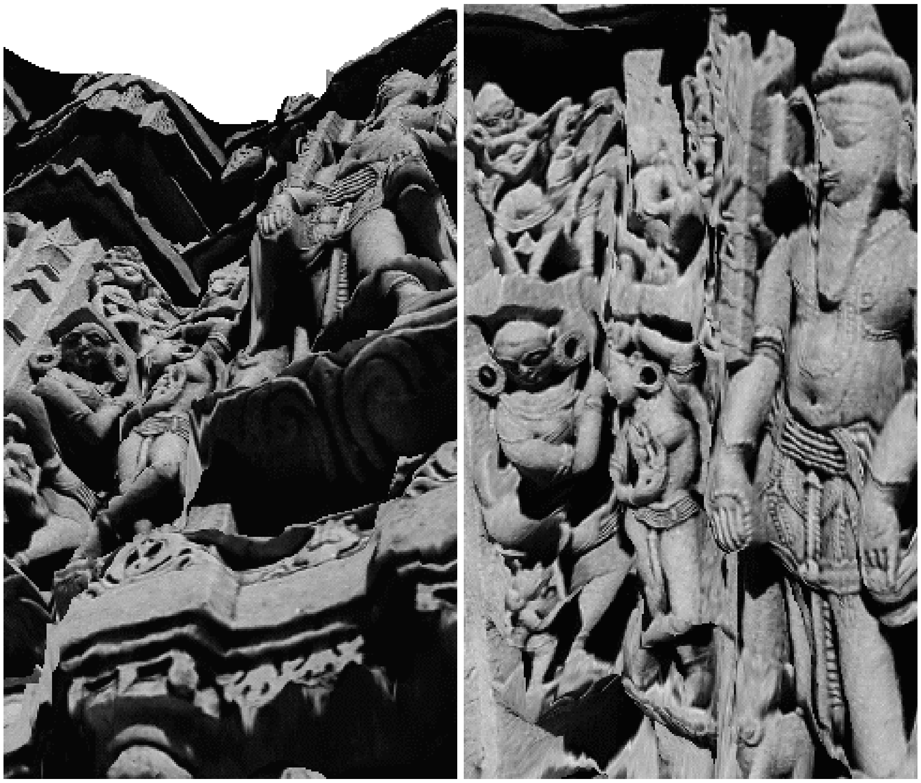

Figure 8.4: *Two detail views of the reconstructed model.*

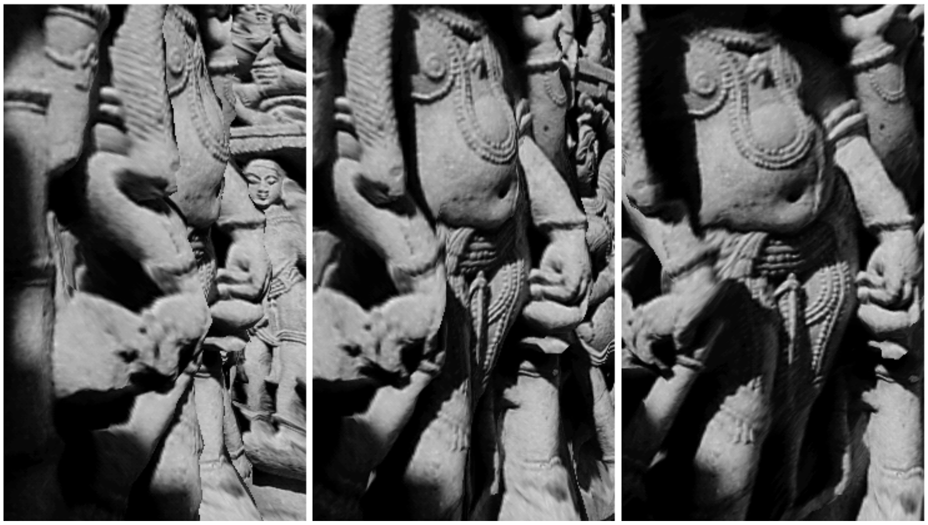

Figure 8.5: *Three rotated views of a detail of the reconstructed model.*

8.3 Acquisition of 3D models from preexisting image sequences

Here the reconstruction of the ancient theater of Sagalassos is shown. Sagalassos is an archaeological site in Turkey. More results obtained at this site are presented in Sections 8.5 and 8.6. The reconstruction is based on a sequence filmed by a cameraman from the BRTN (Belgische Radio en Televisie van de Nederlandstalige gemeenschap) in 1990. The sequence was filmed to illustrate a TV program about Sagalassos. Because of the motion only fields –and not frames– could be used. The resolution of the images we could use was thus restricted to 768×288. The sequence consisted of about hundred images, every tenth image is shown in Figure 8.6. We recorded approximately 3 images per second.

In Figure 8.7 the reconstruction of interest points and cameras is given. This shows that the approach can deal with long image sequences.

Dense depth maps were generated from this sequence and a dense textured 3D surface model was constructed from this. Some views of this model are given in Figure 8.8.

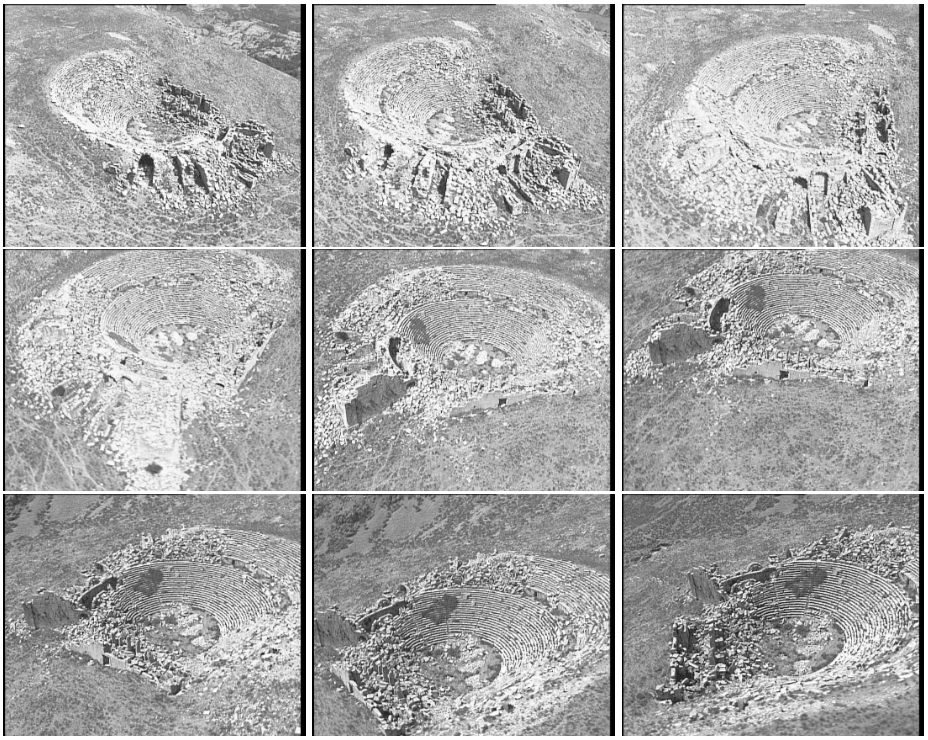

Figure 8.6: *This sequence was filmed from a helicopter in 1990 by a cameraman of the BRT (Belgische Radio en Televisie) to illustrate a TV program on Sagalassos (an archaeological site in Turkey).*

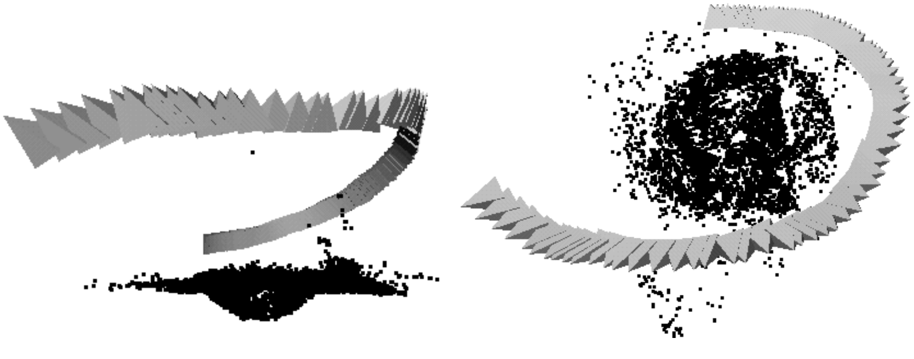

Figure 8.7: *The reconstructed interest points and camera poses recovered from the BRT sequence.*

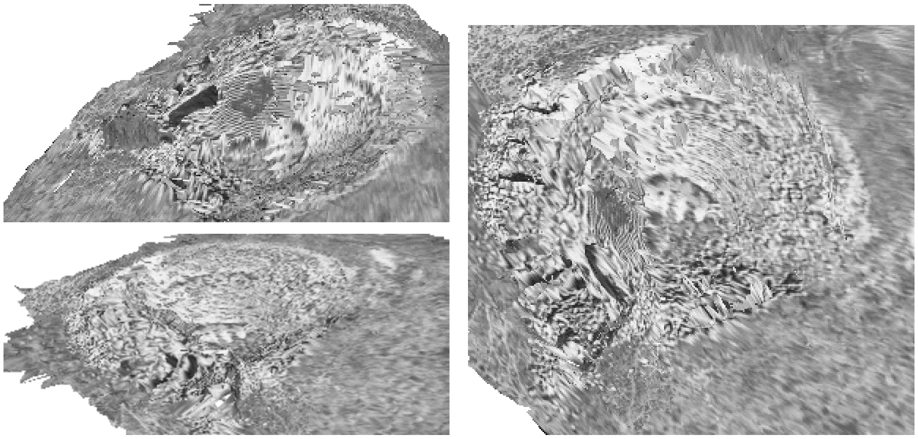

Figure 8.8: *Some views of the reconstructed model of the ancient theater of Sagalassos.*

8.4 Acquisition of plenoptic models

Image-based rendering approaches based on plenoptic modeling [93] have lately received a lot of attention, since they can capture the appearance of a 3D scene from images only, without the explicit use of 3D geometry. Thus one may be able to capture objects with very complex geometry that can not be modeled otherwise. Basically one caches all possible views of the scene and retrieves them during view rendering. Two recent approaches are the lightfield rendering approach [82] and the lumigraph [45].

The problem common to all approaches is the need to calibrate the camera sequence. Typically one uses calibrated camera rigs mounted on a special acquisition device like a robot [82], or a dedicated calibration pattern is used to facilitate calibration [45].

It is clear that in this case the work presented in this dissertation could offer some interesting possibilities. In the case of plenoptic from a hand-held camera, one typically generates very many (hundreds) of images, but with a specific distribution of the camera viewpoints. Since the goal is to capture the appearance of the object from all sides, one will try to sample the viewing sphere, thus generating a two-dimensional mesh of view points.

This application inspired the extension to the initial reconstruction procedure presented in Section 7.3.4. In case where not a 1D sequence is under consideration, but where a 2D sampling of the viewing sphere is executed, it is important to take advantage of the available information. Some preliminary trials with the sequential approach showed problems due to accumulation of errors over very long image sequences (several hundreds of images). This could even lead to the failure of the procedure in some cases. By extending the matching procedure to 2D the approach could take advantage of the resemblance between images for which the indices were far away.

The main problem of the sequential approach in this case is that points which are

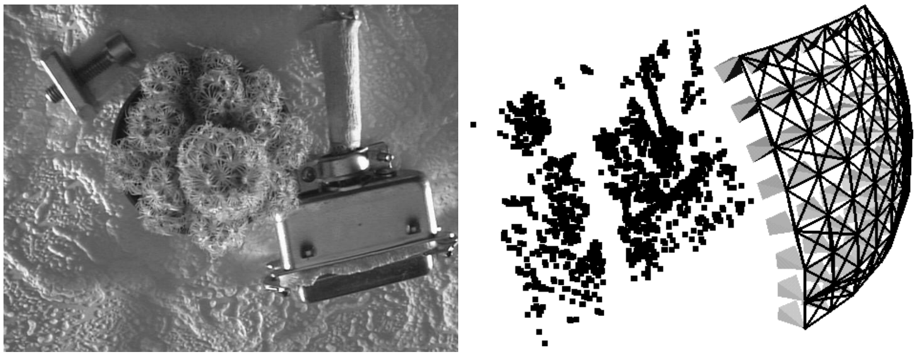

Figure 8.9: *Image of the* sphere *sequence (left) and result of calibration step (right). The cameras are represented by little pyramids. Images which were matched together are connected with a black line.*

visible in many images are getting instantiated over and over. Since these points are not connected, the different instances do not coincide, allowing systematic errors to accumulate. This problem will be illustrated in the following example.

Figure 8.9 shows one of the images of the *sphere* sequence and the recovered camera calibration together with the tracked points. This calibration can then be used to generate a plenoptic representation from the recorded images. Once this is obtained new views can be generated. Figure 8.10 shows all the images in which each 3D point is tracked. The points are in the order that they were instantiated. This explains the upper triangular structure. It is clear that for the sequential approach, even if some points can be tracked as far as 30 images, most are only seen in a few consecutive images. From the results for the extended approach several things can be noticed. The proposed method is clearly effective in the recovery of points which were not seen in the last images, thereby avoiding unnecessary instantiations of new points (the system only instantiated 2170 points instead of 3792 points). The band structure of the appearance matrix for the sequential approach has been replaced by a dense upper diagonal structure. Some points which were seen in the first images are still seen in the last one (more than 60 images further down the sequence). The mesh structure in the upper triangular part reflects the periodicity in the motion during acquisition. On the average, a point is tracked over 9.1 images instead of 4.8 images with the standard approach.

Where the previous sequence was still acquired with a robot arm, mainly to obtain ground truth, the following sequence was acquired with a hand-held camcorder. The complete sequence consists of 190 images. An image of the sequence can be seen in Figure 8.11. The recovered calibration together with the tracked points are also shown in this figure. The statistics of the points were computed for this sequence as well. They can be seen in Figure 8.12.

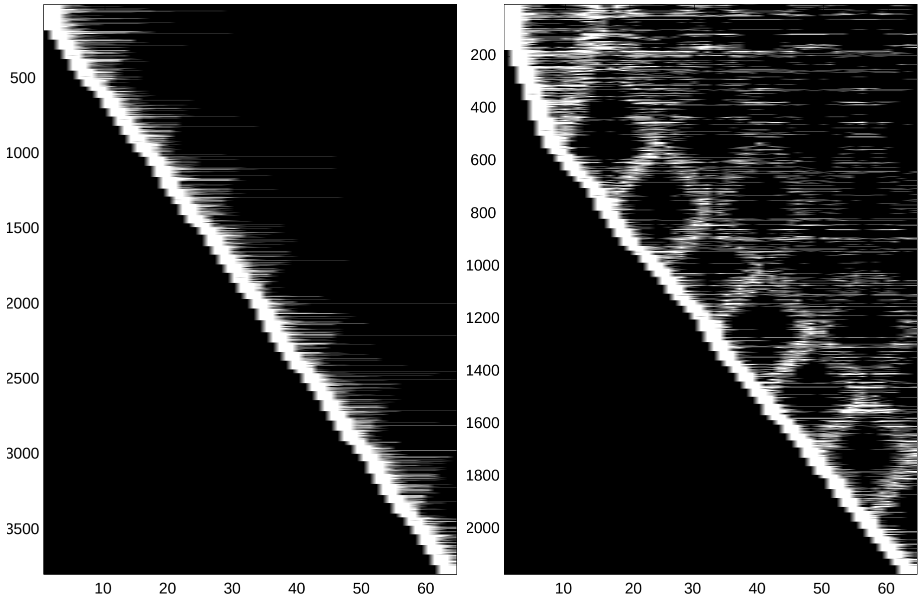

Figure 8.10: *Statistics of the* sphere *sequence. This figure indicates in which images a 3D point is seen. Points (vertical) versus images (horizontal). The results are illustrated for both the sequential approach (left) as the extended approach (right) are illustrated.*

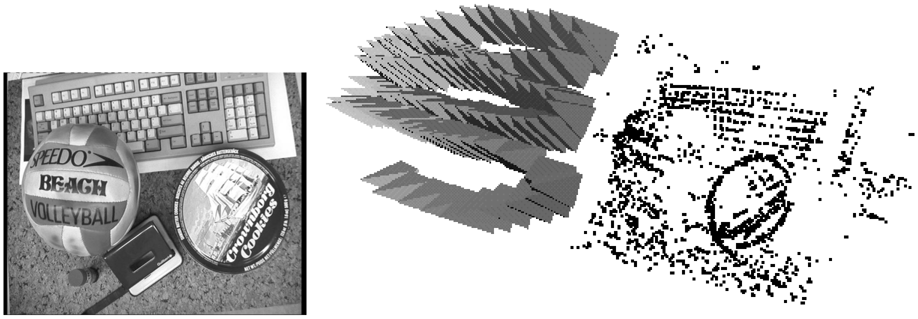

Figure 8.11: *Image of the* desk *sequence (left) and result of calibration step (right). The cameras are represented by little pyramids.*

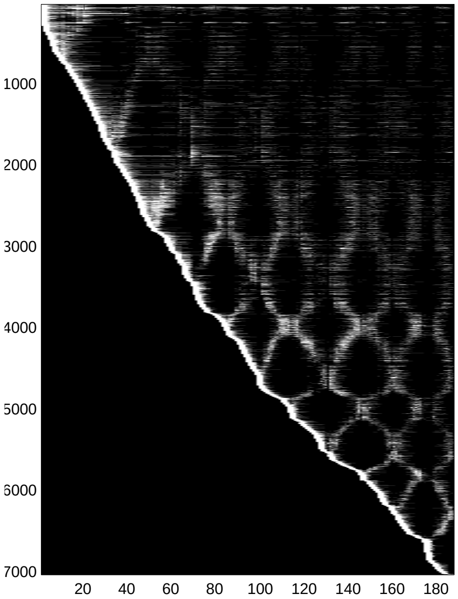

Figure 8.12: *Statistics for* desk *sequence. This figure indicates in which images a 3D point is seen. Points (vertical) versus images (horizontal).*

8.5 Virtualizing archaeological sites

Virtual reality is a technology that offers promising perspectives for archaeologists. It can help in many ways. New insights can be gained by immersion in ancient worlds, unaccessible sites can be made available to a global public, courses can be given "on-site" and different periods or building phases can coexist.

One of the main problems however is the generation of these virtual worlds. They require a huge amount of on-site measurements. In addition the whole site has to be reproduced manually with a CAD- or 3D modeling system. This requires a lot of time. Moreover it is difficult to model complex shapes and to take all the details into account. Obtaining realistic surface texture is also a critical issue. As a result walls are often approximated by planar surfaces, stones often all get the same texture, statues are only crudely modeled, small details are left out, etc.

An alternative approach consists of using images of the site. Some software tools exist, but require a lot of human interaction [109] or preliminary models [26]. Our system offers unique features in this context. The flexibility of acquisition can be very important for field measurements which are often required on archaeological sites. The fact that a simple photo camera can be sufficient for acquisition is an important advantage compared to methods based on theodolites or other expensive hardware. Especially in demanding weather conditions (e.g. dust, wind, heat, humidity).

The ancient site of Sagalassos (south-west Turkey) was used as a test case to illus-

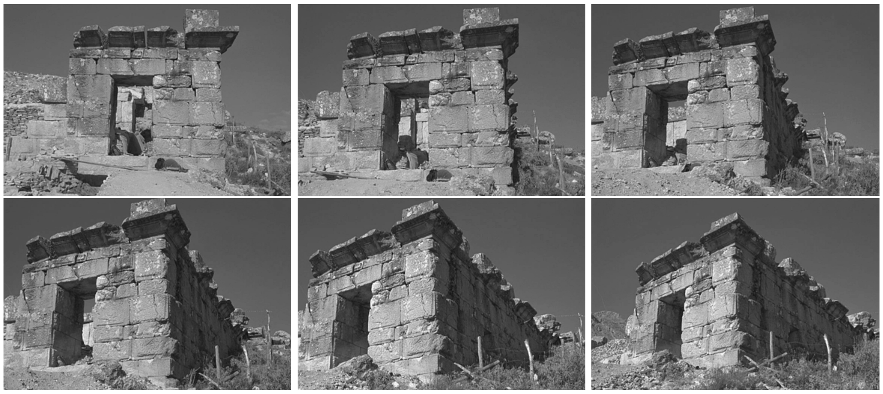

Figure 8.13: *Image sequence which was used to build a 3D model of the corner of the Roman baths*

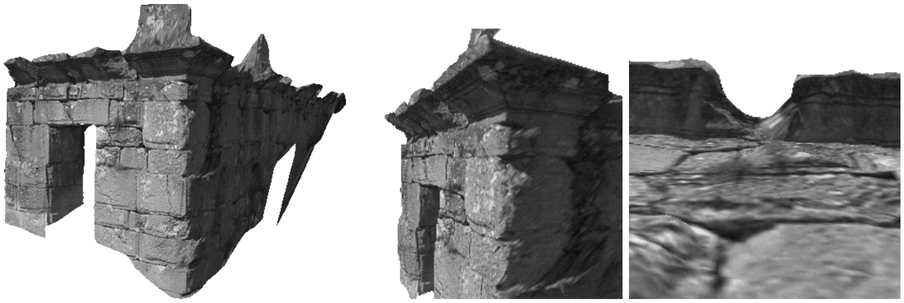

Figure 8.14: *Virtualized corner of the Roman baths, on the right some details are shown*

trate the potential of the approach developed in this work. The images were obtained with a consumer photo camera (digitized on photoCD) and with a consumer digital video camera.

8.5.1 Virtualizing scenes

The 3D surface acquisition technique that we have developed can readily be applied to archaeological sites. The on-site acquisition procedure consists of recording an image sequence of the scene that one desires to *virtualize*. To allow the algorithms to yield good results viewpoint changes between consecutive images should not exceed 5 to 10 degrees. An example of such a sequence is given in Figure 8.13. The result for the image sequence under consideration can be seen in Figure 8.14. An important advantage is that details like missing stones, not perfectly planar walls or symmetric structures are preserved. In addition the surface texture is directly extracted from the images. This does not only result in a much higher degree of realism, but is also

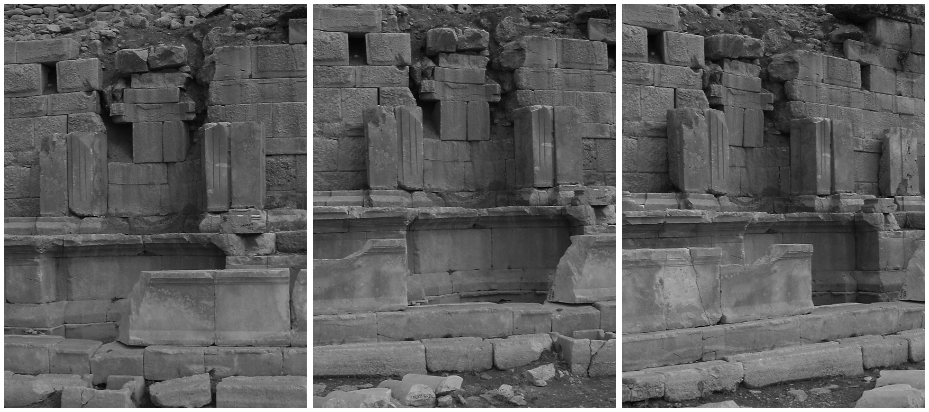

Figure 8.15: *Three of the six images of the Fountain sequence*

important for the authenticity of the reconstruction. Therefore the reconstructions obtained with this system could also be used as a scale model on which measurements can be carried out or as a tool for planning restorations.

As a second example, the reconstruction of the remains of an ancient fountain is shown. In Figure 8.15 three of the six images used for the reconstruction are shown. All images were taken from the same ground level. They were acquired with a digital camera with a resolution of approximately 1500x1000. Half resolution images were used for the computation of the shape. The texture was generated from the full resolution images.

The reconstruction can be seen in Figure 8.16, the left side shows a view with texture, the right view gives a shaded view of the model without texture. In Figure 8.17 two close-up shots of the model are shown.

8.5.2 Reconstructing an overview model

A first approach to obtain a virtual reality model for a whole site consists of taking a few overview photographs from the distance. Since our technique is independent of scale this yields an overview model of the whole site. The only difference is the distance needed between two camera poses. For most active techniques it is impossible to cope with scenes of this size. The use of a stereo rig would also be very hard since a baseline of several tens of meters would be required. Therefore one of the promising applications of the proposed technique is large scale terrain modeling.

In Figure 8.18 3 of the 9 images taken from a hillside near the excavation site are shown. These were used to generate the 3D surface model seen in Figure 8.19. In addition one can see from the right side of this figure that this model could be used to generate a Digital Terrain Map or an orthomap at low cost. In this case only 3 reference measurements –GPS and altitude– are necessary to localize and orient the model in the world reference frame.

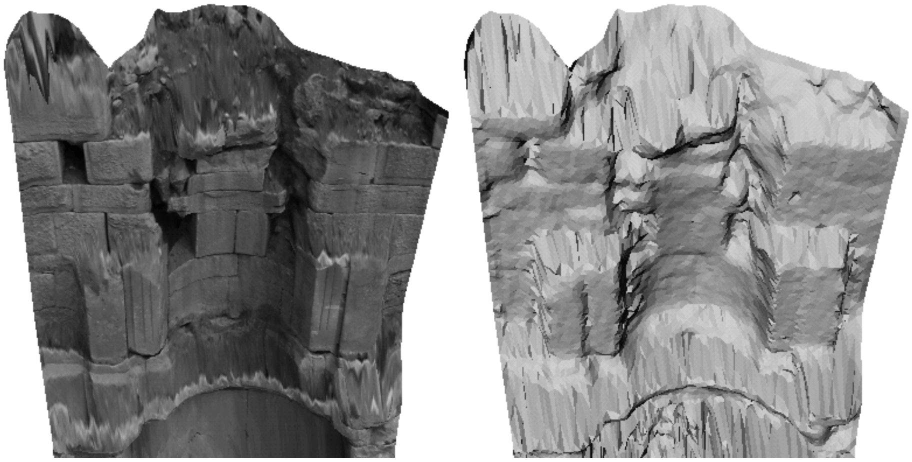

Figure 8.16: *Perspective views of the reconstructed fountain with and without texture*

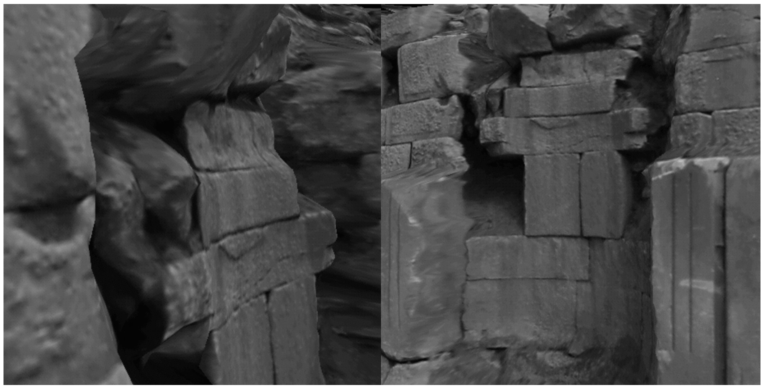

Figure 8.17: *Close-up views of some details of the reconstructed fountain*

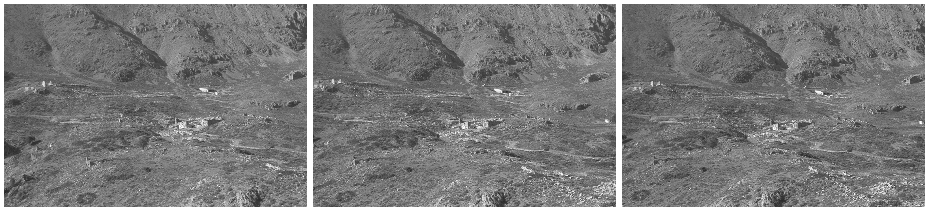

Figure 8.18: *Some of the images of the Sagalassos Site sequence*

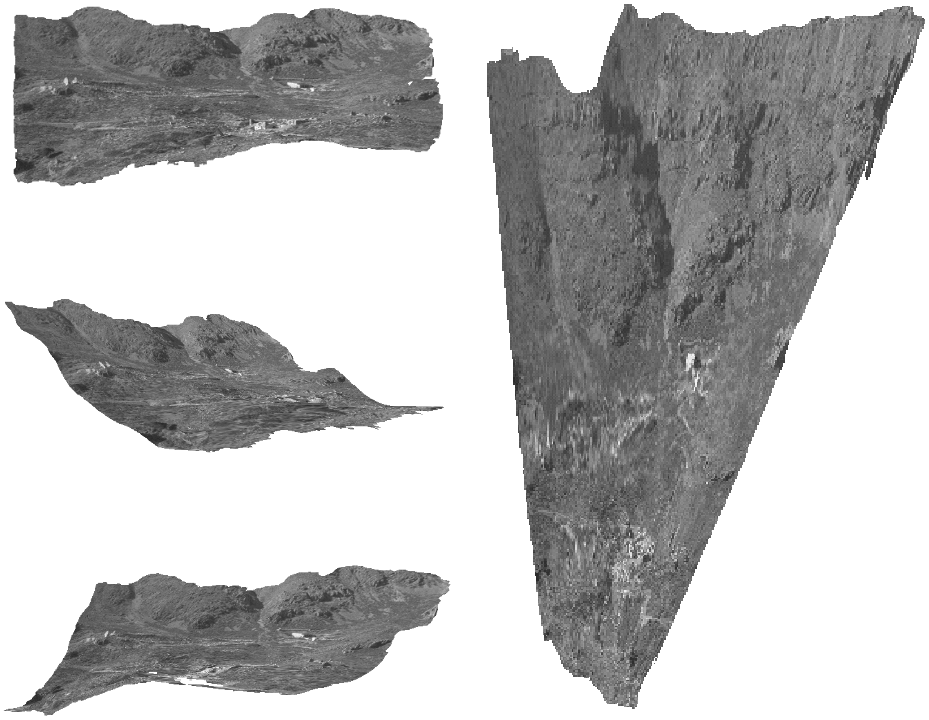

Figure 8.19: *Perspective views of the 3D reconstruction of the Sagalassos site (left). Top view of the reconstruction of the Sagalassos site (right).*

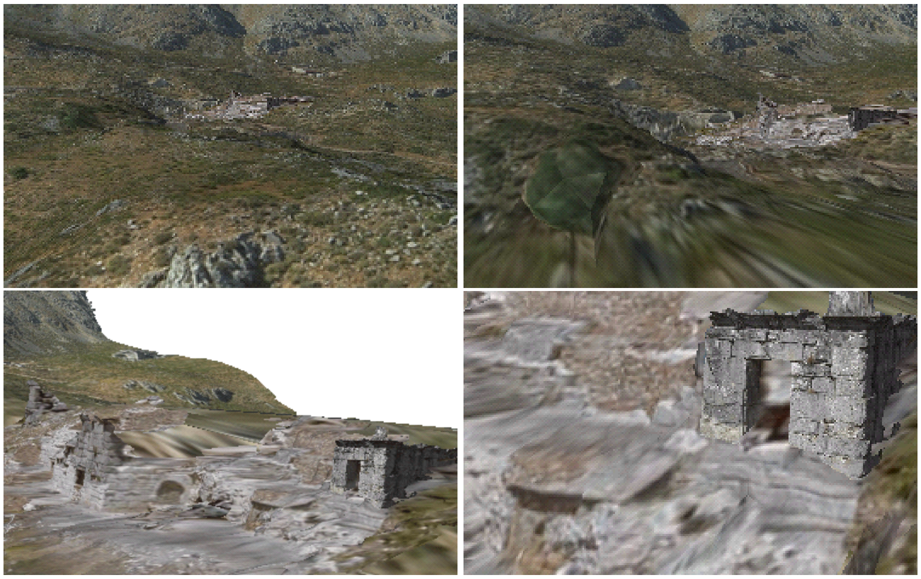

Figure 8.20: *Integration of models of different scales: site of Sagalassos, Roman baths and corner of the Roman baths.*

8.5.3 Reconstructions at different scales

The problem is that this kind of overview model is too coarse to be used for realistic walk-throughs around the site or for looking at specific monuments. Therefore it is necessary to integrate more detailed models into this overview model. This can be done by taking additional image sequences for all the interesting areas on the site. These are used to generate reconstructions of the site at different scales, going from a global reconstruction of the whole site to a detailed reconstruction for every monument.

These reconstructions thus naturally fill in the different levels of details which should be provided for optimal rendering. In Figure 8.20 an integrated reconstruction containing reconstructions at three different scales can be seen.

At this point the integration was done by interactively positioning the local reconstructions in the global 3D model. This is a cumbersome procedure since the 7 degrees of freedom of the similarity ambiguity have to be taken into account. Researchers are working on methods to automate this. Two different approaches are possible. The first approach is based on matching features which are based on both photometric and geometric properties, the second on minimizing a global alignment measure. A combination of both approaches will probably yield the best results.

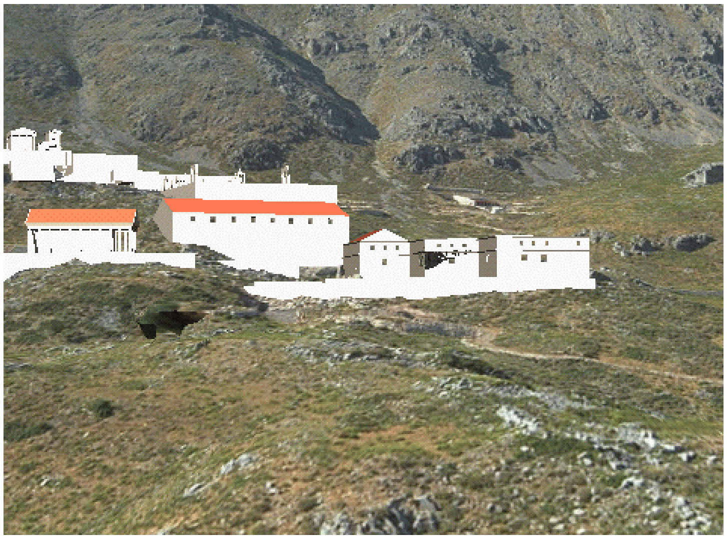

Figure 8.21: *Virtualized landscape of Sagalassos combined with CAD-models of reconstructed monuments*

8.5.4 Combination with other models

An interesting possibility is the combination of these models with other type of models. In the case of Sagalassos some building hypotheses were translated to CAD models. These were integrated with our models. The result can be seen in Figure 8.21. Also other models obtained with different 3D acquisition techniques could easily be integrated.

8.6 More applications in archaeology

Since these 3D models can be generated automatically and the on-site acquisition time is very short, several new applications come to mind. In this section a few possibilities are illustrated.

8.6.1 3D stratigraphy

Archaeology is one of the sciences were annotations and precise documentation are most important because evidence is destroyed during work. An important aspect of this is the stratigraphy. This reflects the different layers of soil that correspond to different time periods in an excavated sector. Due to practical limitations this stratigraphy is often only recorded for some slices, not for the whole sector.

Our technique allows a more optimal approach. For every layer a complete 3D model of the excavated sector can be generated. Since this only involves taking a series of pictures this does not slow down the progress of the archaeological work. In

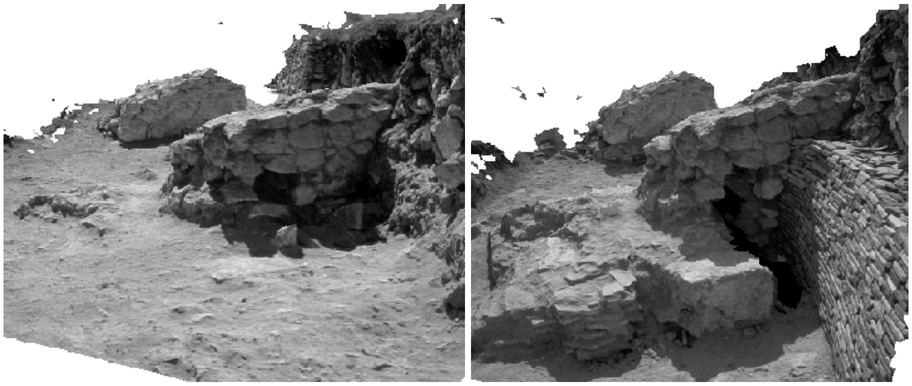

Figure 8.22: *3D stratigraphy, the excavation of a Roman villa at two different moments.*

addition it is possible to model artifacts separately which are found in these layers and to include the models in the final 3D stratigraphy.

This concept is illustrated in Figure 8.22. The excavations of an ancient Roman villa at Sagalassos were recorded with our technique. In the figure a view of the 3D model of the excavation is provided for two different layers.

8.6.2 Generating and testing building hypotheses

The technique also has a lot to offer for generating and testing building hypotheses. Due to the ease of acquisition and the obtained level of detail, one could reconstruct every building block separately. The different construction hypotheses can then interactively be verified on a virtual building site. Some testing could even be automated.

The matching of the two parts of Figure 8.23 for example could be verified through a standard registration algorithm [18]. An automatic procedure can be important when dozens of broken parts have to be matched against each other.

8.7 Applications in other areas

Besides archaeology several other areas require 3D measurements of existing structures. A few possible applications are briefly described here.

8.7.1 Architecture and conservation

As an example a partial 3D reconstruction of the cathedral of Antwerp is shown in Figure 8.24. An important registration project has recently started and the goal is to obtain a 3D model of the cathedral that could be used as a database for future preservation and restoration projects. For some of these projects an accurate texture of the model is crucial, since the level of stone degradation can be deduced from it. In this

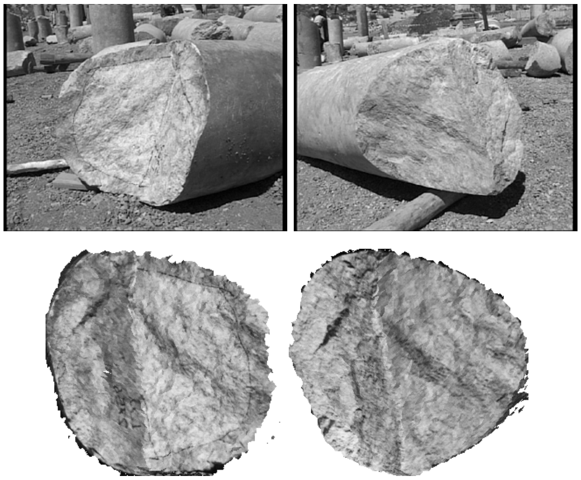

Figure 8.23: *Two images of parts of broken pillars (top) and two orthographic views of the matching surfaces generated from the 3D models (bottom)*

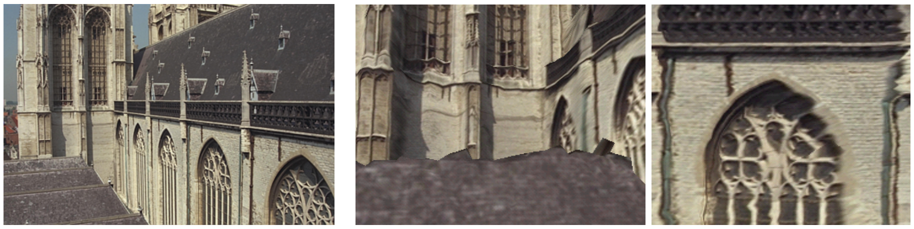

Figure 8.24: *Reconstruction of a part of the cathedral of Antwerp. One of the original images (left), two close-up views of the obtained reconstruction (middle, right).*

context the method proposed in this work has a lot to offer. An interesting approach would consist of combining the existing close-range photogrammetric techniques with our techniques. This could lead to an important increase in productivity without giving in accuracy.

8.7.2 Other applications

The flexibility of the proposed systems allows applications in many domains. In some cases further developments would be required to do so, in others the system (or parts of it) could just be used as is. Some interesting areas are forensics (e.g. crime scene reconstruction), robotics (e.g. autonomous guided vehicles), augmented reality (e.g. camera tracking) or post-production (e.g. generation of virtual sets).

8.8 Conclusion

In this chapter some results were presented in more detail to illustrate the possibilities of this work. It was shown that realistic 3D models of existing monuments could be obtained automatically from a few photographs. The flexibility of the technique allows it to be used on existing photo or video material. This was illustrated through the reconstruction of an ancient theater from a video extracted from the archives of the Belgian television. Our system was adapted to the needs of plenoptic modeling, making it possible to acquire hundreds of images of a scene without significant error accumulation.

The archaeological site of Sagalassos (Turkey) was used as a test case for our system. Several parts of the site were modeled. Since our approach is independent of scale it was also used to obtain a 3D model of the whole site at once. Some potential applications are also illustrated, i.e. 3D stratigraphy and generating/testing building hypotheses.

Some applications to other areas were also briefly discussed.

Chapter 9

Conclusion

9.1 Summary

The work presented in this dissertation deals with the automatic acquisition of realistic 3D models from images. I have tried to develop an approach which allows a maximum of flexibility during acquisition. This work consisted both of developing new theoretical insights and translating these to approaches which work on real image sequences.

This problem was decomposed in a number of tasks. For some of these tasks existing approaches were giving good results and there was no need to develop a new approach to achieve our goal. If possible an existing implementation was used. For self-calibration however no satisfying solution existed. The methods which existed at the start of my work were not giving satisfying results on real image sequences and in addition could only deal with constant camera parameters.

A first important part of this work consisted of developing a self-calibration method which would give good results on real image sequences. Inspired by the stratified approaches which had been proposed for some special motions, I developed a stratified approach which would bridge the gap from projective to metric by first solving for the affine structure. This approach is based on a new constraint for self-calibration, i.e. the modulus constraint. Good results were obtained both on synthetic experiments and on real image sequences.

The traditional assumption made for self-calibration is constant but completely unknown parameters. On the one hand this assumption is restrictive since it does not allow a zoom or even focusing to be used during acquisition. On the other hand typical cameras have rectangular or even square pixels and often the principal point is close to the center of the image. A method was proposed which could deal with all kinds of constraints on the intrinsic camera parameters, i.e. known, constant or varying. Based on this a pragmatic approach was proposed to deal with real sequences acquired with a zooming/focusing camera. The approach was validated on both real and synthetic data.

These methods were combined with other modules to obtain a complete 3D reconstruction system. The input consists of a sequence of images of a scene. The output is a realistic 3D surface model of this scene. The processing is fully automatic. This system offers a very good level of detail and realism, combined with an unprecedented flexibility in acquisition.

This flexibility of the approach was illustrated through the different examples. Realistic 3D models were generated from a few photographs as well as from preexisting video. Scenes ranging from $\approx 1m^3$ to $\approx 1km^3$ were reconstructed. Through a small adaptation the system could be used to record plenoptic functions based on hundreds of images. The system was used with success in the field of archaeology where very promising results could be achieved, enabling new applications.

9.2 Discussion and further research

The last years an ongoing discussion has taken place between protagonists of *calibrated* versus *uncalibrated* systems. In my opinion this is an unnecessary discussion. A whole palette of possibilities exist for 3D acquisition from images. One extreme consists of using a fully calibrated system under very restricted conditions (e.g. controlled lighting, restricted depth, ...). The other extreme consists of going out with a hand-held uncalibrated camera and take a few shots of a scene. In my opinion it is the task of the computer vision community to investigate these different possibilities. There is no such thing as the ideal 3D acquisition system, it all depends on the specificities of the considered application. In this context I think it is an important task for our community to explore the limits of what can be achieved.

What should be mentioned however is that the uncalibrated approach and the use of projective geometry has enabled many new insights. Projective geometry is the natural geometric framework to describe the underlying principles relating scenes to images. Often unnecessary complex equations are used to deal with the Euclidean structure of space while the concepts one is studying are already present at the projective level. Note that the use of projective geometry does not prohibit calibration to be used. It however makes it possible to distinguish easily between what depends on it and what does not.

The system that was described in this dissertation is certainly not an endpoint. As was already described at some points in the text, several possibilities exist to further enhance the system. Several important directions of further research can be identified.

A first important task consist of enhancing the system itself. This is a twofold task. On the one hand the system should be made more robust so that it can deal with a larger class of images. Failure of the algorithms and degenerate cases should be automatically detected. If possible the system should recover from these by launching a more robust or more appropriate version of the algorithm. On the other hand the accuracy of the system can certainly be increased. The relatively simple camera model which is used at this point can be extended. The use of maximum likelihood estimators can be generalized in the different modules.

A second task consists of extending the system. At this point 3D surface models of the observed scene are generated. These surfaces are however still constructed starting from a reference view. The system should fuse these different surface representation into a unique 3D model. This can be achieved by resorting to existing merging approaches which have been developed in the context of calibrated range acquisition. These methods need however to be adapted to the specificities of our approach. An interesting idea consists of using direct feedback from the final 3D model to the images to refine the model, this could be termed *photometric bundle adjustment*. Further possibilities consist of fitting parametric representations to the data (e.g. planes, quadrics, etc.) or to infer higher level representations which would bring us to model based approaches and scene understanding.

This approach is a geometric approach. The system tries to model everything with geometric primitives. Recently image-based approaches received a lot of attention. These approaches try to model the plenoptic function (i.e. the light passing in every direction through every point of the scene). These methods can thus capture the appearance of 3D scenes from images only, without the explicit use of geometry. Thus one may be able to capture very complex geometry and complex lighting that can not be modeled otherwise. These methods however suffer from other disadvantages (e.g. need for calibration, no navigation inside the scene, data intensive, etc.). An interesting approach could therefore consist of combining the two types of approaches. In this context some promising results were already obtained with our system. The idea would be to use geometry when the reprojection of the model in the images is close enough to the original and plenoptic representations otherwise. The use of intermediate representations could also be interesting (e.g. view dependent texturing).

One of the important limitations of most 3D acquisition systems is that only rigid scenes are considered. Dealing with non-rigid subjects could be a very interesting area of further research. This is however also a very complex problem. Up to now existing approaches have mostly been limited to parameterized models which are fitted to the image data. Some of the parameters can then represent a (non-rigid) pose. By exploiting results from independent motion segmentation more could be achieved. Probably more clues will be required to obtain good results (e.g. add contour information).

A last important area of further research consist of tuning the system towards applications. A first general requirement in this context is an intelligent user interface. By intelligent we mean that the necessary failure diagnosis should be done by the system itself which should then go to the user with specific –understandable– requests or suggestions. The system's interface should also enforce the different multi view relations and other constraints in a transparent way. Some further specific developments towards the different areas of application are certainly also worthwhile studying. In this context it is important to keep the system modular so that necessary extensions can easily be integrated.

Bibliography

[1] H. Akaike, “A new look at the statistical model identification”, *IEEE Trans. on Automatic Control*, 19-6, pp 716-723, 1974.

[2] M. Armstrong, A. Zisserman and P. Beardsley, “Euclidean structure from uncalibrated images”, *Proc. British Machine Vision Conference*, 1994.

[3] M. Armstrong, A. Zisserman and R. Hartley, “Euclidean Reconstruction from Image Triplets”, *Computer Vision - ECCV’96*, Lecture Notes in Computer Science, Vol. 1064, Springer-Verlag, pp. 3-16, 1996.

[4] M. Armstrong, *Self-Calibration from Image Sequences*, PhD. Thesis, Department of Engineering Science, University of Oxford, 1996.

[5] A. Azerbayejani, B. Horowitz and A. Pentland, “Recursive estimation of structure from motion using relative orientation constraints”, *Proceedings of the International Conference of Computer Vision and Pattern Recognition*, IEEE Computer Society Press, pp.294-299, June 1993.

[6] H. Baker and T. Binford, “Depth from Edge and Intensity Based Stereo”, *Int. Joint Conf. on Artificial Intelligence*, Vancouver, Canada, pp. 631-636, 1981.

[7] H. Beyer, “Accurate calibration of CCD-cameras”, *Proceedings of the International Conference of Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 1992.

[8] P. Beardsley, A. Zisserman and D. Murray, “Sequential Updating of Projective and Affine Structure from Motion”, *International Journal of Computer Vision* (23), No. 3, Jun-Jul 1997, pp. 235-259.

[9] P. Beardsley, P. Torr and A. Zisserman “3D Model Acquisition from Extended Image Sequences”, *Computer Vision - ECCV’96*, Lecture Notes in Computer Science, Vol. 1065, Springer-Verlag, pp. 683-695, 1996.

[10] W. Boehm and H. Prautzsch, “Geometric Concepts for Geometric Design”, A K Peters, 1994.

[11] D. Bondyfalat and S. Bougnoux, “Imposing Euclidean Constraints During Self-Calibration Processes”, *Proc. SMILE Workshop (post-ECCV’98)*, Lecture Notes in Computer Science, Vol. 1506, Springer-Verlag, pp.224-235, 1998.

[12] B. Boufama, R. Mohr and F. Veillon, “Euclidian Constraints for Uncalibrated Reconstruction”, *Proc. International Conference on Computer Vision*, pp. 466-470, 1993.

[13] B. Boufama and R. Mohr, “Epipole and fundamental matrix estimation using virtual parallax”, *Proc. International Conference on Computer Vision*, pp.1030-1036, 1995

[14] S. Bougnoux, “From Projective to Euclidean Space under any practical situation, a criticism of self-calibration”. *Proc. International Conference on Computer Vision*, Narosa Publishing House, New Delhi /Madras /Bombay /Calcutta /London, pp. 790-796, 1998.

[15] S. Bougnoux and L. Robert, “TotalCalib: a fast and reliable system for off-line calibration of images sequences”, DEMO Session, CVPR’97, June 1997.

[16] D. Brown, “Close-range camera calibration”, Photogrammetric Engineering, 37(8):855-866, 1971.

[17] D. Brown, “The bundle adjustment - progress and prospect”, *XIII Congress of the ISPRS*, Helsinki, 1976.

[18] Y. Chen and G. Medioni, “Object Modeling by Registration of Multiple Range Images”, *Proc. Int. Conf. on Robotics and Automation*, 1991.

[19] D. Capel and A. Zisserman, Automated Mosaicing with Super-resolution Zoom *Proceedings of the Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 885-891, 1998.

[20] I. Cox, S. Hingorani and S. Rao, “A Maximum Likelihood Stereo Algorithm”, *Computer Vision and Image Understanding*, Vol. 63, No. 3, May 1996.

[21] N. Cui, J. Weng and P. Cohen, “Extended Structure and Motion Analysis from Monocular Image Sequences”, *Proc. International Conference on Computer Vision*, pp. 222-229, Osaka, Japan, 1990.

[22] L. de Agapito, E. Hayman and I. Reid “Self-calibration of a rotating camera with varying intrinsic parameters”, *Proceedings of the ninth British Machine Vision Conference*, Sept 1998.

[23] R. Deriche and O. Faugeras, “Tracking Line Segments”, *Computer Vision-ECCV’90*, Lecture Notes in Computer Science, Vol. 427, Springer-Verlag, 1990.

[24] R. Deriche and G. Giraudon, “A computational approach for corner and vertex detection”, *International Journal of Computer Vision*, 1(2):167-187, 1993.

[25] R. Deriche, Z. Zhang, Q.T. Luong and O. Faugeras, “Robust recovery of the epipolar geometry for an uncalibrated stereo rig”, *Computer Vision - ECCV’94*, Lecture Notes in Computer Science, Vol. 801, Springer-Verlag, pp. 567-576, 1994.

[26] P. Debevec, C. Taylor and J. Malik, “Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach”, *Siggraph*, 1996.

[27] P. Debevec, Y. Yu and G. Borshukov, “Efficient View-Dependent Image-Based Rendering with Projective Texture Mapping”, *Proc. SIGGRAPH ’98*, ACM Press, New York, 1998.

[28] F. Devernay and O. Faugeras, “From projective to euclidean reconstruction”, *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 264-269, 1996.

[29] U. Dhond and J. Aggarwal, “Structure from Stereo - A Review”, IEEE Trans. Syst., Man and Cybern. 19, 1489-1510, 1989.

[30] L. Falkenhagen, “Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints”. *Proc. International Workshop on SNHC and 3D Imaging*, Rhodes, Greece, 1997.

[31] L. Falkenhagen, “Depth Estimation from Stereoscopic Image Pairs assuming Piecewise Continuous Surfaces”, *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Productions*, Hamburg, Germany, 1994.

[32] O. Faugeras and G. Toscani, “Camera Calibration for 3D Computer Vision”, *International Workshop on Machine Vision and Machine Intelligence*, pp. 240-247, Tokyo, 1987.

[33] 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”, *Computer Vision – ECCV’98*, vol.1, Lecture Notes in Computer Science, Vol. 1406, Springer-Verlag, pp.36-52, 1998.

[34] O. Faugeras, *Three-Dimensional Computer Vision: a Geometric Viewpoint*, MIT press, 1993.

[35] O. Faugeras, “Stratification of three-dimensional vision: projective, affine, and metric representations”, Journal of the Optical Society of America A, pp. 465–483, Vol. 12, No.3, March 1995.

[36] O. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig”, *Computer Vision - ECCV’92*, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 563-578, 1992.

[37] O. Faugeras, Q.-T. Luong and S. Maybank. “Camera self-calibration: Theory and experiments”, *Computer Vision - ECCV’92*, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 321-334, 1992.

[38] O. Faugeras and S. Maybank, “Motion from point matches: multiplicity of solutions”, *International Journal of Computer Vision*, 4(3):225-246, June 1990.

[39] O. Faugeras and B. Mourrain, “on the geometry and algebra of point and line correspondences between *n* images”, *Proc. International Conference on Computer Vision*, 1995, pp. 951-962.

[40] M. Fischler and R. Bolles, “RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography”, *Commun. Assoc. Comp. Mach.*, 24:381-95, 1981.

[41] A. Fitzgibbon and A. Zisserman, “Automatic camera recovery for closed or open image sequences”, *Computer Vision – ECCV’98*, vol.1, Lecture Notes in Computer Science, Vol. 1406, Springer-Verlag, 1998. pp.311-326, 1998.

[42] A. Fitzgibbon and A. Zisserman, “Automatic 3D Model Acquisition and Generation of New Images from Video Sequences”, *Proceedings of European Signal Processing Conference (EUSIPCO ’98)*, Rhodes, Greece, pp. 1261-1269, 1998.

[43] W. Förstner, “A framework for low level feature extraction” *Computer Vision-ECCV’90*, Lecture Notes in Computer Science, Vol. 427, Springer-Verlag, pp.383-394, 1990.

[44] G. Gimel’farb, “Symmetrical approach to the problem of automatic stereoscopic measurements in photogrammetry”, Cybernetics, 1979, 15(20, 235-247; Consultants Bureau, N.Y.

[45] S. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, “The Lumigraph”, *Proc. SIGGRAPH ’96*, pp 43–54, ACM Press, New York, 1996.

[46] W. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, Massachusetts, 1981.

[47] A. Gruen, “Accuracy, reliability and statistics in close-range photogrammmetry”, *Proceedings of the Symposium of the ISP Commision V*, Stockholm, 1978.

[48] A. Gruen and H. Beyer, “System calibration through self-calibration”, *Proceedings of the Workshop on Calibration and Orientation of Cameras in Computer Vision*, 1992.

[49] G. Golub and C. Van Loan, *Matrix Computations*, John Hopkins University Press, 1983.

[50] C. Harris and M. Stephens, “A combined corner and edge detector”, *Fourth Alvey Vision Conference*, pp.147-151, 1988.

[51] R. Hartley, “Estimation of relative camera positions for uncalibrated cameras”, *Computer Vision - ECCV’92*, Lecture Notes in Computer Science, Vol. 588, Springer-Verlag, pp. 579-587, 1992.

[52] R. Hartley, “Cheirality invariants”, *Proc. D.A.R.P.A. Image Understanding Workshop*, pp. 743-753, 1993.

[53] R. Hartley, “Euclidean reconstruction from uncalibrated views”, in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, pp. 237-256, 1994.

[54] R. Hartley, “Projective reconstruction from line correspondences”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 1994.

[55] R. Hartley, “Self-calibration from multiple views with a rotating camera”, Lecture Notes in Computer Science, Vol. 800-801, Springer-Verlag, pp. 471-478, 1994.

[56] R. Hartley, “A linear method for reconstruction from points and lines”, *Proc. International Conference on Computer Vision*, pp. 882-887, 1995.

[57] R. Hartley, “In defense of the eight-point algorithm”. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580-593, June 1997.

[58] R. Hartley and P. Sturm, “Triangulation”, *Computer Vision and Image Understanding*, 68(2):146-157, 1997.

[59] R. Hartley, “Computation of the Quadrifocal Tensor”, *Computer Vision-ECCV’98*, Lecture Notes in Computer Science, Vol. 1406, Springer-Verlag, pp. 20-35, 1998.

[60] A. Heyden and K. Åström, “Euclidean Reconstruction from Constant Intrinsic Parameters” *Proc. 13th International Conference on Pattern Recognition*, IEEE Computer Soc. Press, pp. 339-343, 1996.

[61] A. Heyden and K. Åström, “Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 438-443, 1997.

[62] A. Heyden, *Geometry and Algebra of Multiple Projective Transformations*, Ph.D.thesis, Lund University, 1995.

[63] A. Heyden, R. Berthilsson, G. Sparr, “An Iterative Factorization Method for Projective Structure and Motion from Image Sequences”, to appear in *Image and Vision Computing*.

[64] A. Heyden and K. Åström, “Minimal Conditions on Intrinsic Parameters for Euclidean Reconstruction”, Asian Conference on Computer Vision, Hong Kong, 1998.

[65] R. Horaud and G. Csurka, “Self-Calibration and Euclidean Reconstruction Using Motions of a Stereo Rig”, *Proc. International Conference on Computer Vision*, Narosa Publishing House, New Delhi /Madras /Bombay /Calcutta /London pp. 96-103, 1998,.

[66] B. Horn, *Robot Vision*, MIT Press, 1986.

[67] M. Irani and S. Peleg, Super resolution from image sequences, *Proc. International Conference on Pattern Recognition*, Atlantic City, NJ, 1990.

[68] D. Jacobs, “Linear Fitting with Missing Data; Applications to Structure-from-Motion and to Characterizing Intensity Images”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society press, pp. 206-212, 1997.

[69] R. Kalman, “A new approach to linear filtering and prediction problems”, *Transactions A.S.M.E., Journal of Basic Engineering*, March 1960.

[70] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice* , Elsevier Science, Amsterdam, 1996.

[71] W. Karl, G. Verghese and A. Willsky, Reconstructing ellipsoids from projections. *CVGIP; Graphical Models and Image Processing*, 56(2):124-139, 1994.

[72] R. Koch, *Automatische Oberflachenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten, PhD thesis*, University of Hannover, Germany, 1996 also published as Fortschritte-Berichte VDI, Reihe 10, Nr.499, VDI Verlag, 1997.

[73] R. Koch, M. Pollefeys and L. Van Gool, Multi Viewpoint Stereo from Uncalibrated Video Sequences. *Proc. European Conference on Computer Vision*, pp.55-71. Freiburg, Germany, 1998.

[74] R. Koch, M. Pollefeys and L. Van Gool, Automatic 3D Model Acquisition from Uncalibrated Image Sequences, Proceedings Computer Graphics International, pp.597-604, Hannover, 1998.

[75] R. Koch, “3-D Surface Reconstruction from Stereoscopic Image Sequences”, *Proc. Fifth International Conference on Computer Vision*, IEEE Computer Soc. Press, pp. 109-114, 1995.

[76] J. Koenderink and A. Van Doorn, “Affine structure from motion”, *Journal of Optical Society of America*, 8(2):377-385, 1991.

[77] E. Kruppa, “Zur ermittlung eines objektes aus zwei perspektiven mit innerer orientierung”, *Sitz.-Ber. Akad. Wiss., Wien, math. naturw. Abt. IIa*, 122:1939-1948, 1913.

[78] S. Laveau and O. Faugeras, “Oriented Projective Geometry for Computer Vision”, in : B. Buxton and R. Cipolla (eds.), *Computer Vision - ECCV’96*, Lecture Notes in Computer Science, Vol. 1064, Springer-Verlag, pp. 147-156, 1996.

[79] S. Laveau, *Géométrie d’un systeème de N caméras. Théorie, estimation et applications*, Ph.D. thesis, Ecole Polytechnique, France, 1996.

[80] J.-M. Lavest, G. Rives and M. Dhome, “3D reconstruction by zooming”, *IEEE Robotics and Automation*, 1993.

[81] R. Lenz and R. Tsai, “Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:713-720, 1988.

[82] M. Levoy and P. Hanrahan, “Lightfield Rendering”, *Proc. SIGGRAPH ’96*, pp 31–42, ACM Press, New York, 1996.

[83] M. Li, “Camera Calibration of a Head-Eye System for Active Vision”, *Computer Vision-ECCV’94*, Lecture Notes in Computer Science, Vol.800, pp. 543-554, Springer-Verlag, 1994.

[84] H. Longuet-Higgins, “A computer algorithm for reconstructing a scene from two projections”, *Nature*, 293:133-135, 1981.

[85] Q.-T. Luong, *Matrice Fondamentale et Autocalibration en Vision par Ordinateur*, PhD thesis, Université de Paris-Sud, France, 1992.

[86] Q.-T. Luong and O. Faugeras, “The fundamental matrix: theory, algorithms, and stability analysis”, *Internation Journal of Computer Vision*, 17(1):43-76, 1996.

[87] Q.-T. Luong and O. Faugeras, “Self Calibration of a moving camera from point correspondences and fundamental matrices”, *Internation Journal of Computer Vision*, vol.22-3, 1997.

[88] Q.-T. Luong and T. Vieville, “Canonic representation for the geometry of multiple projective views”, *Computer Vision - ECCV’94*, Lecture Notes in Computer Science, Vol. 800, Springer-Verlag, pp. 589-600, 1994.

[89] D. Marr and T. Poggio, “A Computational Theory of Human Stereo Vision”, *Proc. Royal Society of London*, Vol. 204 of B, pp. 301-328, 1979.

[90] G. McLean and D. Kotturi, “Vanishing point Detection by Line Clustering”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.17, No. 11, pp.1090-1095, Nov. 1995.

[91] P. McLauchlan and D. Murray, “A unifying framework for structure from motion recovery from image sequences”, *Proc. International Conference on Computer Vision*, IEEE Computer Cosiety Press, pp. 314-320, 1995.

[92] P. McLauchlan and D. Murray, “Active camera calibration for a Head-Eye platform using the variable State-Dimension filter”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):15-22, 1996.

[93] L. McMillan and G. Bishop, “Plenoptic modeling: An image-based rendering system”, *Proc. SIGGRAPH’95*, pp. 39-46, 1995.

[94] S. Maybank, *Theory of reconstruction from image motion*, Springer-Verlag, Berlin, 1993.

[95] S. Maybank and O. Faugeras, “A theory of self-calibration of a moving camera”, *International Journal of Computer Vision*, 8:123-151, 1992.

[96] R. Mohr, F. Veillon and L. Quan, “Relative 3D reconstruction using multiple uncalibrated images”, *Proc. International Conference on Computer Vision*, IEEE Computer Soc. Press, pp.543-548, 1993.

[97] R. Mohr, B. Boufama and P. Brand, “Accurate projective reconstruction”, in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, pp. 297-316, 1994.

[98] T. Moons, “A Guided Tour Through Multiview Relations”, *Proc. SMILE Workshop (post-ECCV’98)*, Lecture Notes in Computer Science 1506, Springer-Verlag, pp.304-346, 1998.

[99] T, Moons, L. Van Gool, M. Proesmans and E. Pauwels, “Affine reconstruction from perspective image pairs with a relative object-camera translation in between”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no.1, pp. 77-83, Jan. 1996.

[100] T. Moons, L. Van Gool, M. Van Diest, and E. Pauwels, “Affine reconstruction from perspective image pairs”, in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, pp. 297–316, 1994.

[101] T. Moons, L. Van Gool and M. Pollefeys, “Geometrical structure from perspective image pairs”, in : F.Dillen, L.Vrancken, L.Verstraelen, and I. Van de Woestijne (eds.), *Geometry and topology of submanifolds, VII* World Scientific, pp.305-308, 1996.

[102] A. Morgan, *Solving polynomial systems using continuation for engineering and scientific problems*, Prentice-Hall Englewood Cliffs (N.J.), 1987.

[103] D. Morris and T. Kanade, "A Unified Factorization Algorithm for Points, Line Segments and Planes with Uncertainty Models", *Proc. International Conference on Computer Vision*, Narosa Publishing House, pp 696-702, 1998.

[104] M. Mühlich and R. Mester, "The Role of Total Least Squares in Motion Analysis", *Proc. ECCV'98*, pp. 305-321, 1998.

[105] J. Mundy and A. Zisserman, "Machine Vision", in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, 1994.

[106] E. Ofek, E. Shilat, A. Rappopport and M. Werman, "Highlight and Reflection Independent Multiresolution Textures from Image Sequences", *IEEE Computer Graphics and Applications*, vol.17 (2), March-April 1997.

[107] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter-scanline Search Using Dynamic Programming", *IEEE Trans. on Pattern Analysis and Machine Intelligence* 7(2), 139-154, 1985.

[108] M. Okutomi and T. Kanade, "A Locally Adaptive Window for Signal Processing", *International Journal of Computer Vision*, 7, 143-162, 1992.

[109] PhotoModeler, by Eos Systems Inc., `http://www.photomodeler.com/`.

[110] S. Pollard, J. Mayhew and J. Frisby, "PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit", *Perception* 14(4), 449-470, 1985.

[111] M. Pollefeys and L. Van Gool, "Stratified self-calibration with the modulus constraint", accepted for publication in *IEEE transactions on Pattern Analysis and Machine Intelligence*.

[112] M. Pollefeys, R. Koch and L. Van Gool. "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters", accepted for publication in *International Journal of Computer Vision*.

[113] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, "An Automatic Method for Acquiring 3D models from Photographs: applications to an Archaeological Site", accepted for *Proc. ISPRS International Workshop on Photogrammetric Measurements, Object Modeling and Documentation in Architecture and Industry*, july 1999.

[114] M. Pollefeys, M. Proesmans, R. Koch, M. Vergauwen and L. Van Gool, "Detailed model acquisition for virtual reality", in J. Barcelo, M. Forte and D. Sanders (eds.), *Virtual Reality in Archaeology*, to appear, ArcheoPress, Oxford.

[115] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, "Automatic Generation of 3D Models from Photographs", *Proceedings Virtual Systems and MultiMedia*, 1998.

[116] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, “Virtualizing Archaeological Sites”, *Proceedings Virtual Systems and MultiMedia*, 1998.

[117] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, “Metric 3D Surface Reconstruction from Uncalibrated Image Sequences”, *Proc. SMILE Workshop (post-ECCV’98)*, Lecture Notes in Computer Science, Vol. 1506, pp.138-153, Springer-Verlag, 1998.

[118] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, “Flexible acquisition of 3D structure from motion”, *Proceedings IEEE workshop on Image and Multidimensional Digital Signal Processing*, pp.195-198, Alpbach, 1998.

[119] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, “Flexible 3D Acquisition with a Monocular Camera”, *Proceedings IEEE International Conference on Robotics and Automation*, Vol.4, pp.2771-2776, Leuven, 1998.

[120] M. Pollefeys, R. Koch and L. Van Gool, “Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters”, *Proc. International Conference on Computer Vision*, Narosa Publishing House, pp.90-95, 1998.

[121] M. Pollefeys, L. Van Gool and M. Proesmans, “Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths”, *Computer Vision - ECCV’96*, Lecture Notes in Computer Science, Vol. 1064, Springer-Verlag, pp. 31-42, 1996.

[122] M. Pollefeys, L. Van Gool and A. Oosterlinck, “The Modulus Constraint: A New Constraint for Self-Calibration”, *Proc. 13th International Conference on Pattern Recognition*, IEEE Computer Soc. Press, pp. 349-353, 1996.

[123] M. Pollefeys and L. Van Gool, “A stratified approach to self-calibration”, *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 407-412, 1997.

[124] M. Pollefeys and L. Van Gool, “Self-calibration from the absolute conic on the plane at infinity”, *Proc. Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, Vol. 1296, Springer-Verlag, pp. 175-182, 1997.

[125] M. Pollefeys, L. Van Gool and T. Moons. “Euclidean 3D reconstruction from stereo sequences with variable focal lengths”, *Recent Developments in Computer Vision*, Lecture Notes in Computer Science, Vol.1035, Springer-Verlag, pp. 405-414, 1996.

[126] M. Pollefeys, L. Van Gool and T. Moons. “Euclidean 3D reconstruction from stereo sequences with variable focal lengths”, *Proc.Asian Conference on Computer Vision*, Vol.2, pp.6-10, Singapore, 1995

[127] M. Pollefeys, L. Van Gool and A. Oosterlinck, “Euclidean self-calibration via the modulus constraint”, in F.Dillen, L.Vrancken, L.Verstraelen, and I. Van de Woestijne (eds.), *Geometry and topology of submanifolds, VIII*, World Scientific, Singapore, New Jersey, London, Hong Kong, pp.283-291, 1997.

[128] M. Pollefeys, R. Koch and L. Van Gool. “Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters”, Technical Report Nr. KUL/ESAT/MI2/9707, MI2-ESAT, K.U.Leuven, Belgium, 1997.

[129] M. Pollefeys and L. Van Gool. “A stratified approach to metric self-calibration”, Technical Report Nr. KUL/ESAT/MI2/9702, MI2-ESAT, K.U.Leuven, Belgium, 1997.

[130] M. Pollefeys, L. Van Gool and Andre Oosterlinck. “Self-calibration with the modulus constraint” , Technical Report Nr. KUL/ESAT/MI2/9609, MI2-ESAT, K.U.Leuven, Belgium, 1996.

[131] M. Pollefeys, L. Van Gool, M. Proesmans. “Euclidean 3D reconstruction from image sequences with variable focal lengths”, Technical Report Nr. KUL/ESAT/MI2/9508, MI2-ESAT, K.U.Leuven, Belgium, 1995.

[132] P. Pritchett and A. Zisserman, “Wide Baseline Stereo Matching”, *Proc. International Conference on Computer Vision*, Narosa Publishing House, pp. 754-760, 1998.

[133] P. Pritchett and A. Zisserman, “Matching and Reconstruction from Widely Separate Views”, *Proc. SMILE Workshop (post-ECCV’98)*, Lecture Notes in Computer Science, Vol. 1506, Springer-Verlag, pp.78-92, 1998.

[134] M. Proesmans, L. Van Gool and A. Oosterlinck, “Determination of optical flow and its discontinuities using non-linear diffusion”, *Computer Vision - ECCV’94*, Lecture Notes in Computer Science, Vol. 801, Springer-Verlag, pp. 295-304, 1994.

[135] L. Robert, C. Zeller, O. Faugeras, and M. Hébert. “Applications of non-metric vision to some visually-guided robotics tasks”. In *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, Y. Aloimonos, (ed.), chapter 5, pages 89-134. Lawrence Erlbaum Associates, 1997.

[136] P. Rousseeuw, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.

[137] S. Roy, J. Meunier and I. Cox, “Cylindrical Rectification to Minimize Epipolar Distortion”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.393-399, 1997.

[138] F. Schaffalitzky and A. Zisserman, “Geometric Grouping of Repeated Elements within Images”, *Proc. 9th British Machine Vision Conference*, pp 13-22, 1998.

[139] C. Schmid and R. Mohr, “Local Greyvalue Invariants for Image Retrieval”, *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol.19, no.5, pp 872-877, may 1997.

[140] C. Schmid and A. Zisserman, “Automatic Line Matching across Views”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp 666-671, 1997.

[141] C. Schmid, R. Mohr and C. Bauckhage, “Comparing and Evaluating Interest Points”, *Proc. International Conference on Computer Vision*, Narosa Publishing House, pp. 230-235, 1998.

[142] J.G. Semple and G.T. Kneebone, *Algebraic Projective Geometry*, Oxford University Press, 1952.

[143] A. Shashua, “Omni-Rig Sensors: What Can be Done With a Non-Rigid Vision Platform?” *Proc. of the Workshop on Applications of Computer Vision (WACV)*, Princeton, Oct. 1998.

[144] A. Shashua, “Trilinearity in visual recognition by alignment”, *Computer Vision - ECCV’94*, Lecture Notes in Computer Science, Vol. 801, Springer-Verlag, pp. 479-484, 1994.

[145] A. Shashua and S. Avidan, The Rank 4 Constraint in Multiple View Geometry, *Proc. Computer Vision-ECCV’96*, Springer-Verlag, April 1996.

[146] H.-Y. Shum, M. Han and R. Szeliski, “Interactive construction of 3D models from panoramic mosaics”, *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, IEEE comp. soc. press, pp. 427-433, 1998.

[147] C. Slama, *Manual of Photogrammetry*, American Society of Photogrammetry, Falls Church, VA, USA, 4th edition, 1980.

[148] M. Spetsakis and J. Aloimonos, “Structure from motion using line correspondences”, *International Journal of Computer Vision*, 4(3):171-183, 1990.

[149] G. Stein, “Lens Distortion Calibration Using Point Correspondences”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp 602-608, 1997.

[150] P. Sturm, *Vision 3D non-calibrée: contributions à la reconstruction projective et études des mouvements critiques pour l’auto-calibrage*, Ph.D. Thesis, INP de Grenoble, France , 1997.

[151] P. Sturm and L. Quang, “Affine stereo calibration”, *Proceedings Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, Vol. 970, Springer-Verlag, pp. 838-843, 1995.

[152] P. Sturm, "Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction", *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 1100-1105, 1997.

[153] P. Sturm, "Critical motion sequences and conjugacy of ambiguous Euclidean reconstructions", *Proc. SCIA - 10th Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, pp. 439-446, 1997.

[154] P. Sturm and B. Triggs. "A factorization based algorithm for multi-image projective structure and motion". *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. 1064, Springer-Verlag, pp. 709-720, 1996.

[155] P. Sturm, "Note 28: Critical Motion Sequences for Cameras with Free Focal Length", University of Reading, August 1998.

[156] R. Szeliski and S. Kang, "Recovering 3D shape and motion from image streams using non-linear least-squares", DEC technical report 93/3, DEC, 1993.

[157] TargetJr, `http://www.targetjr.org/`.

[158] C. Taylor, P. Debevec and J. Malik, "Reconstructing Polyhedral Models of Architectural Scenes from Photographs", *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. 1065, vol.II, pp 659-668, 1996.

[159] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization approach", *International Journal of Computer Vision*, 9(2):137-154, 1992.

[160] P. Torr and A. Zisserman, "Robust parametrization and computation of the trifocal tensor", *Image and Vision Computing*, 15(1997) 591-605.

[161] P. Torr and A. Zisserman, "Robust Computation and Parameterization of Multiple View Relations", *Proc. International Conference on Computer Vision*,Narosa Publishing house, pp 727-732, 1998.

[162] P. Torr, P. Beardsley and D. Murray, "Robust Vision", *Proc. British Machine Vision Conference*, 1994.

[163] P. Torr, *Motion Segmentation and Outlier Detection*, PhD Thesis, Dept. of Engineering Science, University of Oxford, 1995.

[164] P. Torr, A. Fitzgibbon and A. Zisserman, "Maintaining Multiple Motion Model Hypotheses Over Many Views to Recover Matching and Structure", *Proc. International Conference on Computer Vision*, Narosa Publishing house, pp 485-491, 1998.

[165] B. Triggs, "The geometry of projective reconstruction I: Matching constraints and the joint image", *Proc. International Conference on Computer Vision*, IEEE Computer Soc. Press, pp. 338-343, 1995.

[166] B. Triggs, “The Absolute Quadric”, *Proc. 1997 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Soc. Press, pp. 609-614, 1997.

[167] B. Triggs, “Autocalibration from planar scenes”, *Computer Vision – ECCV’98*, vol.1, Lecture Notes in Computer Science, Vol. 1406, Springer-Verlag, pp 89-105, 1998.

[168] R. Tsai and T. Huang, “Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces”, *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol.6, pp.13-27, Jan. 1984.

[169] R. Tsai, “An efficient and accurate camera calibration technique for 3D machine vision”, *Proc. Computer Vision and Pattern Recognition*, 1986.

[170] R. Tsai. “A versatile camera calibration technique for high-accuracy 3D machine vision using off-the-shelf TV cameras and lenses”. *IEEE Journal of Robotics and Automation*, RA-3(4):323-331, August 1987.

[171] T. Tuytelaars and L. Van Gool, “Content-based Image Retrieval based on Local Affinely Invariant Regions”, accepted for the *Third International Conference on Visual Information Systems, VISUAL99*, to be held in Amsterdam, June 2-4, 1999.

[172] T. Tuytelaars, M. Proesmans, L. Van Gool, “The cascaded Hough transform as support for grouping and finding vanishing points and lines”, *Proceedings international workshop on Algebraic Frames for Perception-Action Cycle*, Lecture Notes in Computer Science, Vol. 1315, Springer-Verlag, pp. 278-289, 1997.

[173] L. Van Gool, T. Moons, M. Proesmans and M. Van Diest, “Affine reconstruction from perspective image pairs obtained by a translating camera”, *Proceedings International Conference on Pattern Recognition* vol. I, pp. 290-294, Jerusalem, Israel, 1994.

[174] L. Van Gool, F. Defoort, R. Koch, M. Pollefeys, M. Proesmans and M. Vergauwen, “3D modeling for communications”, *Proceedings Computer Graphics International*, pp.482-487, Hannover, 1998.

[175] T. Vieville and D. Lingrad, “Using singular displacements for uncalibrated monocular vision systems”, *Computer Vision - ECCV’96*, Lecture Notes in Computer Science, Vol. 1064, Springer-Verlag, pp. 207-216, 1996.

[176] H. Wang and M. Brady, “Corner detection: some new results”, *IEE Colloquium Digest of Systems Aspects of Machine Vision*, pp. 1.1-1.4, London, 1992.

[177] J. Weng, P. Cohen and M. Herniou, “Camera calibration with distortion models and accuracy evaluation”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(10):965-980, 1992.

[178] J. Weng, T. Huang and N. Ahuja, “Motion ans structure from image sequences”, Springer-Verlag, Series in Information Science, Vol. 29, 1993.

[179] R. Willson and S. Shafer, “A Perspective Projection Camera Model for Zoom Lenses”, *Proceedings Second Conference on Optical 3-D Measurement Techniques*, Zurich Switzerland, October 1993.

[180] R. Willson, “Modeling and Calibration of Automated Zoom Lenses” *Proceedings of the SPIE 2350:Videometrics III*, Boston MA, October 1994, pp.170-186.

[181] R. Willson, *Modeling and Calibration of Automated Zoom Lenses*, Ph.D. thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, January 1994.

[182] R. Willson and S. Shafer, “What is the Center of the Image?”, *Journal of the Optical Society of America A*, Vol. 11, No. 11, pp.2946-2955, November 1994.

[183] C. Zeller, *Calibration projective, affine et Euclidienne en vision par ordinateur et application a la perception tridimensionnelle*, Ph.D. Thesis, Ecole Polytechnique, France, 1996.

[184] C. Zeller and O. Faugeras, “Camera self-calibration from video sequences: the Kruppa equations revisited”. INRIA, Sophia-Antipolis, France, Research Report 2793, 1996.

[185] C. Zeller and O. Faugeras, “Projective, affine and metric measurements from video sequences”, *Proc. of the International Symposium on Optical Science, Engineering and Instrumentation*, 1995.

[186] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, “A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry”, *Artificial Intelligence Journal*, Vol.78, pp.87-119, October 1995.

[187] Z. Zhang, “On the Epipolar Geometry Between Two Images with Lens Distortion”, *Proc. International Conference on Pattern Recognition*, IEEE Computer Soc. Press, A80.13, 1996.

[188] A. Zisserman, P. Beardsley and I. Reid, “Metric calibration of a stereo rig”, *Proceedings IEEE Workshop on Representation of Visual Scenes*, Cambridge, pp. 93-100, 1995.

[189] A. Zisserman, D. Liebowitz and M. Armstrong, “Resolving ambiguities in auto-calibration”, *Phil. Trans. R. Soc. Lond.*, A(1998) 356, 1193-1211.

Appendix A

The modulus constraint

A.1 Derivation of the modulus constraint

The roots of equation (5.2) must obey $|\lambda_1| = |\lambda_2| = |\lambda_3|$. In this appendix a necessary condition is derived. A third order polynomial can be written as follows.

$$l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = l_3(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \tag{A.1}$$

From equation (A.1) the following relations follow:

$$\lambda_1 + \lambda_2 + \lambda_3 = -\frac{l_2}{l_3} \tag{A.2}$$

$$\lambda_1(\lambda_2 + \lambda_3) + \lambda_2\lambda_3 = \frac{l_1}{l_3} \tag{A.3}$$

$$\lambda_1\lambda_2\lambda_3 = -\frac{l_0}{l_3} \tag{A.4}$$

At least one of the roots must be real, therefore it can be assumed that λ_1 is real (λ_2 and λ_3 can then be either real or complex). If the roots have the same moduli the following equation must be satisfied.

$$\lambda_1^2 = \lambda_2\lambda_3 \tag{A.5}$$

Rewriting (A.3) using (A.2) and (A.5) yields

$$\lambda_1(-\frac{l_2}{l_3} - \lambda_1) + \lambda_1^2 = \frac{l_1}{l_3}$$

or

$$\lambda_1 = -\frac{l_1}{l_2} \tag{A.6}$$

substituting (A.5) in (A.4) implies

$$\lambda_1^3 = -\frac{l_0}{l_3} \tag{A.7}$$

Eliminating λ_1 from the equations (A.6) and (A.7) gives a necessary condition that is only depending on l_3, l_2, l_1, l_0.

$$l_3 l_1^3 = l_2^3 l_0 \tag{A.8}$$

Note that equation A.6 and thus also equation A.8 are only necessary conditions for the homography to have eigenvalues corresponding to a rotation matrix. These equations can also be satisfied for three real eigenvalues. When multiple solutions persist, solutions for which no two eigenvalues are conjugate can be ruled out.

A.2 Expressions for l_3, l_2, l_1 and l_0

In this section expressions for l_3, l_2, l_1, l_0 will be derived. They will be expressed in terms of $\pi_\infty = [p_X p_Y p_Z]^\top$ and the projective calibration. Starting from equation 5.2 a similar but simpler equation can be derived which avoids the occurrence of the matrix inversion.

$$\begin{array}{rcl}
\det\left(\mathbf{H}_{ij} - \lambda\mathbf{I}\right) & = & \det\left(\mathbf{H}_{1j}\mathbf{H}_{1i}^{-1} - \lambda\mathbf{I}\right) = \det\left(\mathbf{H}_{1i}^{-1}\right)\det\left(\mathbf{H}_{1j} - \lambda\mathbf{H}_{1i}\right) = 0 \\
 & \Updownarrow & \\
\det\left(\mathbf{H}_{1j} - \lambda\mathbf{H}_{1i}\right) & = & 0
\end{array}$$

The following notations are used to simplify the expressions: $\mathbf{H}_{1j} = [\mathtt{h}_1\mathtt{h}_2\mathbf{h}_3]$, $\mathbf{H}_{1i} = [\mathbf{h}'_1\mathbf{h}'_2\mathbf{h}'_3]$, $\mathsf{e}_{1j} = \mathsf{e}$ and $\mathsf{e}_{1i} = \mathsf{e}'$, $|\mathbf{H}|$ means the determinant of $\mathbf{H}$.

$$\begin{array}{rcl}
\det\left(\mathbf{H} - \lambda\mathbf{H}'\right) & = & |\mathtt{h}_1 - \lambda\mathtt{h}'_1\, \mathtt{h}_2 - \lambda\mathtt{h}'_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| \\
 & = & |\mathtt{h}_1\, \mathtt{h}_2 - \lambda\mathtt{h}'_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| - \lambda\,|\mathtt{h}'_1\, \mathtt{h}_2 - \lambda\mathtt{h}'_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| \\
 & = & \begin{array}{ll} |\mathtt{h}_1\, \mathtt{h}_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| & -\lambda\,|\mathtt{h}_1\, \mathtt{h}'_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| \\ -\lambda\,|\mathtt{h}'_1\, \mathtt{h}_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| & +\lambda^2\,|\mathtt{h}'_1\, \mathtt{h}'_2\, \mathtt{h}_3 - \lambda\mathtt{h}'_3| \end{array} \\
 & = & \begin{array}{llll} |\mathtt{h}_1\mathtt{h}_2\mathtt{h}_3| & -\lambda|\mathtt{h}'_1\mathtt{h}_2\mathtt{h}_3| & +\lambda^2|\mathtt{h}_1\mathtt{h}'_2\mathtt{h}'_3| & -\lambda^3|\mathtt{h}'_1\mathtt{h}'_2\mathtt{h}'_3| \\ & -\lambda|\mathtt{h}_1\mathtt{h}'_2\mathtt{h}_3| & +\lambda^2|\mathtt{h}'_1\mathtt{h}_2\mathtt{h}'_3| & \\ & -\lambda|\mathtt{h}_1\mathtt{h}_2\mathtt{h}'_3| & +\lambda^2|\mathtt{h}'_1\mathtt{h}'_2\mathtt{h}_3| & \end{array}
\end{array} \tag{A.9}$$

In the above expressions $\mathbf{H} + \mathsf{e}\pi_\infty^\top$ or a similar expression should be substituted to $[\mathtt{h}_1\mathtt{h}_2\mathtt{h}_3]$, $[\mathtt{h}'_1\mathtt{h}_2\mathtt{h}_3]$, ..., $|\mathtt{h}'_1\mathtt{h}'_2\mathtt{h}'_3|$. Therefore the determinant of $\mathbf{H} + \mathsf{e}\pi_\infty^\top$ should also be factorized. The other determinants can be factorized in a similar way.

$$\begin{array}{rcl}
\det(\mathbf{H} + \mathsf{e}\pi_\infty^\top) & = & |\mathtt{h}_1 + p_X\mathsf{e}\, \mathtt{h}_2 + p_Y\mathsf{e}\, \mathtt{h}_3 + p_Z\mathsf{e}| \\
 & = & |\mathtt{h}_1\, \mathtt{h}_2 + p_Y\mathsf{e}\, \mathtt{h}_3 + p_Z\mathsf{e}| + p_X\,|\mathsf{e}\, \mathtt{h}_2 + p_Y\mathsf{e}\, \mathtt{h}_3 + p_Z\mathsf{e}| \\
 & = & \begin{array}{lll} |\mathtt{h}_1\, \mathtt{h}_2\, \mathtt{h}_3 + p_Z\mathsf{e}| & +p_X\,|\mathsf{e}\, \mathtt{h}_2\, \mathtt{h}_3 + p_Z\mathsf{e}| & \\ & +p_Y\,|\mathtt{h}_1\, \mathsf{e}\, \mathtt{h}_3 + p_Z\mathsf{e}| & +p_X p_Y\, \underbrace{|\mathsf{e}\, \mathsf{e}\, \mathtt{h}_3 + p_Z\mathsf{e}|}_{=0} \end{array} \\
 & = & |\mathtt{h}_1\, \mathtt{h}_2\, \mathtt{h}_3| + p_X\,|\mathsf{e}\, \mathtt{h}_2\, \mathtt{h}_3| + p_Y\,|\mathtt{h}_1\, \mathsf{e}\, \mathtt{h}_3| + p_Z\,|\mathtt{h}_1\, \mathtt{h}_2\, \mathsf{e}|
\end{array}$$

It follows from this expression that the coefficients of λ^0 and λ^3 of eq.A.9 are first order polynomials in $\pi_\infty = [p_X\, p_Y\, p_Z]^\top$. For λ^1 and λ^2 the derivation is a bit more tedious.

For the third order terms we still have two equal columns (ex. $p'_X p_Y p_Z |\mathtt{e}'\mathtt{e}\mathtt{e}|$) which means that this determinant vanishes. Some second order terms of the factorization do not vanish at first sight. These are the terms where both $\mathtt{e}$ and $\mathtt{e}'$ appear in the determinants. They can be grouped in pairs (ex. coefficient of λ):

$$\begin{array}{l} p_X p_Y (|\mathtt{e}\mathtt{e}'\mathtt{h}_3| + |\mathtt{e}'\mathtt{e}\mathtt{h}_3|) \\ p_X p_Z (|\mathtt{e}\mathtt{h}_2\mathtt{e}'| + |\mathtt{e}'\mathtt{h}_2\mathtt{e}|) \\ p_Y p_Z (|\mathtt{h}_1\mathtt{e}\mathtt{e}'| + |\mathtt{h}_1\mathtt{e}'\mathtt{e}|) \end{array}$$

All these terms vanish because permutating 2 rows of a determinant changes the sign of that determinant ($|\mathtt{e}'\mathtt{e}\mathtt{h}_3| = -|\mathtt{e}\mathtt{e}'\mathtt{h}_3|$).

This finally yields the following expressions for the 1^{st} order terms:

$$\begin{array}{l} |\mathtt{h}'_1\,\mathtt{h}_2\,\mathtt{h}_3| + |\mathtt{h}_1\,\mathtt{h}'_2\,\mathtt{h}_3| + |\mathtt{h}_1\,\mathtt{h}_2\,\mathtt{h}'_3| \\ +p_X(|\mathtt{e}'\,\mathtt{h}_2\,\mathtt{h}_3| + |\mathtt{e}\,\mathtt{h}'_2\,\mathtt{h}_3| + |\mathtt{e}\,\mathtt{h}_2\,\mathtt{h}'_3|) \\ +p_Y(|\mathtt{h}'_1\,\mathtt{e}\,\mathtt{h}_3| + |\mathtt{h}_1\,\mathtt{e}'\,\mathtt{h}_3| + |\mathtt{h}_1\,\mathtt{e}\,\mathtt{h}'_3|) \\ +p_Z(|\mathtt{h}'_1\,\mathtt{h}_2\,\mathtt{e}| + |\mathtt{h}_1\,\mathtt{h}'_2\,\mathtt{e}| + |\mathtt{h}_1\,\mathtt{h}_2\,\mathtt{e}'|) \end{array}$$

For the second order terms the accents should be inverted.

In conclusion the modulus constraint can be expressed as $l_3 l_1^3 = l_2^3 l_0$ with

$$\begin{array}{rl} l_3 = & - |\mathtt{h}'_1\,\mathtt{h}'_2\,\mathtt{h}'_3| - p_X|\mathtt{e}'\,\mathtt{h}'_2\,\mathtt{h}'_3| - p_Y|\mathtt{h}'_1\,\mathtt{e}'\,\mathtt{h}'_3| - p_Z|\mathtt{h}'_1\,\mathtt{h}'_2\,\mathtt{e}'| \\ l_2 = & \quad (|\mathtt{h}_1\mathtt{h}'_2\mathtt{h}'_3| + |\mathtt{h}'_1\mathtt{h}_2\mathtt{h}'_3| + |\mathtt{h}'_1\mathtt{h}'_2\mathtt{h}_3|) \\ & +p_X\ (|\mathtt{e}\,\mathtt{h}'_2\mathtt{h}'_3| + |\mathtt{e}'\mathtt{h}_2\mathtt{h}'_3| + |\mathtt{e}'\mathtt{h}'_2\mathtt{h}_3|) \\ & +p_Y\ (|\mathtt{h}_1\mathtt{e}'\mathtt{h}'_3| + |\mathtt{h}'_1\mathtt{e}\,\mathtt{h}'_3| + |\mathtt{h}'_1\mathtt{e}'\mathtt{h}_3|) \\ & +p_Z\ (|\mathtt{h}_1\mathtt{h}'_2\mathtt{e}'| + |\mathtt{h}'_1\mathtt{h}_2\mathtt{e}'| + |\mathtt{h}'_1\mathtt{h}'_2\mathtt{e}|) \\ l_1 = & -(|\mathtt{h}'_1\mathtt{h}_2\mathtt{h}_3| + |\mathtt{h}_1\mathtt{h}'_2\mathtt{h}_3| + |\mathtt{h}_1\mathtt{h}_2\mathtt{h}'_3|) \\ & -p_X\ (|\mathtt{e}'\mathtt{h}_2\mathtt{h}_3| + |\mathtt{e}\,\mathtt{h}'_2\mathtt{h}_3| + |\mathtt{e}\,\mathtt{h}_2\mathtt{h}'_3|) \\ & -p_Y\ (|\mathtt{h}'_1\mathtt{e}\,\mathtt{h}_3| + |\mathtt{h}_1\mathtt{e}'\mathtt{h}_3| + |\mathtt{h}_1\mathtt{e}\,\mathtt{h}'_3|) \\ & -p_Z\ (|\mathtt{h}'_1\mathtt{h}_2\mathtt{e}| + |\mathtt{h}_1\mathtt{h}'_2\mathtt{e}| + |\mathtt{h}_1\mathtt{h}_2\mathtt{e}'|) \\ l_0 = & \quad |\mathtt{h}_1\,\mathtt{h}_2\,\mathtt{h}_3| + p_X\,|\mathtt{e}\,\mathtt{h}_2\,\mathtt{h}_3| + p_Y\,|\mathtt{h}_1\,\mathtt{e}\,\mathtt{h}_3| + p_Z\,|\mathtt{h}_1\,\mathtt{h}_2\,\mathtt{e}| \quad . \end{array} \tag{A.10}$$

A.3 Expressions for l_3, l_2, l_1 and l_0 for a varying focal length

Here equation (5.15) is elaborated in detail:

$$\det\left(\mathbf{K}_f^{-1}\tilde{\mathbf{P}}_A - \lambda I\right) = l_3\lambda^3 + l_2\lambda^2 + l_1\lambda + l_0 = 0$$

yields with p_{ij} as the coefficients of the affine camera projection matrix $\mathbf{P}_A$

$$\begin{array}{rcl} l_3 & = & -1 \\ l_2 & = & (u_x p_{31} + u_y p_{32} + p_{33})f + p_{11} + p_{22} - u_x p_{31} - u_y p_{32} \\ l_1 & = & (u_x(p_{21}p_{32} - p_{31}p_{22}) + u_y(p_{31}p_{12} - p_{11}p_{32}) \\ & & \quad +p_{31}p_{13} + p_{32}p_{23} - p_{11}p_{33} - p_{22}p_{33})f \\ & & +u_x(p_{31}p_{22} - p_{21}p_{32}) + u_y(p_{11}p_{32} - p_{31}p_{12}) + p_{21}p_{12} - p_{11}p_{22} \\ l_0 & = & \left| \begin{array}{ccc} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{array} \right| f \end{array} \tag{A.11}$$

In this case it is interesting to analyze the solutions of the modulus constraint. The constraint $l_3(f)l_1(f)^3 = l_2(f)^3 l_0(f)$ was obtained by imposing equal moduli to the eigenvalues of $\mathbf{K}_f^{-1}\tilde{\mathbf{P}}_A$ (the *modulus* constraint). If f is a real solution then $-f$ will also be a solution. Changing the sign of the focal length is equivalent to a point reflection of the image around the principal point, which means that the moduli of the eigenvalues of $\mathbf{K}_f^{-1}\tilde{\mathbf{P}}_A$ will stay the same (only signs can change). *What does this mean for the coefficients of equation (5.16)?* We choose λ_1 and $-\lambda_1$ to be the real roots[1].

$$\begin{array}{rcl} a_4(f^2 - \lambda_1^2)(f^2 + bf + c) & = & 0 \\ a_4(f^4 + bf^3 + (c - \lambda_1^2)f^2 - \lambda_1^2 bf - \lambda_1^2 c) & = & 0 \end{array} \tag{A.12}$$

From equation (A.12) one can easily obtain λ_1 which is the desired solution for f_3.

$$f = \pm\sqrt{\frac{a_1}{a_3}} \tag{A.13}$$

Here a_1 and a_3 are the coefficients of the first order resp. third order term of equation (A.12). These coefficients are obtained by filling in $l_1(f)$, $l_2(f)$, $l_3(f)$, $l_4(f)$ from equation (A.11) in equation (A.8).

---

[1] A different real root λ_2 would imply $-\lambda_2$ to be a solution too. This would lead to $b = 0$ and thus also $p_Z = 0$ and $a_1 = 0$ in eq.(5.16). In practice we only encountered 4 real roots for pure translation. Three were identical and one had opposite sign.

Appendix B

Self-calibration from rectangular pixels

In this appendix the proof of Theorem 6.1 is given. Before starting the actual proof a lemma will be given. This lemma gives a way to check for the absence of skew from the coefficients of $\mathbf{P}$ directly without needing the factorization. A camera projection matrix can be factorized as follows $\mathbf{P} = [\mathbf{H}|\mathbf{e}] = \mathbf{K}\,[\mathbf{R}| - \mathbf{R}\mathbf{t}]$. In what follows $\mathbf{h}_i$ and $\mathbf{r}_i$ denote the rows of $\mathbf{H}$ and $\mathbf{R}$.

Lemma B.1 *The absence of skew is equivalent with* $(\mathbf{h}_1 \times \mathbf{h}_3)(\mathbf{h}_2 \times \mathbf{h}_3) = 0$.

Proof: It is always possible to factorize $\mathbf{H}$ as $\mathbf{K}\mathbf{R}$. Therefore the following can be written:

$$\begin{array}{l} (\mathbf{h}_1 \times \mathbf{h}_3)(\mathbf{h}_2 \times \mathbf{h}_3) \\ = ((f_x\mathbf{r}_1 + s\mathbf{r}_2 + u\mathbf{r}_3) \times \mathbf{r}_3)((f_y\mathbf{r}_2 + v\mathbf{r}_3) \times \mathbf{r}_3) \\ = ((f_x\mathbf{r}_1 + s\mathbf{r}_2) \times \mathbf{r}_3)(f_y\mathbf{r}_2 \times \mathbf{r}_3) \\ = -f_x f_y \mathbf{r}_2\mathbf{r}_1 + s f_y \mathbf{r}_1\mathbf{r}_1 = s f_y \ . \end{array}$$

Because $f_y \neq 0$ this concludes the proof. □

Equipped with this lemma the following theorem can be proven.

Theorem 6.1 *The class of transformations which preserves the absence of skew is the group of similarity transformations.*

Proof: It is easy to show that the similarity transformations preserve the calibration matrix $\mathbf{K}$ and hence also the orthogonality of the image plane:

$$\begin{array}{l} \mathbf{K}[\mathbf{R}| - \mathbf{R}\mathbf{t}] \left[\begin{array}{cc} \mathbf{R}' & \sigma^{-1}\mathbf{t}' \\ 0 & \sigma^{-1} \end{array} \right] \\ = \mathbf{K}[\mathbf{R}\mathbf{R}'|\sigma^{-1}(\mathbf{R}\mathbf{t}' - \mathbf{R}\mathbf{t})] \ . \end{array}$$

Therefore it is now sufficient to prove that the class of transformations which preserve the condition $(\mathbf{h}_1 \times \mathbf{h}_3)(\mathbf{h}_2 \times \mathbf{h}_3) = 0$ is at most the group of similarity transformations. To do this a specific set of positions and orientations of cameras can be chosen, since the absence of skew is supposed to be preserved for *all possible* views. In general $\mathbf{P}$ can be transformed as follows:

$$\mathbf{P}' = [\mathbf{H}|\mathbf{e}] \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d \end{bmatrix} = \left[\mathbf{HA} + \mathbf{ec}^\top | \mathbf{Hb} + \mathbf{e}d\right]$$

If $\mathbf{t} = 0$ then $\mathbf{H}' = \mathbf{KRA}$ and thus

$$\begin{aligned} &(\mathbf{h}'_1 \times \mathbf{h}'_3)(\mathbf{h}'_2 \times \mathbf{h}'_3) \\ &= ((f_x\mathbf{r}_1 + u\mathbf{r}_3)\mathbf{A} \times \mathbf{r}_3\mathbf{A})\,((f_y\mathbf{r}_2 + v\mathbf{r}_3)\mathbf{A} \times \mathbf{r}_3\mathbf{A})\,. \end{aligned}$$

Therefore the condition of the lemma is equivalent with

$$(\mathbf{r}_1\mathbf{A} \times \mathbf{r}_3\mathbf{A})(\mathbf{r}_2\mathbf{A} \times \mathbf{r}_3\mathbf{A}) = 0 \ .$$

Choosing for the rotation matrices $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$, rotations of 90^o around the x-, y- and z-axis, imposes the following equations to hold:

$$\begin{aligned} (\mathbf{a}_1 \times \mathbf{a}_2)(\mathbf{a}_3 \times \mathbf{a}_2) &= 0\,, \\ (\mathbf{a}_3 \times \mathbf{a}_1)(\mathbf{a}_2 \times \mathbf{a}_1) &= 0\,, \\ (\mathbf{a}_2 \times \mathbf{a}_3)(\mathbf{a}_1 \times \mathbf{a}_3) &= 0\,. \end{aligned} \tag{B.1}$$

Hence $(\mathbf{a}_1 \times \mathbf{a}_2)$, $(\mathbf{a}_1 \times \mathbf{a}_3)$ and $(\mathbf{a}_2 \times \mathbf{a}_3)$ define a set of 3 mutually orthogonal planes where $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ form the intersection and are therefore also orthogonal.

Choosing $\mathbf{R}_4$ and $\mathbf{R}_5$ as $\mathbf{R}_1$ and $\mathbf{R}_2$ followed by a rotation of 45^o around the z-axis, the following two equations can be derived:

$$\begin{aligned} ((\mathbf{a}_1 + \mathbf{a}_3) \times \mathbf{a}_2)\,((\mathbf{a}_1 - \mathbf{a}_3) \times \mathbf{a}_2) &= 0 \\ ((\mathbf{a}_3 + \mathbf{a}_2) \times \mathbf{a}_1)\,((\mathbf{a}_3 - \mathbf{a}_2) \times \mathbf{a}_1) &= 0 \ . \end{aligned} \tag{B.2}$$

Carrying out some algebraic manipulations and using $\mathbf{a_1} \perp \mathbf{a_2} \perp \mathbf{a_3}$ this yields the following result:

$$|\mathbf{a}_1|^2 = |\mathbf{a}_2|^2 = |\mathbf{a}_3|^2 \ .$$

These results mean that $\mathbf{A} = \sigma\mathbf{R}$ with σ a scalar and $\mathbf{R}$ an orthonormal matrix. The available constraints are not sufficient to impose $\det\mathbf{R} = 1$, therefore mirroring is possible.

Choose $\mathbf{R}_6 = \mathbf{R}_1$ and $\mathbf{t}_6^\top = [1\,0\,0]$, then $\left((\mathbf{a}_1 + \mathbf{c}^\top) \times \mathbf{a}_2\right)(\mathbf{a}_3 \times \mathbf{a}_2) = 0$ must hold. Using (B.1) and $\mathbf{a}_2 \times \mathbf{a}_3 \sim \mathbf{a}_1$ this condition is equivalent with $(\mathbf{c} \times \mathbf{a}_2)\mathbf{a}_1 = 0$. Writing $\mathbf{c}$ as $c_1\mathbf{a_1} + c_2\mathbf{a_2} + c_3\mathbf{a_3}$ this boils down to $c_3 = 0$. Taking $\mathbf{R}_7 = \mathbf{R}_2$, $\mathbf{t}_7 = [0\,0\,1]^\top$, $\mathbf{R}_8 = \mathbf{R}_3$ and $\mathbf{t}_8 = [0\,1\,0]^\top$ leads in a similar way to $c_2 = 0$ and $c_1 = 0$ and therefore to $\mathbf{c}^\top = [0\,0\,0]$.

In conclusion the transformation $\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{c}^\top & d \end{bmatrix}$ is restricted to the following form $\begin{bmatrix} \sigma\mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}$ which concludes the proof. □

Remark that 8 views were needed in this proof. This is consistent with the counting argument of Section 6.2.1.

Appendix C

Planar sections of imaginary cones

In this appendix it is shown how planar sections of imaginary cones can be represented through real entities. This makes it possible to use (classical) geometric insight to deal with complex problems which include imaginary objects.

We define the mapping between an imaginary ellipse $\mathbf{E}_I$ and a real ellipse $\mathbf{E}_R$ as follows:

$$\mathbf{E}_I = \mathbf{H}_A^\top \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{H}_A \leftrightarrow \mathbf{E}_R = \mathbf{H}_A^\top \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{H}_A \tag{C.1}$$

where $\mathbf{T}_A$ represents an arbitrary 2D affine transformation. Let us also define the mapping between an imaginary cone and a family of real ellipsoids:

$$\mathbf{C}_I = \mathbf{T}_A^\top \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{T}_A \leftrightarrow \mathbf{C}_R(\lambda) = \mathbf{T}_A^\top \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\lambda^2 \end{bmatrix} \mathbf{T}_A \tag{C.2}$$

where $\mathbf{T}_A$ represents an arbitrary 3D affine transformation. Let us also define the ratio of intersection between a plane and an ellipsoid as 1 when the plane passes through the center of the ellipse, as 0 when it is tangent and linearly (with the distance between the planes) in between.

Theorem C.1 *The intersection of the imaginary cone $\mathbf{C}_I$ with plane Π can be obtained as $\mathbf{E}_I$ corresponding to $\mathbf{E}_R$ which is the intersection of the ellipsoid $\mathbf{C}_R(\lambda)$ where λ is determined so that the ratio of intersection is $\frac{1}{\sqrt{2}}$.*

Proof: Let us first prove this for the simple case where $\mathbf{C}_I : X^2 + Y^2 + Z^2 = 0$ and $\mathbf{\Pi} : Z = 1$:

$$\begin{array}{rcl} \mathbf{C}_I : X^2 + Y^2 + Z^2 = 0 & \leftrightarrow & \mathbf{C}_R : X^2 + Y^2 + Z^2 = \sqrt{2}^2 \\ \mathbf{E}_I : X^2 + Y^2 + 1 = 0 & \leftrightarrow & \mathbf{E}_R : X^2 + Y^2 + 1 = \sqrt{2}^2 \end{array} \tag{C.3}$$

All other relative positions of planes and cones can be achieved by applying affine transformations to the geometric entities defined above. Since tangency, intersections and ratios of lengths along some line are affine invariants it follows that the mapping is valid for all imaginary cones and all planes. □

This therefore gives us an easy way to visualize the planar intersection of imaginary cones. This is useful to get some intuitive insight in some type of critical motion sequences. See for example Figure 6.5 where these concepts were applied.

Appendix D

Generalized rectification

The traditional rectification scheme consists of transforming the image planes so that the corresponding space planes are coinciding. This is however not possible when the epipole is located in the image. Even when this is not the case the image can still become very big (i.e. if the epipole is close to the image). This is the reason why Roy, Meunier and Cox [137] proposed a cylindrical rectification scheme.

The procedure described in [137], however, is relatively complex and some important implementation details were not discussed. Many operations are performed in 3D while everything can be done in the images. Additionally, advantage was not taken from the fact that the two halves of an epipolar plane are completely separate (i.e. should never be matched). In fact in the uncalibrated case this could even give problems in some cases since they in fact assume that the cameras are oriented correctly, which is in general not guaranteed for a projective reconstruction.

Here we present a very simple algorithm for rectification which works optimally in all possible cases. It only requires the oriented fundamental matrix between the two images.

D.1 Oriented epipolar geometry

The epipolar geometry describes the relations that exist between two images. Every point in a plane that passes through both centers of projection will be projected in each image on the intersection of this plane with the corresponding image plane. Therefore these two intersection lines are said to be in epipolar correspondence.

This geometry can easily be recovered from the images as seen in Section 7.3.1. The epipolar geometry is described by the following equation:

$$\mathtt{m}'\mathbf{F}\mathtt{m} = 0 \tag{D.1}$$

where $\mathtt{m}$ and $\mathtt{m}'$ are homogenous representations of corresponding image points and $\mathbf{F}$ is the fundamental matrix. This matrix has rank two, the right and left nullspace corresponds to the epipoles $\mathtt{e}$ and $\mathtt{e}'$ which are common to all epipolar lines. The epipolar line corresponding to a point $\mathtt{m}$ is thus given by $\mathtt{l}' \sim \mathbf{F}\mathtt{m}$ with $\sim$ meaning equality up to a non-zero scale factor (a strictly positive scale factor when oriented geometry is used, see further).

Epipolar line transfer The transfer of corresponding epipolar lines is described by the following equations:

$$\mathtt{l}' \sim \mathbf{H}^{-\top}\mathtt{l} \text{ or } \mathtt{l} \sim \mathbf{H}^{\top}\mathtt{l}' \tag{D.2}$$

with $\mathbf{H}$ a homography for an arbitrary plane. As seen in Section 3.3.1 a valid homography can immediately be obtained from the fundamental matrix:

$$\mathbf{H} = [\mathtt{e}']_{\times}\mathbf{F} + \mathtt{e}'\pi^{\top} \tag{D.3}$$

with $\mathtt{l}$ a random vector for which $\det\mathbf{H} \neq 0$ so that $\mathbf{H}$ is invertible. If one disposes of camera projection matrices an alternative homography is easily obtained as (see equation (3.24)):

$$\mathbf{H}^{-\top} = \left(\mathbf{P}'^{\top}\right)^{\dagger}\mathbf{P}^{\top} \tag{D.4}$$

where † indicates the Moore-Penrose pseudo inverse.

Orienting epipolar lines The epipolar lines can be oriented such that the matching ambiguity is reduced to half epipolar lines instead of full epipolar lines. This is important when the epipole is in the image. This fact was ignored in the approach of Roy et al. [137].

Figure D.1 illustrates this concept. Points located in the right halves of the epipolar planes will be projected on the right part of the image planes and depending on the orientation of the image in this plane this will correspond to the right or to the left part of the epipolar lines. These concepts are explained more in detail in the work of Laveau [78] on oriented projective geometry (see also [52]).

In practice this orientation can be obtained as follows. Besides the epipolar geometry one point match is needed (note that 7 or more matches were needed anyway to determine the epipolar geometry). An oriented epipolar line $\mathtt{l}$ separates the image plane into a positive and a negative region:

$$f_{\mathtt{l}}(\mathtt{m}) = \mathtt{l}^{\top}\mathtt{m} \text{ with } \mathtt{m} = [x\,y\,1]^{\top} \tag{D.5}$$

Note that in this case the ambiguity on $\mathtt{l}$ is restricted to a strictly positive scale factor. For a pair of matching points $(\mathtt{m}, \mathtt{m}')$ both $f_{\mathtt{l}}(\mathtt{m})$ and $f_{\mathtt{l}'}(\mathtt{m}')$ should have the same sign . Since $\mathtt{l}'$ is obtained from $\mathtt{l}$ through equation (D.2), this allows to determine the sign of $\mathbf{H}$. Once this sign has been determined the epipolar line transfer is oriented. We take the convention that the positive side of the epipolar line has the positive region of the image to its right. This is clarified in Figure D.2.

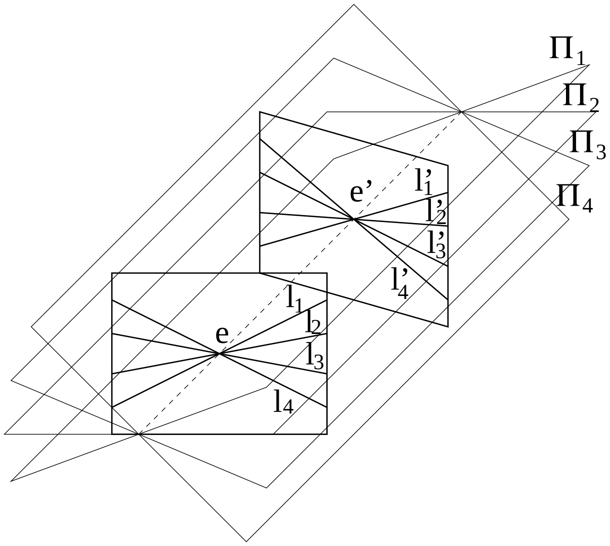

Figure D.1: *Epipolar geometry with the epipoles in the images. Note that the matching ambiguity is reduced to half epipolar lines.*

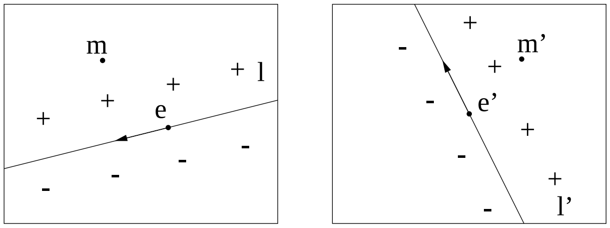

Figure D.2: *Orientation of the epipolar lines.*

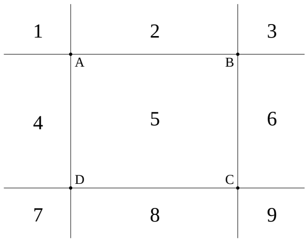

Figure D.3: *the extreme epipolar lines can easily be determined depending on the location of the epipole in one of the 9 regions. The image corners are given by* `A`, `B`, `C`, `D`.

D.2 Generalized Rectification

The key idea of our new rectification method consists of reparameterizing the image with polar coordinates (around the epipoles). Since the ambiguity can be reduced to half epipolar lines only positive longitudinal coordinates have to be taken into account. The corresponding half epipolar lines are determined through equation (D.2) taking orientation into account.

The first step consists of determining the common region for both images. Then, starting from one of the extreme epipolar lines, the rectified image is build up line by line. If the epipole is in the image an arbitrary epipolar line can be chosen as starting point. In this case boundary effects can be avoided by adding an overlap of the size of the matching window of the stereo algorithm (i.e. use more than 360 degrees). The distance between consecutive epipolar lines is determined independently for every half epipolar line so that no pixel compression occurs. This non-linear warping allows to obtain the minimal achievable image size without losing image information.

The different steps of this method are described more in detail in the following paragraphs.

Determining the common region Before determining the common epipolar lines the extremal epipolar lines for a single image should be determined. These are the epipolar lines that touch the outer image corners. The different regions for the position of the epipole are given in Figure D.3. The extremal epipolar lines always pass through corners of the image (e.g. if the epipole is in region 1 the area between `eB` and `eD`). The extreme epipolar lines from the second image can be obtained through the same procedure. They should then be transfered to the first image. The common region is then easily determined as in Figure D.4.

Determining the distance between epipolar lines To avoid losing pixel information the area of every pixel should be at least preserved when transformed to the rectified image. The worst case pixel is always located on the image border opposite to the epipole. A simple procedure to compute this step is depicted in Figure D.5. The same procedure can be carried out in the other image. In this case the obtained epipolar line

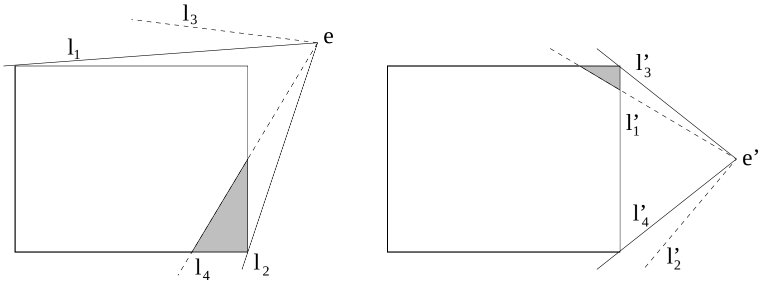

Figure D.4: *Determination of the common region. The extreme epipolar lines are used to determine the maximum angle.*

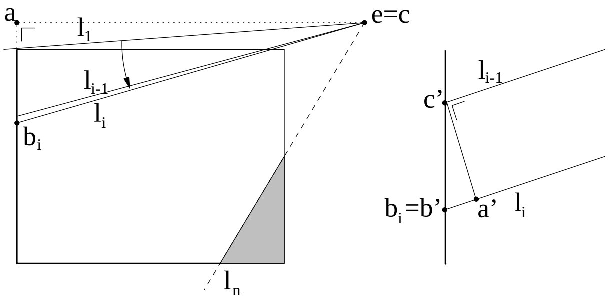

Figure D.5: *Determining the minimum distance between two consecutive epipolar lines. On the left a whole image is shown, on the right a magnification of the area around point* $\mathtt{b}_i$ *is given. To avoid pixel loss the distance* $|\mathtt{a}'\mathtt{c}'|$ *should be at least one pixel. This minimal distance is easily obtained by using the congruence of the triangles* `abc` *and* $\mathtt{a}'\mathtt{b}'\mathtt{c}'$*. The new point* `b` *is easily obtained from the previous by moving* $\frac{|\mathtt{bc}|}{|\mathtt{ac}|}$ *pixels (down in this case).*

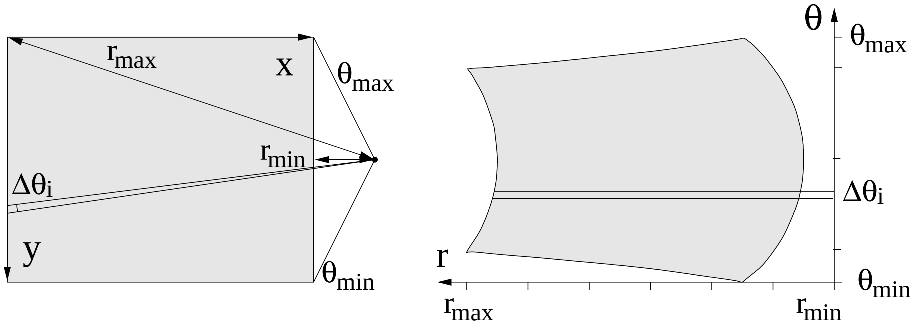

Figure D.6: *The image is transformed from (x,y)-space to (r,θ)-space. Note that the θ-axis is non-uniform so that every epipolar line has an optimal width (this width is determined over the two images).*

should be transferred back to the first image. The minimum of both displacements is carried out.

Constructing the rectified image The rectified images are build up row by row. Each row corresponds to a certain angular sector. The length along the epipolar line is preserved. Figure D.6 clarifies these concepts. The coordinates of every epipolar line are saved in a list for later reference (i.e. transformation back to original images). The distance of the first and the last pixels are remembered for every epipolar line. This information allows a simple inverse transformation through the constructed look-up table.

Note that an upper bound for the image size is easily obtained. The height is bound by the contour of the image $2 \times (W + H)$. The width is bound by the diagonal $\sqrt{W^2 + H^2}$. Note that the image size is uniquely determined with our procedure and that it is the minimum that can be achieved without pixel compression.

Transfering information back Information about a specific point in the original image can be obtained as follows. The information for the corresponding epipolar line can be looked up in the table. The distance to the epipole should be computed and substracted from the distance for the first pixel of the image row. The image values can easily be interpolated for higher accuracy.

To warp a complete image back a more efficient procedure than a pixel-by-pixel warping can be designed. The image can be reconstructed radially (i.e. radar like). All the pixels between two epipolar lines can then be filled in at once from the information that is available for these epipolar lines. This avoids multiple look-ups in the table.

Nederlandse samenvatting

Zelf-calibratie en metrische 3D reconstructie uit ongecalibreerde beeldsequenties

Inleiding

Het bekomen van drie-dimensionele (3D) modellen uit beelden is een uitdagend onderwerp. In het gebied van computervisie wordt dit onderwerp reeds vele jaren onderzocht. Er bestaan veel toepassingen die gebaseerd zijn op dit soort modellen. Vroeger werden vooral robotica en inspectie toepassingen aangepakt. In deze context is nauwkeurigheid meestal de belangrijkste factor en wordt meestal gebruik gemaakt van dure toestellen die onder welbepaalde –restrictieve– omstandigheden werken.

Tegenwoordig is er meer en meer interesse vanuit de multimedia- en computervisualisatie-gemeenschap. De evolutie van de computers is zo dat vandaag de dag zelfs een standaard persoonlijke computer probleemloos complexe 3D modellen kan visualiseren. Veel computerspelen worden in complexe 3D werelden geplaatst. Op Internet geraakt het gebruik van 3D modellen en omgevingen meer en meer ingeburgerd. Deze evolutie wordt echter vertraagd door de moeilijkheid om zulke 3D modellen te bekomen. Hoewel het gemakkelijk is om met behulp van modelleringspakketten interactief eenvoudige 3D modellen te genereren, vragen complexe scènes behoorlijk veel tijd. Daarenboven wenst men meestal bestaande objecten of scènes te modelleren. In dit geval vergt het bekomen van een realistische 3D model meestal erg veel moeite en is het resultaat vaak teleurstellend.

Er bestaat een groeiende vraag naar systemen die bestaande objecten of scènes kunnen *virtualiseren*. In dit geval zijn de vereisten echter heel verschillend van wat voor de vroegere, industriële, toepassingen vereist was. Het belangrijkste aspect is nu de visuele kwaliteit van het 3D model. Ook de randvoorwaarden zijn verschillend. Er is een grote vraag naar eenvoudige acquisitie procedures die gebruik maken van standaard foto- en videoapparatuur. Dit verklaart het succes van de QuickTime VR technologie die eenvoudige acquisitie met snelle weergave combineert. Let wel dat in dit geval geen 3D informatie wordt opgenomen en dat men slechts kan *rondkijken* en niet *rondwandelen*.

In dit werk werd onderzocht hoe ver men kon gaan in het ontwerp van eenvoudige en flexibele procedures voor automatische acquisitie van realistische 3D modellen. Hiertoe werd een systeem ontworpen dat in staat is om getextureerde metrische 3D modellen uit sequenties van beelden te halen die met een met de hand vastgehouden camera werden opgenomen. Omwille van de tijdsbeperking van het project moesten een aantal keuzes gemaakt worden. Het systeem werd opgebouwd door een aantal geavanceerde algoritmes te combineren met een aantal nieuwe componenten die in dit project ontworpen werden. Sommige algoritmes uit de huidige stand van zaken werden aangepast of uitgebreid om in het systeem te passen.

In de onderzoeksgemeenschap werd heel wat moeite gedaan om vanuit een ongecalibreerde beeldsequentie de calibratie van de camera-opstelling te bekomen tot op een willekeurige projectieve transformatie en reeds vele jaren gebeurt er heel wat werk om dichte correspondentie-kaarten te bekomen voor gecalibreerde camera-opstellingen. Er was echter een missende schakel. Hoewel de mogelijkheid tot zelf-calibratie (i.e. het beperken van de ambiguïteit van projectief tot metrisch) aangetoond was, gaven practische algoritmes geen bevredigende resultaten. Daarenboven bleven bestaande theorie en algoritmes beperkt tot constante intrinsieke camera-parameters. Daardoor was het niet mogelijk om gebruik te maken van de zoom en focus mogelijkheden die aanwezig zijn op de meeste camera's.

In deze context besliste ik om mij te concentreren op het zelf-calibratie aspect. Algoritmes die goede resultaten gaven op echte beeldsequenties waren vereist. Het bleek ook interessant om te onderzoeken in welke mate door deze algoritmes rekening kon gehouden worden met variërende camera parameters, bv. een varierende focale lengte zodat zoom en focus konden gebruikt worden.

Projectieve meetkunde

Het werk dat in deze thesis wordt gepresenteerd steunt op meetkunde en, meer bepaald, op projectieve meetkunde. Dit is namelijk het natuurlijke kader om het beeldvormingsproces in te beschrijven. Verschillende geometrische entiteiten, zoals punten, lijnen, vlakken, kegelsneden en kwadrieken, worden gedefinieerd. Het effect van transformaties op deze entiteiten wordt besproken. Een aantal andere eigenschappen worden eveneens afgeleid.

De stratificatie van 3D meetkunde

Heel wat aandacht gaat naar de stratificatie van de meetkunde. Hiermee wordt bedoeld dat de euclidische meetkunde kan gezien worden als bestaande uit een projectief, affien, metrisch en euclidisch stratum. Projectieve meetkunde wordt gebruikt omwille van de eenvoud van het formalisme. Bijkomende structuur en eigenschappen kunnen dan worden toegevoegd aan de hand van deze hiërarchie van meetkundige strata.

Het concept van stratificatie is rechtstreeks gerelateerd met de transformatie-groepen die ageren op geometrische entiteiten en die eigenschappen van configuraties van deze elementen onveranderd laten. Met het projectieve stratum komt de groep van projectieve transformaties overeen, met het affiene stratum is dat de groep van affiene transformaties, met het metrische stratum de groep van de similariteiten en met het Euclidische stratum de groep van Euclidische transformaties. Het is belangrijk om op te merken dat deze groepen subgroepen zijn van elkaar, bv. de metrische groep is een subgroep van de affiene groep, beiden zijn dan weer subgroepen van de projectieve groep.

Een belangrijk aspect gerelateerd met deze groepen zijn de respectievelijke invarianten. Een *invariant* is een eigenschap van een configuratie van geometrische entiteiten die niet verandert onder invloed van transformaties die tot de bijhorende groep behoren. Invarianten komen dus overeen met metingen die men kan doen in een welbepaald stratum van de geometrie. Deze invarianten zijn dikwijls gerelateerd tot entiteiten die –als geheel– onveranderd blijven onder transformaties van deze specifieke groep. Deze entiteiten spelen een cruciale rol in dit werk. Het terugvinden ervan laat immers toe om de structuur van de geometrie tot op een hoger niveau van de stratificatie te brengen.

In de volgende paragrafen worden de verschillende strata afzonderlijk besproken. De geassocieerde transformatie-groepen, hun invarianten en de overeenkomstige invariante entiteiten worden voorgesteld.

Projectief stratum Het eerste stratum is het projectieve stratum. Deze bezit de minste structuur en heeft dus het minste invarianten en de meest algemene transformatiegroep. De groep van projectieve transformaties of collineaties is de meest algemene groep van lineaire transformaties. Een projectieve transformatie heeft 15 vrijheidsgraden.

Het samenvallen, het raken en de collineariteit zijn projectief invariant. Verder is ook de kruisverhouding van vier collineaire punten (of de equivalenten voor rechten en vlakken) invariant voor de groep van projectieve transformaties.

Affien stratum Het volgende stratum in deze stratificatie is het affiene stratum. Dit stratum heeft meer structuur dan de projectieve maar minder dan de metrische of Euclidische strata. Affiene meetkunde verschilt van projectieve meetkunde door het identificeren van een specifiek vlak, nl. het *vlak op oneindig*. Daar een affiene transformatie het vlak op oneindig op zichzelf afbeeldt, heeft het slechts 12 vrijheidsgraden.

Alle projectieve invarianten zijn a fortiori affiene eigenschappen. Voor de –meer restricitieve– affiene groep wordt parallellisme toegevoegd als een nieuwe invariante eigenschap. Lijnen of vlakken die hun doorsnede in het vlak op oneindig hebben worden als *parallel* aanzien. De *verhouding van lengtes volgens een bepaalde richting* vormt een nieuwe invariante eigenschap voor dit stratum.

Om vertrekkende van een projectieve voorstelling van een scène tot een affiene voorstelling te komen moet men dus het vlak op oneindig localiseren. Daar waar vertrekkende van een Euclidische voorstelling de positie van het vlak op oneindig gekend is, is dat in het algemeen niet zo wanneer men slechts beschikt over een projectieve voorstelling. Het kennen van een aantal affiene eigenschappen van de scène (bv. parallelle lijnen) laat echter wel toe om de positie van dit vlak terug te vinden.

Metrisch stratum Het metrische stratum komt overeen met de groep van de similariteiten. Deze transformaties komen overeen met Euclidische transformaties aangevuld met een schaal aanpassing. Wanneer geen absolute metingen mogelijk zijn, is dit het hoogste niveau van geometrische structuur dat uit beelden kan gehaald worden. In de filmindustrie wordt dankbaar gebruik gemaakt van deze eigenschap. Ze laat immers toe schaalmodellen aan te wenden voor speciale effecten. Een metrische transformatie heeft 7 vrijheidsgraden (3 voor orientatie, 3 voor translatie en 1 voor schaal).

In dit geval zijn er twee nieuwe invariante eigenschappen: *relatieve lengtes* en *hoeken*. Zoals in het affiene geval, zijn deze nieuwe invariante eigenschappen gerelateerd met een invariante geometrische entiteit. Buiten het invariant laten van het vlak op oneindig, laat een metrische transformatie eveneens een specifieke kegelsnede invariant, nl. de *absolute kegelsnede*. De overeenkomstige duale kegelsnede en duale kwadriek zijn eveneens invariant. Naargelang het probleem kan aan de ene of de andere voorstelling de voorkeur gegeven worden.

Om de metrische structuur van een scène te bekomen uit de projectieve of affiene structuur moet men beschikken over een aantal metrische metingen om de absolute kegelsnede te localiseren. Vermits deze kegelsnede in het vlak op oneindig ligt kan men deze eenvoudiger bekomen als dit vlak reeds voordien gelocaliseerd werd.

Euclidisch stratum Om volledig te zijn wordt Euclidische meetkunde ook kort besproken. Veel verschil is er niet met de metrische groep. Het verschil is dat de absolute schaal vast ligt en dat dus niet enkel relatieve lengtes, maar ook *absolute lengtes* invariant zijn. Een Euclidische transformatie heeft 6 vrijheidsgraden, 3 voor orientatie en 3 voor translatie.

Cameramodel en verband tussen meerdere beelden

Alvorens te bespreken hoe 3D informatie uit beelden kan bekomen worden is het belangrijk om te weten hoe beelden worden gevormd. Eerst wordt het gebruikte cameramodel voorgesteld en dan worden een aantal belangrijke verbanden tussen meerdere zichten gegeven.

Cameramodel

In dit werk wordt het perspectief cameramodel gebruikt. Het beeldvormingsproces wordt volledig bepaald door het vastleggen van het projectie centrum en het retinaal vlak. De projectie van een 3D punt wordt dan bepaald als het snijpunt van de lijn die het 3D punt verbindt met het projectiecentrum en het retinaal vlak. De coördinaten in het beeld hangen dan af van de parametrisatie van het retinaal vlak. De meeste camera's worden relatief goed beschreven door dit model. In sommige gevallen moeten bijkomende effecten in rekening worden gebracht (bv. radiale distortie).

Een perspectiefcamera wordt door een aantal parameters beschreven. Deze bestaan uit twee categorieën. De extrinsieke camera parameters en de intrinsieke camera parameters De extrinsieke parameters beschrijven de positie en orientatie van de camera.

De intrinsieke parameters beschrijven het projectieproces zelf. De belangrijkste parameter is de *focale lengte* (de afstand tussen het projectie centrum en het beeldvlak uitgedrukt in pixelafmetingen). Verder zijn er ook de coördinaten van het *focaal punt* (het punt van het beeldvlak dat het dichtst bij het projectiecentrum ligt) en de vorm van een pixel (meestal vierkant, maar kan ook een rechthoek of zelfs een parallellogram zijn). Deze parameters worden meestal samengebracht in een 3×3 bovendriehoeksmatrix $\mathbf{K}$.

projectie matrix Gebruik makende van homogene coördinaten kan het projectieproces volledig beschreven worden m.b.v. een 3×4 projectie matrix. De bovenvermelde parameters kunnen hieruit bekomen worden door QR-decompositie. In dit werk worden eveneens een aantal interessante verbanden tussen de camera projectie matrix en homografieën tussen vlakken besproken.

Verband tussen meerdere zichten

Er bestaan heel wat interessante verbanden tussen meerdere zichten van een scène. Een aantal van deze verbanden worden besproken in de volgende paragrafen.

Twee zichten Voor een perspectiefcamera moet het 3D punt dat overeenkomt met een bepaald beeldpunt, gelegen zijn op de rechte die het beeldpunt verbindt met het projectiecentrum. In een tweede zicht zal het overeenkomstige beeldpunt gelegen zijn op de projectie van deze rechte.

In feite moeten de projecties van alle punten gelegen in een vlak dat door beide projectiecentra gaat, gelegen zijn op de intersectie van dit vlak met het respectievelijke beeldvlak. Van deze rechten zegt men dat ze in *epipolaire correspondentie* verkeren. Door van deze eigenschap gebruik te maken kan het correspondentie probleem tussen twee zichten tot één dimensie herleid worden.

Wiskundig worden deze verbanden beschreven door de *fundamentele matrix* Deze matrix geeft voor elk beeldpunt de overeenkomstige epipolaire lijn in het andere beeld (nl. de lijn waar het overeenkomstige punt op gelegen moet zijn). Deze matrix kan dus gebruikt worden om het zoeken van corresponderende punten te vergemakkelijken. Omgekeerd kunnen een aantal corresponderende punten gebruikt worden om deze matrix te bepalen. Hiervoor volstaan 7 puntcorrespondenties (nl. het aantal vrijheidgraden van de fundamentele matrix). Deze matrix vertoont ook een aantal interessante verbanden met de homografieën voor vlakken. Deze worden eveneens kort besproken in dit werk.

Drie zichten Wanneer drie zichten beschouwd worden, kan men natuurlijk beeldparen vormen en de epipolaire beperkingen gebruiken. Een beeldpunt in een derde beeld kan dan voorspeld worden als de intersectie van beide epipolaire lijnen in het derde beeld. Deze strategie werkt echter niet voor punten die coplanair zijn met de drie projectie centra. Na expliciete reconstructie aan de hand van de beeldpunten in de eerste twee beelden kan men de projectie in het derde beeld echter wel bekomen. Dit betekent dus dat niet alle verbanden tussen drie beelden beschreven worden door de fundamentele matrix. De projectie van een rechte in een beeld kan eveneens voorspeld worden uit de projecties in twee andere beelden.

Al deze verbanden worden wel beschreven door de *trifocal tensor*. Dit is een $3 \times 3 \times 3$ tensor met 18 vrijheidsgraden. Deze kan bepaald worden uit 6 of meer corresponderende punten tussen drie zichten. Lijnen geven ook beperkingen die toelaten om de trifocale tensor te bepalen.

Meerdere zichten Een beeldpunt heeft twee vrijheidgraden. n afbeeldingen van hetzelfde 3D punt hebben echter geen $2n$ vrijheidsgraden, maar slechts 3. Er moeten dus $2n-3$ onafhankelijke beperkingen bestaan tussen deze afbeeldingen. Voor lijnen, die eveneens 2 vrijheidsgraden hebben in een beeld, maar 4 in de ruimte, moeten n projecties aan $2n-4$ beperkingen voldoen.

Zelf-calibratie

Een van de belangrijkste bijdragen van dit werk situeert zich op het vlak van zelf-calibratie. Alvorens de specifieke algoritmes die ontwikkeld werden, te bespreken, worden de algemene concepten aangebracht en worden een aantal methodes voorgesteld die door anderen ontwikkeld werden.

Projectieve ambiguïteit

Veronderstel dat een aantal beelden van een statische scène gegeven zijn. Als de calibratie, positie en orientatie gekend zijn, is het mogelijk om de geobserveerde punten te reconstrueren. Twee (of meer) corresponderende beeldpunten zijn hiervoor voldoende. Deze reconstructie wordt bekomen als het snijpunt van de rechten die overeenkomen met de respectievelijke beeldpunten. Deze reconstructie is eenduidig bepaald in de ruimte.

In het ongecalibreerde geval zijn zowel calibratie, orientatie en positie van de camera onbekend. In het meest algemene geval heeft men dus geen enkele beperking op de projectiematrix. Daar de scène eveneens als ongekend beschouwd wordt, zijn er ook geen beperkingen voor de positie van de 3D punten. Voor een bepaalde reconstructie die compatibel is met de geobserveerde punten geldt dat elke projectieve transformatie van deze reconstructie eveneens een geldige recontructie is. Zonder bijkomende beperkingen is de reconstructie dus enkel bepaald tot op een willekeurige projectieve transformatie na. Dit wordt een *projectieve reconstructie* genoemd.

Alhoewel deze reconstructie voldoende kan zijn voor een aantal toepassingen, is deze voor heel wat andere toepassingen niet bruikbaar. Voor visualisatie heeft men bijvoorbeeld minstens een metrische reconstructie nodig. Om de reconstructie van projectief naar metrisch te brengen, zijn ofwel een aantal metrische eigenschappen van de scène vereist, ofwel een aantal beperkingen op de calibratie van de camera. Dit laatste kan bestaan uit een aantal beperkingen op de intrinsieke of extrinsieke camera parameters.

Calibratie

Standaard calibratie technieken zijn gebaseerd op Euclidische of metrische kennis van de scène of camera, of op de kennis van de beweging van de camera. Een eerste mogelijkheid bestaat erin om eerst een projectieve reconstructie te bekomen en deze dan a posteriori te transformeren naar een metrische (of zelfs Euclidische) reconstructie. De traditionele aanpak bestaat er echter in om rechtstreeks een metrische (of Eucldische) reconstructie te bepalen.

Zelf-calibratie

In veel gevallen zijn de specifieke waarden voor de intrinsieke of extrinsieke camera parameters niet gekend. Meestal zijn deze parameters echter ook niet volledig vrij. Deze beperkingen kunnen gebruikt worden om een metrische calibratie van de camera opstelling te bekomen. Dit wordt *zelf-calibratie* genoemd. Het traditionele zelf-calibratie probleem is meer beperkt. In dit geval veronderstelt men dat alle intrinsieke parameters constant blijven. De exacte waarden voor de intrinsieke parameters zijn echter niet gekend en de beweging van de camera is niet beperkt. Dit komt overeen met een ongekende camera die vrij wordt verplaatst (bv. manueel vastgehouden). Dit probleem werd door heel wat onderzoekers bestudeerd.

In een aantal belangrijke gevallen is de beweging van de camera beperkt. Deze extra informatie kan gebruikt worden om eenvoudiger algoritmes te ontwikkelen. In dit geval is het echter niet steeds mogelijk om alle parameters te bepalen, vermits beperkte bewegingssequenties niet steeds genoeg informatie opleveren voor zelf-calibratie. Een aantal interessante gevallen die verder besproken worden zijn: pure translatie, pure rotatie en vlakke beweging.

In een aantal gevallen waar de beweging niet algemeen genoeg is, kan het zijn dat zelf-calibratie niet in staat zijn om de ambiguïteit op de reconstructie te beperken tot metrisch. Dit probleem van kritische bewegingssequenties wordt eveneens besproken. Een aantal nieuwe resultaten worden gepresenteerd.

Algemene beweging

Er bestaan heel wat methodes voor zelf-calibratie. Deze kunnen echter eenvoudig in een aantal klassen onderverdeeld worden. Een eerste klasse vertrekt van een projectieve reconstructie en probeert de absolute kegelsnede te indentificeren als de enige kegelsnede die aan een aantal beperkingen beantwoord. Typisch moeten de projecties van deze kegelsnede identiek zijn voor alle beelden, daar deze rechtstreeks gerelateerd zijn met de intrinsieke camera parameters die constant verondersteld worden.

Een tweede klasse methodes is eveneens gebaseerd op de absolute kegelsnede, maar de vergelijkingen worden beperkt tot de epipolaire geometrie. Het voordeel is dat enkel de fundamentele matrices nodig zijn. Anderzijds vertoont deze methodes ook een aantal belangrijke nadelen.

Naast deze twee klassen van methodes bestaan er eveneens een aantal methodes die de projectie matrices factorizeren en rechtstreeks opleggen dat de intrinsieke camera parameters constant moeten zijn.

Beperkte bewegingen

Zoals reeds gezegd, kunnen beperkte bewegingen voor zelf-calibratie heel interessant zijn. Deze bewegingen kunnen resulteren in eenvoudiger algoritmes. Anderzijds is het dan niet steeds mogelijk om alle parameters eenduidig te bepalen. Een paar interessante voorbeelden worden verder uitgewerkt.

Pure translatie In het geval van pure translatie blijft het vlak op oneindig niet enkel als geheel onveranderd, maar ook punt per punt. Dit laat toe om zeer eenvoudig dit vlak terug te vinden en zo onmiddelijk een affiene reconstructie te bekomen. Vermits in dit geval alle kegelsneden in het vlak op oneindig constant blijven, is het niet mogeliijk om een metrische reconstructie te bekomen.

Pure rotatie In dit geval hangt de verplaatsing van de beeldpunten enkel af van de rotatie van de camera en niet van de afstand van de punten tot de camera. Daardoor is er geen onderscheid tussen punten van het vlak op oneindig en andere. Alle beeldpunten ondergaan dezelfde homografie. Deze homografie is dus ook geldig voor het vlak op oneindig zodat men eenvoudige beperkingen krijgt voor de intrinsieke camera parameters. Merk echter wel op dat in dit geval geen enkele 3D informatie kan bekomen worden en dat deze strategie dus enkel kan dienen om de parameters van de camera te bepalen en niet om een scene te reconstrueren.

Vlakke beweging Met een vlakke beweging wordt een beweging bedoeld waarvoor alle verplaatsingen tot een welbepaald vlak beperkt blijven en alle rotatiesassen loodrecht op dit vlak staan. In dit geval zijn er ook een aantal entiteiten die onveranderd blijven. In het vlak op oneindig is er de horizon (nl. de doorsnede van het bewegingsvlak met het vlak op oneindig) en het vluchtpunt van de rotatieas. Door deze entiteiten te bepalen kan men redelijk eenvoudig de positie van het vlak op oneindig bekomen. Uiteindelijk kan men op één parameter na een metrisch reconstructie bekomen.

Kritische bewegingssequenties

Zoals reeds gebleken is uit de vorige paragrafen, kan een beperkte bewegingssequentie tot gevolg hebben dat de metrische structuur van de scène niet volledig bepaald kan worden. In het algemeen wordt zelf-calibratie bekomen door de kegelsnede terug te vinden die dezelfde projectie heeft in alle beelden. Enkel de absolute kegelsnede beantwoordt in het algemeen aan deze beperking. Omgekeerd kan het echter wel zijn dat voor een specifieke beweging meerdere kegelsneden hieraan beantwoorden. In dat geval is het niet mogelijk om een onderscheid te maken tussen de echte absolute kegelsnede en andere mogelijke kandidaten.

Recent werd een volledige classificatie afgeleid van de verschillende types van beweging die aanleiding geven tot meer dan één potentiële absolute kegelsnede. Het basisidee bestaat erin om het probleem om te draaien en te zien welke bewegingen het beeld van een welbepaalde kegelsnede onveranderd laten.

Deze classificatie kan enerzijds preventief gebruikt worden. Men kan ervoor zorgen dat de beweging van de camera bij de opnames niet beperkt blijft tot één van de kritische bewegingssequenties.

Anderzijds werd ook een theorema afgeleid dat toelaat om na zelf-calibratie te verifiëren of men al dan niet te maken had met een kritische bewegingssequentie. In het kader van mijn werk werd een eenvoudiger bewijs afgeleid voor deze stelling. De stelling luidt als volgt:

Theorema 4.1 *Laat S een bewegingssequentie voorstellen die kritisch is ten aanzien van de duale kwadriek* $\mathbf{\Phi}^*$*, en laat* $\mathbf{P}_{Ei}$ *de originele camera projectiematrices zijn voor de verschillende zichten van S. Laat* $\mathbf{T}$ *een projectieve transformatie zijn die* $\mathbf{\Phi}^*$ *op de absolute duale kwadriek* $\mathbf{\Omega}^*$ *afbeeldt en, laat* $\mathbf{P}_{Pi} = \mathbf{P}_{Ei}\mathbf{T}^{-1}$ *de projectie matrices zijn na transformatie door* $\mathbf{T}$*. Dan bestaat er een Euclidische transformatie tussen elke twee* $\mathbf{P}_{Pi}$*.*

Hieruit kan men afleiden dat de bekomen beweging eveneens een kritische beweging is van dezelfde klasse als de oorspronkelijke beweging. Daardoor is het mogelijk om uit om het even welke potentieele metrische reconstructie het type –al dan niet kritische– bewegingssequentie te bepalen. Hierdoor kan men dan ook de ambiguïteit op de reconstructie bepalen.

Een andere interessante vraag die nog niet beantwoord was, is: “*Wat kunnen we doen met een ambigue reconstructie?*”. Het volgende theorema, dat ik heb afgeleid, geeft hier een antwoord op. In het volgende theorema wordt met $C(S)$ de verzameling van potentieele absolute kwadrieken voor de bewegingssequentie S bedoeld.

Theorema 4.2 *Laat S een kritische bewegingssequentie zijn en laat* $\mathbf{P}_{Ei}$ *de overeenkomstige projectie matrices zijn. Laat* $\mathbf{\Phi}^*$ *een willekeurig element zijn van* $C(S)$ *en laat* $\mathbf{T}$ *een willekeurige projectieve transformatie zijn die* $\mathbf{\Phi}^*$ *op* $\mathbf{\Omega}^*$ *afbeeldt. Laat* S_P *de door* $\mathbf{T}$ *getransformeerde bewegingssequentie voorstellen en laat* $\mathbf{P}_{Pi} = \mathbf{P}_{Ei}\mathbf{T}^{-1}$*. Laat* M *een Euclidische beweging voorstellen waarvoor* $C(S_P \bigcup M) = C(S_P)$ *en laat* $\mathbf{P}_{Pnew}$ *de corresponderende projectie matrix zijn. Dan bestaat er een Euclidische transformatie tussen* $\mathbf{P}_{Enew} = \mathbf{P}_{Pnew}\mathbf{T}$ *en elke andere* $\mathbf{P}_{Ei}$*.*

Dit laat dus toe om zelf in het geval van een ambigue reconstructie nieuwe zichten te genereren zonder distorties. Dit kan namelijk wanneer de beweging beperkt blijft tot de specifieke kritische bewegingssequentie. Bijvoorbeeld als een model bekomen werd door een vlakke beweging uit te voeren met de camera, kan men correcte nieuwe zichten genereren voor zover men de translaties van de virtuele camera in het vlak uitvoert en de rotaties rond assen loodrecht erop laat gebeuren.

Gestratifieerde zelf-calibratie

De laatste jaren werden heel wat methodes voorgesteld om de calibratie van een camera te bekomen uit correspondenties tussen verschillende beelden van dezelfde scène. Deze methodes zijn gebaseerd op de rigiditeit van de scène en op het constant zijn van de intrinsieke camera parameters. De meeste bestaande technieken vertrekken van een projectieve reconstructie en proberen onmiddelijk de intrinsieke camera parameters te bepalen. Al deze methodes moeten echter op één of andere manier rekening houden met het affiene stratum van de geometrie (nl. de positie van het vlak op oneindig).

Bij de Kruppa vergelijkingen wordt de positie van het vlak op oneindig uit de vergelijkingen geëlimineerd. Deze oplossing vertoont een aantal belangrijke gebreken. De meeste andere technieken trachten de onbekende parameters van het affiene en metrische stratum tegelijkertijd uit de zelf-calibratie vergelijkingen op te lossen. Dit resulteert meestal in een complex optimisatieprobleem dat niet altijd convergeert.

Dit probleem zette ons aan om een gestratifieerde aanpak uit te bouwen. In dit geval bepaalt men, vertekkende van een projectieve reconstructie, eerst het affiene stratum en gaat men dan pas ver tot het metrische stratum. Een gelijkaardige aanpak werd reeds voorgesteld voor het geval men beschikt over een pure translatiebeweging waaruit rechtstreeks het affiene stratum kan bepaald worden. In het algemeen is zulke beweging echter moeilijk te garanderen. Succesvolle gestratifieerde strategieën werden eveneens voorgesteld voor de zelf-calibratie van vaste stereo-opstellingen.

In dit werk werd een gestratifieerde zelf-calibratiemethode uitgewerkt die niet enkel zeer goede resultaten oplevert voor experimenten op synthetische gegevens, maar ook op echte beeldsequenties die met een hand-gehouden videocamera opgenomen werden. Deze methode wordt in de volgende paragrafen voorgesteld. Het centrale concept is de *modulus beperking*.

De modulus beperking

Een gestratifieerde aanpak voor zelf-calibratie vereist een methode om het vlak op oneindig van de andere vlakken te differentiëren. Een eigenschap van de homografieën voor dit vlak zal hiervoor gebruikt worden. Deze homografie kan zowel geschreven worden als functie van de euclidische entiteiten, als in functie van projectieve entiteiten en een aantal onbekenden.

In het eerste geval laat ons dit toe om een modulusbeperking af te leiden. De homografieën voor het vlak op oneindig moeten namelijk altijd geconjugeerd zijn met rotatiematrices op een schaalfactor na. Dit betekent dat de modulus van alle eigenwaarden gelijk moet zijn.

Deze beperking kan opgelegd worden aan de projectieve vorm van de homografie van het vlak op oneindig. In dit geval geeft deze vergelijking een beperking op de onbekende positie van het vlak op oneindig. Waneer genoeg vergelijkingen voorhanden zijn kan men hieruit de positie van het vlak op oneindig bepalen.

De modulus beperking levert een vierdegraadsvergelijking op voor elk paar van beelden. Het aantal onbekenden is drie, nl. de positie van het vlak op oneindig.

Zelf-calibratie met constante intrinsieke parameters

Met behulp van de modulus beperking heb ik een gestratifieerde zelf-calibratieprocedure uitgewerkt. Een minimum van drie beelden is noodzakelijk om het vlak op oneindig te kunnen bepalen (in dit geval kunnen de paren 1-2, 1-3 en 2-3 gebruikt worden). In het minimale geval is het gebruik van een continuatie algoritme aangewezen zodat alle mogelijke oplossingen gevonden worden. Na eliminatie van onmogelijke oplossingen (bv. imaginair), wordt de meest waarschijnlijke oplossing geselecteerd. Dit kan uitgesteld worden tot na de metrische calibratie.

Indien men over meer zichten beschikt, kunnen alle beperkingen gecombineerd worden in een optimisatie criterium. Het minimisatie algoritme kan geïnitialiseerd worden met de resultaten van de continuatiemethode. In dit geval zal er in het algemeen slechts één oplossing zijn.

Eens het vlak op oneindig gelocaliseerd is, kan men de intrinsieke camera parameters eenvoudig bepalen door een lineaire stelsel van vergelijkingen op te lossen. Om een hogere nauwkeurigheid te bekomen is het echter aangewezen om in een laatste stap een globale minimisatie te doen op de parameters van het affiene en het metrische stratum.

Uit onze experimenten blijkt dat deze zelf-calibratie strategie zeer goede resultaten oplevert. Deze methode werd eveneens vergeleken met een aantal alternatieve technieken. De Kruppa methode blijkt in de praktijk zeer slechte resultaten op te leveren (wat trouwens ook theoretisch verklaard kan worden). De andere technieken vertrekken van een ruwe schatting van de parameters en proberen onmiddelijk één of andere vorm van globale minimisatie over alle parameters van het probleem. Op het vlak van nauwkeurigheid zijn deze methodes vergelijkbaar met de gestratifieerde aanpak, maar op het vlak van robuustheid presteerde onze aanpak tijdens de experimenten duidelijk beter.

De werkbaarheid van onze aanpak werd eveneens aangetoond aan de hand van een aantal beeldsequenties van het Arenberg kasteel. Metingen tonen aan dat de metrische eigenschappen van de scène in de reconstructie teruggevonden worden (bv. orthogonaliteit).

Andere toepassingen

De modulus beperking blijkt eveneens zeer geschikt te zijn voor een aantal meer specifieke zelf-calibratie problemen. Twee gevallen worden in dit werk besproken.

Twee beelden en twee vluchtpunten Voor één beeldpaar beschikt men slechts over één modulus beperking. Dit is niet voldoende om zelf-calibratie toe te laten. Twee vluchtpunten op zich zijn eveneens onvoldoende om de positie van het vlak op oneindig te bepalen. Door beide beperkingen te combineren heeft men echter wel voldoende informatie om het vlak op oneindig te localiseren.

In dit geval blijkt de modulus beperking heel eenvoudig toe te passen. Deze methode werd uitgewerkt en zowel op synthetische data als op echte beelden toegepast. In dit laatste geval hebben we gebruik gemaakt van een algoritme dat automatisch vluchtpunten uit beelden localiseert (in heel wat scènes kan men automatisch 2 maar niet 3 vluchtpunten localiseren). De resultaten zijn goed, maar de methode blijkt zeer gevoelig te zijn aan ruis op de metingen. Dit is te verwachten voor een methode die werkt met een minimale hoeveelheid informatie.

Een variërende focale lengte De modulus beperking kan eveneens voor andere doeleinden dan affiene calibratie gebruikt worden. De beperking is gebaseerd op twee voorwaarden: de affiene calibratie *en* constante intrinsieke camera parameters. In plaats van de beperking te gebruiken om de affiene calibratie te bekomen, kan men in het het traditionele geval –waar affiene calibratie door pure translatie bekomen wordt– deze beperkingen gebruiken om het variëren van een parameter op te vangen. De meest praktische toepassing is het variëren van de focal lengte. Dit laat toe om de zelf-calibratie te bekomen ondanks het gebruik van zoom en auto-focus. Deze techniek werd geïmplementeerd en geëvalueerd op zowel reëele als synthetische gegevens.

Een vaste stereo opstelling met variërende focale lengte Voor een vaste stereo opstelling is het mogelijk om variërende focale lengtes toe te laten. In dit geval wordt echter geen gebruik gemaakt van de modulus beperking, maar van de epipolaire geometrie. Voor een vaste stereo opstelling is de epipolaire geometrie normaal gezien constant. Wanneer de focale lengte variëert zal de afstand tussen de epipool en het centrum van het beeld volgens dezelfde verhouding veranderen. Door het beeld te herschalen zodat de epipool terug op zijn plaats komt, is het mogelijk om de verandering in focale lengte te compenseren. Daarna kunnen de standaard algoritmes voor vaste stereo opstellingen gebruikt worden.

Flexiebele zelf-calibratie

De laatste jaren werd de mogelijkheid tot zelf-calibratie van cameras door heel wat onderzoekers bestudeerd. Meestal werden constante maar volledig ongekende parameters verondersteld. Het belangrijke nadeel hiervan is dat het tijdens opnames niet toegelaten is om te zoomen of te focussen. Anderzijds is het voorgestelde perspectief camera model meestal te algemeen vergeleken met de klassen van bestaande cameras. Meestal kan men veronderstellen dat pixels rechthoekig –of zelf vierkantig– zijn. Daardoor kan een meer pragmatische aanpak gevolgd worden. Door een aantal parameters als gekend te veronderstellen, kan men andere parameters laten variëren tijdens de opnames.

Theorie

Alvorens een practisch algoritme voor te stellen, worden een aantal theoretische aspecten van het zelf-calibratie probleem voor variërende camera parameters besproken.

Een tel argument Een projectieve transformatie heeft 15 vrijheidsgraden, terwijl een metrisch transformatie er slechts 7 telt. Er zijn dus 8 onafhankelijke vergelijkingen nodig om de ambiguïteit van projectief tot metrisch te beperken. Elke gekende parameter levert één vergelijking per beeld op, een constante maar ongekende parameter levert één vergelijking minder op. Het volgende telargument wordt dus bekomen (met n het aantal beelden):

$$n \times (\#gekend) + (n-1) \times (\#vast) \geq 8$$

Dit laat toe om te bepalen wat de minimale lengte is waarvoor zelf-calibratie bekomen kan worden. Dit is natuurlijk enkel geldig wanneer men niet met een kritische bewegingssequentie te maken heeft.

Een geometrische interpretatie van zelf-calibratie beperkingen Om een beter inzicht te krijgen in het zelf-calibratie probleem werd een geometrische interpretatie afgeleid voor de verschillende zelf-calibratie beperkingen. De beperking van rechthoekige pixels en gekende parameters voor het focale punt komen overeen met het opleggen van orthogonaliteit tussen twee vlakken in de ruimte. De beperkingen van een gekende focale lengte of van een gekende verhouding tussen hoogte en breedte van pixels komt neer op een gekende verhouding tussen twee –typisch orthogonale– vectoren in de ruimte.

Minimale voorwaarden voor zelf-calibratie Er werd aangetoond dat zelf voor het minimale geval waar slechts wordt aangenomen dat de pixels rechthoekig zijn, zelf-calibratie in principe mogelijk is. Dit werd gedaan aan de hand van het volgende theorema.

Theorema 6.1 *De klasse van transformaties die rechthoekigheid van pixels bewaard is de groep van de similariteiten*

Gebruik makende van de concepten van de vorige paragraaf werd het besluit van dit theorema veralgemeend tot eender welke gekende intrinsieke parameter. Het bewijs dat geleverd wordt, is puur geometrisch en wordt geïllustreerd in Figuur 6.1 en 6.2.

Zelf-calibratie van een camera met variërende intrinsieke parameters

Zoals reeds hoger vermeld werd, zijn zelf-calibratie algoritmes meestal gebaseerd op de absolute kegelsnede. De projecties hiervan zijn rechtstreeks gerelateerd met de intrinsieke camera parameters.

De aanpak die we voorstellen bestaat erin om de beperkingen op de intrinsieke camera parameters te vertalen naar beperkingen op het beeld van de absolute kegelsnede. Dit gebeurt eenvoudig door middel van een parametrisatie.

Door middel van de projectie vergelijking worden de beperkingen voor elk beeld van de absolute kegelsnede teruggeprojecteerd naar beperkingen voor de absolute kegelsnede zelf. Deze wordt dan gevonden als de kegelsnede die het best aan alle beperkingen beantwoordt.

Niet-lineaire methode Een niet-lineair algoritme kan eenvoudig opgezet worden. In dit geval brengt men alle beperkingen samen in één optimisatie-criterium. De nauwkeurigheid die hiermee behaald wordt is zeer goed, maar men moet wel beschikken over een initialisatie. Hiervoor werd een lineair algoritme opgesteld dat een benaderende oplossing oplevert.

Lineaire methode Wanneer een aantal veronderstellingen worden gedaan (nl. rechthoekige pixels en het focaal punt gekend), kan men het stelsel van vergelijkingen tot een lineair stelsel herleiden. Bijkomende beperkingen op de intrinsieke parameters resulteren dan eveneens in lineaire vergelijkingen. Zelfs wanneer aan deze veronderstellingen niet perfect voldaan wordt, laat deze methode meestal toch toe om een goede initialisatie te bekomen voor de het oplossen van het niet-lineaire stelsel van vergelijkingen.

Kritische bewegingssequenties

In de vorige paragrafen werd een methode gepresenteerd die verschillende types beperkingen op de intrinsieke camera parameters kan combineren. In dit geval kunnen kritische bewegingssequenties natuurlijk ook voorkomen. Welk type kritische bewegingssequenties kunnen voorkomen hangt af van welke specifieke beperkingen gebruikt worden voor zelf-calibratie. De twee uitersten zijn: volledig gekende parameters met bijna geen kritische gevallen; en helemaal geen beperkingen waarvoor alle bewegingssequenties kritisch zijn.

Vertrekkende van de analyse voor constante parameters, kan men relatief eenvoudig rekening houden met een aantal gekende intrinsieke camera parameters. Enkel de kritische sequenties die effectief de opgelegde waarden van de gekende parameters hebben, zijn dan nog kritisch. Dit wil zeggen dat in dit geval dezelfde klassen van kritische bewegingssequenties blijven bestaan, maar dat de ambiguïteit op de reconstructie kleiner is.

Het is moeilijker om rekening te houden met variërende intrinsieke camera-parameters. Een gelijkaardige analyse kan echter uitgevoerd worden als voor constante parameters: *Gegeven een specifiek kegelsnede, welke combinaties van posities en orientatie beantwoorden aan alle beperkingen?* Door dit probleem op te lossen voor de verschillende types kegelsneden, bekomt men de klassen van kritische bewegingssequenties voor de veronderstelde beperkingen.

Kritische bewegingssequenties voor een variërende focale lengte Eén van de meest interessante gevallen is het geval waar alle parameters gekend zijn behalve de focale lengte die vrij kan variëren. De verschillende klassen van kritische bewegingen werden afgeleid: **F1:** rotatie rond de optische as en translaties, **F2:** hyperbolische en/of elliptische beweging, **F3:** voorwaartse beweging, **F4:** twee posities met willekeurige rotaties, **F4.1:** pure rotaties.

In appendix werd eveneens een methode afgeleid om deze kritische bewegingen te visualizeren. Dit laat toe om een intuïtief inzicht te verwerven in dit complexe probleem.

Daar het lineaire algoritme de planariteit van de duale absolute kwadriek niet oplegt, zijn er in dit geval extra kritische bewegingssequenties. Deze treden op wanneer de camera een vast punt fixeert tijdens de opnames.

Het ontdekken van kritische bewegingssequenties Er werd een methode afgeleid om voor een specifieke bewegingssequentie numeriek na te kijken of de bewegingssequentie kritisch of quasi-kritisch was. Dit kan zeer nuttig zijn om te beslissen of er bijkomende (benaderende) beperkingen moeten opgelegd worden aan de parameters om toch een (quasi-)metrische reconstructie te bekomen.

Selectie van beperkingen

In deze paragraaf worden een aantal algemene principes voorgesteld om vanuit de opnames zelf automatisch te bepalen welke beperkingen van toepassing zijn. Men kan het vergelijken met automatische modelselectie (bv. Akaike's informatiecriterium), maar deze principes zijn in dit geval niet rechtstreeks toepasbaar.

Voor het lineaire algoritme werd een pragmatische oplossing voorgesteld om het probleem van de bijkomende kritische bewegingen op te vangen door indien nodig –automatisch– de focale lengte vast te leggen op een benaderende waarde. Dit blijkt in de praktijk goed te werken.

Experimenten

Een aantal experimenten werden uitgevoerd om de validiteit van de voorgestelde zelf-calibratie algoritmes aan te tonen. Hiervoor werden zowel synthetische gegevens als echte beeldsequenties gebruikt. De resultaten tonen aan dat de algoritmes er effectief in slagen om de metrische structuur van de opgenomen scènes te reconstrueren.

Metrische 3D reconstructie

Het bekomen van 3D modellen van objecten is een onderwerp dat in het domein van computervisie reeds vele jaren bestudeerd wordt. Een paar jaren geleden was de belangrijkste toepassing robot navigatie en visuele inspectie. Tegenwoordig is dit veranderd. Er is meer en meer vraag naar 3D modellen vanuit computer visualizatie, virtuele realiteit en communicatie. Dit heeft tot gevolg dat een aantal aspecten van het reconstructie probleem veranderen. De visuele kwaliteit van de modellen wordt één van de belangrijkste punten.

Voor deze nieuwe toepassingen zijn de opname omstandigheden en het niveau van expertise van de gebruikers vaak heel anders dan wat verwacht wordt om de meeste bestaande 3D opname technieken te gebruiken. De bestaande technieken vereisen vaak het uitvoeren van complexe calibratie procedures bij elke opname. Er is ook een belangrijke vraag naar flexibiliteit tijdens de opnames. Calibratie procedures moeten afwezig zijn, of tot een minimum beperkt blijven.

Daarenboven zijn heel wat bestaande technieken opgebouwd rond gespecialiseerde hardware (bv. laser range scanners of stereo opstellingen) wat resulteert in een hoge kost voor deze systemen. Veel nieuwe toepassingen vergen echter goedkope opname technieken. Dit stimuleert het gebruik van standaard foto- of video camera's.

In dit werk werd een systeem uitgebouwd dat 3D modellen genereert van het oppervlak van objecten uit een sequentie van beelden die met een gewone camera werden opgenomen. De gebruiker neemt de beelden op door vrij rond het object te bewegen. Noch de camerabeweging, noch de camera instellingen moeten gekend zijn. Het bekomen model is een schaalmodel van het oorspronkelijke object (nl. een *metrische* reconstructie). De textuur wordt eveneens rechtstreeks uit de beelden gehaald.

Andere onderzoekers hebben reeds een aantal systemen voorgesteld om 3D modellen uit beelden te halen. Deze systemen hebben echter allemaal een aantal belangrijke beperkingen in vergelijking tot ons systeem. Ofwel wordt slechts rekening gehouden met een orthografisch camera model, ofwel is er a priori informatie nodig over de scène. Meestal worden slechts een beperkt aantal punten gereconstrueert en niet een volledig oppervlakte model.

Ons systeem gebruikt een perspectief camera en heeft geen specifieke a priori informatie nodig, noch over de scène, noch over de camera. Het systeem is gebaseerd op recente algoritmes voor *projectieve reconstructie*, *zelf-calibratie* en *dichte diepteschatting*. In de volgende paragraaf wordt het systeem meer in detail besproken.

Het 3D reconstructie systeem

Het systeem bestaat uit een aantal modules. Door het systeem wordt geleidelijk aan meer en meer informatie over de scène en de camera opstelling bekomen.

Projectieve reconstructie In een eerste stap worden de verschillende beelden ten opzichte van elkaar gerelateerd. Dit wordt paarsgewijs gedaan door de epipolaire geometrie te bepalen. Een initiële reconstructie wordt dan gemaakt aan de hand van de eerste twee beelden. Voor de volgende beelden wordt eerst de camerapose ten opzichte van de reeds bekomen projectieve reconstructie geschat. Voor elk van deze beelden worden dan de punten die in correspondentie werden gebracht met een punt uit het vorige beeld, gereconstrueerd, verfijnd of verbeterd. Daardoor is het niet nodig dat de oorspronkelijke punten zichtbaar blijven tijdens de gehele sequenties. Voor sequenties waar punten verdwijnen en later weer zichtbaar worden, werd het algoritme aangepast zodat hier –op een efficiente manier– mee rekening werd gehouden. In dit geval wordt het huidige beeld niet enkel gerelateerd met het vorige beeld, maar ook met een aantal andere zichten. Het resultaat van deze volledige procedure is typisch een paar honderd tot een paar duizend gereconstrueerde punten en de (projectieve) pose van de cameras. Deze reconstructie is enkel bepaald tot op een willekeurige projectieve transformatie na.

Van een projectieve naar een metrische reconstructie De volgende stap bestaat erin om deze ambiguïteit te beperken tot op een willekeurige metrische transformatie na. Bij een projectieve reconstructie is niet enkel de scène, maar ook de camera vervormd. Vermits dit algoritme omgaat met onbekende scènes is het niet mogelijk om automatisch vervormingen van de scène vast te stellen. Alhoewel de camera eveneens niet gekend is, bestaan er in de meeste gevallen toch wel een aantal beperkingen op de intrinsieke camera parameters (bv. rechthoekige of vierkante pixels, een constante verhouding tussen de zijden van de pixels, het focaal punt in het midden van het beeld). Een vervorming van de camera's resulteert meestal in het niet meer opgaan van één of meer van deze beperkingen. Een metrische reconstructie en calibratie worden bekomen door de projectieve reconstructie te transformeren zodat alle beperkingen op de intrinsieke camera parameters zo goed mogelijk opgaan. Praktische methoden hiervoor werden uitvoerig besproken in dit werk.

Dichte diepte schatting Na de procedures van de vorige paragrafen te hebben uitgevoerd, beschikt men over een gecalibreerde beeldsequentie. De relatieve positie en orientatie van de camera's is gekend voor alle zichten. Deze calibratie vergemakkelijkt het zoeken naar correspondenties en laat ons toe om stereo-algoritmes te gebruiken die ontwikkeld werden voor gecalibreerde systemen. Hiermee kunnen voor bijna elke pixel de corresponderende pixels in andere beelden bepaald worden.

Deze correspondenties laten dan toe om door triangulatie de afstand tussen de camera en het oppervlak van het geobserveerd object te bepalen. Deze resultaten worden verfijnd door de resultaten voor meerdere zichten te combineren.

Opbouw van het 3D model Een dicht metrisch 3D oppervlaktemodel wordt bekomen door de berekende diepte te gebruiken om in de ruimte een driehoekennet op te bouwen dat het oorspronkelijke oppervlak benadert. De textuur wordt uit de beelden gehaald en op het oppervlak geplaatst.

Overzicht van het systeem In Figuur D.7 wordt een overzicht van het uitgebouwde systeem gegeven. Het bestaat uit een aantal onafhankelijke modules die de nodige informatie doorspelen naar de volgende modules. De eerste module bepaalt een projectieve reconstructie van de camera-opstelling tegelijkertijd met een spaarse reconstructie van de scène. De volgende module berekent de metrische calibratie van de reconstructie aan de hand van zelf-calibratie. Vervolgens worden dichte correspondentie kaarten bepaald. Uiteindelijk worden alle resultaten geintegreerd in een getextureerd 3D oppervlakte model van de opgenomen scène.

Implementatie

Er worden kort een aantal aspecten van de implementatie toegelicht. Alhoewel de vereiste rekentijd sterk afhangt van de gebruikte beeldsequentie, worden toch een aantal typische uitvoeringstijden gegeven. Om een volledig 3D model op te bouwen uit een tiental beelden heeft een standaard werkstation ongeveer 1 uur rekentijd nodig.

Een paar mogelijke verbeteringen

Het systeem werkt reeds op heel wat beeldsequenties, maar er zijn gevallen waarvoor het systeem faalt. Er worden een aantal suggesties gemaakt om bepaalde problemen op te lossen. Sommigen hiervan werden reeds geïmplementeerd.

Bepaling van puntcorrespondenties Dit is een van de meest kritische punten van het systeem. Wanneer twee opeenvolgende beelden te erg van elkaar verschillen, faalt het hele systeem. Een manier om dit op te vangen bestaat erin om gebruikersinteractie

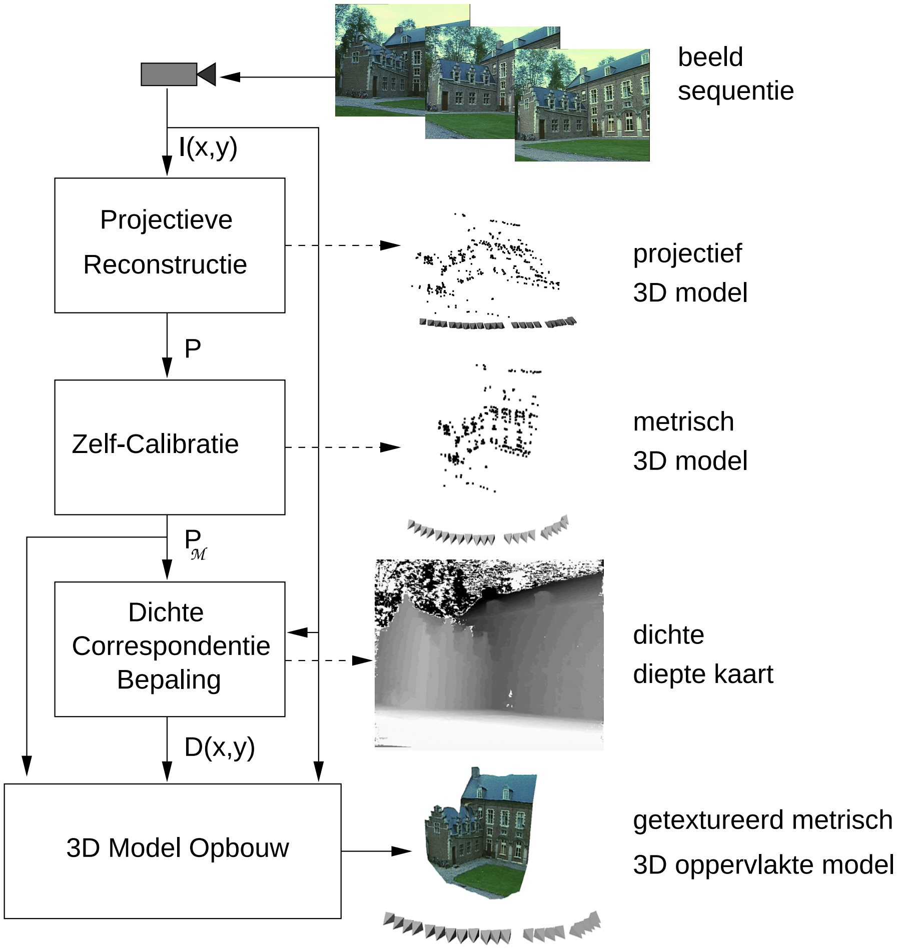

Figure D.7: Overzicht van het systeem: Uit de beelden ($I(x, y)$) wordt de projectieve reconstructie berekend; De projectieve cameramatrices $\mathbf{P}$ worden doorgegeven aan de zelf-calibratie module die een metrische calibratie $\mathbf{P}_M$ aflevert; De volgende module gebruikt deze om dichte correpondentie kaarten $D(x, y)$ te bepalen; Alle deze resultaten worden samengebracht in de laatste module tot een getextureerd 3D oppervlaktemodel. Rechts worden de resultaten van de verschillende modules weergegeven: de voorlopige reconstructies (zowel projectief als metrisch) worden weergegeven door een puntenwolk, de camera's worden weergegeven door kleine pyramides, de resultaten van de dichte correspondentiebepaling worden geaccumuleerd in een dichte correspondentiekaart (licht betekent dichtbij en donker betekent veraf).

toe te laten. Recent werden echter een aantal interessante resultaten behaald op het vlak van automatische puntcorrespondentie bepaling. Het integreren van deze nieuwe technieken in het systeem zal resulteren in het verhogen van de robuustheid en zal de beperkingen tijdens de opnames verminderen. Het verschil tussen opeenvolgende zichten mag groter zijn.

Voor zeer verschillende beelden bestaan er twee technieken. De eerste bestaat erin om de beelden eerst te transformeren met behulp van een globale of lokale homografie zodat het verschil tussen de beelden wordt verkleind. Een andere aanpak bestaat erin om een meer algemene similariteitmeting te gebruiken die bv. niet enkel invariant voor translatie , maar ook voor rotatie, schaal en belichting. Uiteindelijk kunnen er ook andere problemen optreden zoals repetitieve structuren die –wanneer men er niet op een hoger niveau rekening mee houdt– de correspondetie bepaling in de war kunnen sturen.

Projectieve reconstructie De methode die op dit moment gebruikt wordt voor de projectieve reconstructie is niet volledig optimaal. Het grootste nadeel is dat het algoritme faalt wanneer de initialisatie aan de hand van de eerste beelden mislukt.

Recent werden echter technieken voorgesteld die niet meer van een specifiek paar van beelden afhankelijk zijn voor de initialisatie van de reconstructie. Daarenboven maken deze methodes enkel en alleen gebruik van projectieve meetkunde (in de 3D ruimte). Het op deze manier aanpassen van de module voor projectieve reconstructie zou de robuustheid van ons systeem moeten verbeteren.

Het overleven van vlakke delen Dit is ook een belangrijk probleem. Projectieve posebepaling werkt niet als alle zichtbare punten in een vlak liggen. Er werd een mogelijke strategie voorgesteld die zelf-calibratie vroeger inschakelt. Voor vlakke delen hoeft dan slechts een metrische pose geschat te worden (wat wel mogelijk is voor vlakke scènes).

Veralgemeende rectificatie De meeste stereo algoritmes werken op gerectificeerde beelden. Dit betekent dat men er door transformatie van de oorspronkelijke beelden voor zorgt dat overeenkomstige rijen van het beeld in epipolaire correspondentie staan. Dit laat toe om zeer efficiënte algoritmes te bekomen.

Bij de standaard techniek worden gerectificeerde beelden bekomen door de epipool naar oneindig te transformeren, wat geen probleem is voor de meeste stereo opstellingen. In ons geval kunnen we echter niet garanderen dat er geen belangrijke voorwaartse verplaatsing plaatsgrijpt tussen twee opeenvolgende beelden zodat de epipool in het beeld terecht komt. In dit geval is de standaard techniek niet toepasbaar. Recent werd een methode voorgesteld die dit geval wel aankon, maar deze methode heeft een aantal belagrijke nadelen.

Er werd een nieuwe techniek voorgesteld voor rectificatie. Deze eenvoudige en efficiënte techniek werd geïmplementeerd. Het basis idee bestaat erin om polaire coordinaten te gebruiken rond de epipolen. De grootte van de resulterende beelden is optimaal (zo klein mogelijk zonder pixels te comprimeren).

Resultaten en toepassingen

Het systeem werd toegepast op heel wat verschillende beeldsequenties. In de volgende paragrafen worden een aantal voorbeelden gegeven van toepassingen waarvoor het systeem gebruikt werd. De bijhorende figuren kunnen teruggevonden worden in Hoofdstuk 8.

Acquisitie van 3D modellen met foto's

De belangrijkste toepassing van ons systeem is het verkrijgen van 3D modellen uit beelden. Eén van de eenvoudigste methodes om een 3D model van een bestaande scène te bekomen bestaat er dus in om een aantal foto's vanuit verschillende standpunten te nemen. Deze kunnen dan automatisch verwerkt worden door het systeem tot een realistisch 3D model van de beschouwde scène.

De mogelijkheden van ons systeem op dit vlak werden geïllustreerd met behulp van een zeer fijn afgewerkt deel van een Jain tempel in Ranakpur (India). Er werden 5 foto's gebruikt. Het algoritme verwerkte deze automatisch tot een gedetaileerd 3D oppervlakte model van het opgenomen stuk tempel.

Acquisitie van 3D modellen uit bestaande beeldsequenties

Dankzij de flexibiliteit van het systeem is het zelfs mogelijk om bestaand filmmateriaal te gebruiken om 3D modellen te genereren. Deze mogelijkheid werd geïllustreerd met behulp van een opname van het antieke theater van Sagalassos dat een paar jaar geleden door de BRTN werd opgenomen. Deze opname kon door het systeem probleemloos tot een 3D model verwerkt worden.

Acquisitie van plenoptische modellen

De laatste jaren werd heel wat onderzoek gedaan naar plenoptische modellen. In plaats van een scène geometrisch te modelleren, gaat men hier rechtstreeks het uitzicht van een scène trachten op te nemen. Dit gebeurt door het licht dat in elke richting door elk punt van de scène passeert, op te meten. Typisch wordt zo'n model bekomen door een enorme hoeveelheid beelden op te nemen van een scène. Eén van de belangrijkste problemen hierbij is het bepalen van de positie, orientatie en de calibratie van de camera voor elk zicht. Totnogtoe gebeurde dit meestal door de camera te monteren op een robotarm of door een calibratieobject op te nemen in de scène.

Na een aanpassing van ons systeem kon het probleemloos gebruikt worden voor de acquisitie van plenoptisch modellen. Er werd bijvoorbeeld een beeldsequentie van meer dan 180 beelden verwerkt. Het systeem kon de pose en calibratie van alle zichten berekenen. Dit kon dan verder gebruikt worden om een plenoptisch model te construeren.

Het virtualizeren van archaeologische sites

Archaeologie is een ideaal toepassingsgebied voor onze techniek. Er zijn belangrijke toepassingen voor 3D modellen van opgegraven sites. Zowel voor metingen als om gebruik te maken van de nieuwe mogelijkheden die virtuele realiteit toelaat.

Virtualizeren van scènes Door een aantal foto's te nemen van een gebouw, steen of andere object op een archaeologische site kan men er met ons systeem een 3D model mee genereren. Men bekomt dus een virtuele copie van de opgenomen scène. Dit werd in Sagalassos (Turkije) met succes toegepast op heel wat scènes: fontein, heroon, theater, stenen, etc.

Het reconstrueren van een globaal model Ons systeem is onafhankelijk van de schaal van de op te nemen scène. Het enige verschil is de afstand die tussen het opnemen van twee opeenvolgende foto's afgelegd moet worden. Daardoor kan men ook zeer grote scènes reconstrueren. In Sagalassos heeft dit ons toegelaten om de hele site in één keer op te nemen vanaf een nabij gelegen heuvel.

Reconstructies op verschillende schalen Dit overzichtsmodel is echter niet voldoende gedetailleerd om de verschillende monumenten die op de site te vinden zijn realistisch weer te geven. Daarom werden hiervan aparte reconstructies gemaakt (uit lokale beeldsequenties) en deze werden geïntegreerd in het globale model.

Combinatie met andere modellen De modellen die met onze techniek bekomen worden kunnen eenvoudig gecombineerd worden met andere 3D modellen. Zo werden bijvoorbeeld 3D CAD modellen van monumenten uit Sagalassos probleemloos in het globale 3D model van Sagalassos geïntegreerd.

Andere toepassingen in archaeologie

In het domein van archaeologie zijn nog heel wat andere interessante toepassingen mogelijk van onze techniek.

3D stratigrafie Tijdens opgravingen is het heel belangrijk om zoveel mogelijk informatie op te nemen. Later zal dit niet meer mogelijk zijn. Momenteel wordt enkel een profiel van de verschillende stratigrafische grondlagen opgemeten. Het zou echter veel interessanter zijn mocht de stratigrafie van een bepaalde sector volledig in 3D opgenomen worden.

Testen die uitgevoerd werden met onze techniek tonen aan dat dit haalbaar is. Daarenboven is het zelfs zo dat de opname tijd op de site zelf korter wordt. Het gaat namelijk sneller om een paar foto's te nemen dan om een profiel op te meten.

Het genereren en testen van constructiehypothesen Een andere toepassing bestaat erin om gebroken bouwelementen of delen van ingestorte gebouwen weer samen te brengen dankzij 3D copieën die veel gemakkelijker gemanipuleerd kunnen worden. Bovendien zou men zelf automatische registratie technieken kunnen toepassen.

Toepassingen in andere gebieden

Naast archaeologie bestaan er nog heel wat andere domeinen die 3D metingen van bestaande 3D structuren vergen. Heel wat interessante toepassingen kunnen gevonden worden in het domein van architectuur en conservatie. Zo werden een aantal testen uitgevoerd op de kathedraal van Antwerpen.

De flexibiliteit van de voorgestelde techniek maakt toepassingen mogelijk in vele domeinen. In een aantal gevallen moet het systeem uitgebreid worden, in andere gevallen kan het systeem (of een deel ervan) onmiddelijk gebruikt worden. Een aantal interessante domeinen zijn forensisch onderzoek (bv. virtuele reconstructies van een misdaadscène), robotica (bv. 3D modellering van de omgeving voor autonome voertuigen), virtueel uitgebreide realiteit (bv. camera positie bepaling) of post-productie (bv. virtuele opname sets genereren).

Besluit

Het werk dat werd voorgesteld in deze thesis handelt over de automatische acquisitie van realistische 3D modellen uit beelden. Ik heb getracht om een systeem te ontwikkelen dat een maximum aan flexibiliteit toelaat tijdens de opnames. Dit werk bestond zowel uit het ontwikkelen van nieuwe theoretische concepten als uit het vertalen van deze concepten naar technieken die werken op echte beeldsequenties.

Dit probleem werd onderverdeeld in een aantal deeltaken. Voor een aantal van deze taken bestonden reeds oplossingen die voldoening geven en was er dus geen nood aan het ontwikkelen van een nieuwe techniek om ons doel te bereiken. Indien mogelijk werd een bestaande implementatie gebruikt. Voor zelf-calibratie echter leverden alle bestaande methodes bij de aanvang van mijn werk geen goede resultaten op voor echte beeldsequenties. Ze konden bovendien niet omgaan met variërende cameraparameters.

Een eerste belangrijk deel van dit werk bestond er dus in om een zelf-calibratie algoritme te ontwikkelen dat goede resultaten gaf op echte beeldsequenties. Geïnspireerd door de gestratifieerde technieken die goede resultaten gaven voor bepaalde specifieke bewegingen, ontwikkelde ik een techniek die de metrische structuur uit de projectieve bekomt door eerst de affiene structuur te bepalen. Deze techiek is gebaseerd op een nieuwe zelf-calibratie-beperking: de modulus beperking. Goede resultaten werden bekomen, zowel tijdens synthetische experimenten als op echte beelden.

De veronderstelling die traditioneel bij zelf-calibratie gemaakt wordt, is die dat de intrinsieke camera parameters constant maar volledig ongekend zijn. Enerzijds is deze veronderstelling beperkend vermits het gebruik van zoom en focus uitgesloten worden. Anderzijds is ze te algemeen omdat de meeste camera’s rechthoekige of zelfs vierkante pixels hebben en het focaal punt in de buurt van het centrum van het beeld ligt. Een methode werd voorgesteld die met alle mogelijke beperkingen op de intrinsieke cameraparameters kon rekening houden, nl. gekend, constant of variërend. Deze methode liet ons toe om een algoritme te ontwikkelen dat werkt op echte beeldsequenties die opgenomen worden met een camera waarvan de zoom en focus zonder probleem gebruikt mag worden. Deze methode werd gevalideerd op reële en synthetische gegevens.

Deze methodes werden gecombineerd met andere modules om een volledig 3D reconstructiesysteem te bouwen. Het systeem vertrekt van een beeldsequentie van een scène. Het resultaat is een realistisch getextureerd 3D oppervlakte model van deze scène. De verwerking is volledig automatisch. Dit systeem biedt een zeer goed niveau van detail en realisme aan, gecombineerd met een voorheen nooit bereikte flexibiliteit voor de acquisitie.

Deze flexibiliteit van het systeem werd geïllustreerd aan de hand van de verschillende voorbeelden. Realistische 3D modellen werden zowel bekomen uit een paar foto's als uit bestaande video opnames. Scènes van zeer verschillende groottes werden gereconstrueerd (van één enkele steen tot een archaeologische site die zich over verschillende vierkante kilometers uitstrekt). Na een kleine aanpassing kon het systeem gebruikt worden om plenoptische modellen, gebaseerd op honderden beelden op te nemen. Het systeem werd met succes gebruikt in het domein van archeologie waar veelbelovende resultaten behaald werden. Het systeem laat zelfs een aantal nieuwe toepassingen toe die voorheen onhaalbaar waren.

Discussie en verder onderzoek

Sinds een aantal jaren is er een discussie aan de gang tussen de voor- en tegenstanders van *gecalibreerde* versus *ongecalibreerde* systemen. Volgens mij is dit een onnodige discussie. Een heel palet aan mogelijkheden bestaat voor de acquisitie van 3D informatie uit beelden. Aan het ene uiterste vind men een volledig gecalibreerd systeem dat enkel onder zeer beperkende omstandigheden werkt (bv. gecontroleerde belichting, beperkte dieptevariatie). Aan het andere uiterste zou men een systeem kunnen hebben dat genoeg heeft aan een paar willekeurige zichten van een object. Volgens mij is het de taak van de computervisie-gemeenschap om deze verschillende alternatieven te onderzoeken. Het ideale 3D acquisitiesysteem bestaat niet, alles hangt af van de specifieke toepassing die men voor ogen heeft. In deze context denk ik dat het een belangrijke taak van onze gemeenschap is om de grenzen te onderzoeken van wat bereikt kan worden.

Wat echter wel moet gezegd worden is dat de ongecalibreerde aanpak en het gebruik van projectieve meetkunde heel wat nieuwe inzichten heeft mogelijk gemaakt. Projectieve meetkunde is het natuurlijke kader om de onderliggende principes te beschrijven die beelden en scène relateren. Dikwijls worden onnodig vergelijkingen gebruikt om rekening te houden met Euclidische aspecten terwijl de eigenschappen die men bestudeert reeds aanwezig zijn op het projectieve niveau. Merk trouwens op dat het gebruik van projectieve meetkunde calibratie niet uitsluit. Het laat echter wel toe om een duidelijk onderscheid te maken tussen wat ervan afhangt en wat niet.

Het systeem dat in dit werk beschreven werd, is zeker geen eindpunt. Zoals reeds op een aantal plaatsen werd opgemerkt, bestaan er meerdere mogelijkheden om het systeem verder te verbeteren. Verschillende richtingen voor verder onderzoek kunnen geïdentificeerd worden.

Een eerste belangrijke taak bestaat erin om het systeem zelf te verbeteren. Deze taak bestaat uit twee delen. Eenerzijds zou het systeem meer robuust moeten gemaakt worden zodat het op een grotere klasse beelden toepasbaar is. Het falen van een algoritme zou automatisch gedetecteerd moeten worden. Indien mogelijk zou het systeem dit probleem te boven moeten komen door een meer robuust of meer geschikt algoritme te gebruiken. Anderzijds kan de nauwkeurigheid van het systeem zeker opgedreven worden. Het relatief eenvoudige camera model dat momenteel gebruikt wordt kan verfijnd worden. Het gebruik van grootste waarschijnlijkheid schatters (maximum likelihood estimators) kan veralgemeend worden in de verschillende modules.

Een tweede taak bestaat erin om het systeem uit te breiden. Op dit ogenblik worden 3D oppervlakte modellen van de scène gegenereerd. Deze oppervlakken worden echter nog steeds opgebouwd vanuit een referentie zicht. Het systeem zou deze verschillende voorstellingen moeten samenbrengen in een unieke 3D voorstelling. Dit kan gebeuren door beroep te doen op bestaande fusie technieken die ontwikkeld werden in de context van gecalibreerde diepte bepaling. Deze methoden dienen echter aangepast te worden aan de specifieke eigenschappen van onze aanpak. Een interessant idee bestaat erin om het uiteindelijke 3D model te verfijnen door het rechtstreeks te vergelijken met de oorspronkelijke beelden. Dit zou kunnen bestempeld worden als *fotometrische bundelaanpassing*. Bijkomende mogelijkheden zijn het fitten van parametrisch voorstellingen aan de data (bv. vlakken, kwadrieken, etc) of het infereren van hogere orde voorstellingen wat ons zou leiden tot modelgebaseerde strategieën of tot scène-analyse.

Onze aanpak is een geometrisch aanpak. Het systeem tracht alles te modelleren aan de hand van geometrische primitieven. Recent kennen beeldgebaseerde voorstellingen heel wat belangstelling. Deze technieken trachten de plenoptische functie (i.e. het licht dat in elke richting door elk punt van de scène passeert) te berekenen. Deze technieken slagen er dus in om het uitzicht van een 3D scène te modelleren zonder expliciet gebruik te maken van geometrie. Het wordt dus mogelijk om zeer complexe vormen en complexe lichteffecten die anders niet konden gemodelleerd worden toch voor te stellen. Deze technieken hebben echter eveneens een aantal andere nadelen: de nood aan gekende calibratie en pose, geen mogelijkheid tot navigatie in de scène, data intensief, enz. Een interessante aanpak zou erin kunnen bestaan om de twee technieken te combineren. In deze context werden reeds een aantal veelbelovende resultaten behaald met ons systeem. Een mogelijke aanpak zou het geometrische model gebruiken wanneer de terugprojectiefout in het beeld klein genoeg is en zou anders beroep doen op een plenoptische voorstelling. Het gebruik van tussenliggende voorstellingen kan eveneens interessant zijn (bv. gezichtspuntafhankelijke texturen).

Eén van de belangrijkste beperkingen van de meeste 3D opname-systemen is dat slechts statische scènes opgenomen kunnen worden. Het aanpakken van niet-statische onderwerpen is zeker een heel interessant maar ook een zeer complex probleem.

Totnogtoe zijn bestaande systemen meestal beperkt tot het fitten van parametrische modellen aan de beelddata. Sommige parameters kunnen dan een (niet-rigiede) pose voorstellen. Door gebruik te maken van resultaten op het bepalen van onafhankelijke beweging kan men zeker meer bereiken. Waarschijnlijk zullen echter meerdere aanwijzingen moeten gecombineerd worden om goede resultaten te bekomen (bv. toevoegen van contour informatie).

Een laatste belangrijk gebied voor verder onderzoek is het aanpassen van het systeem voor gebruik in nieuwe toepassingen. Een eerste algemene behoefte in deze context is een intelligente gebruikersinterface. Met intelligent wordt hier bedoeld dat het systeem zelf de nodige diagnose moet kunnen stellen als er iets mis gaat en dan met –verstaanbare– vragen of suggesties naar de gebruiker toe gaat. De interface van het systeem zou eveneens de verschillende verbanden tussen meerdere zichten en andere beperkingen op een voor de gebruiker transparante wijze moet opleggen. Een aantal meer specifieke ontwikkelingen naar verschillende toepassingsdomeinen zijn zeker ook interessant om te onderzoeken. In deze context is het belangrijk om het systeem modulair te houden zodat uitbreidingen en aanpassingen eenvoudig geïntegreerd kunnen worden.

Curriculum Vitae

Marc Pollefeys was born on 1 may 1971 in Anderlecht, Belgium. He studied at the Sint-Jan Berchmanscollege in Brussels. In 1994 he received a Masters degree in Electrical Engineering from the K.U.Leuven.

From 1994 to 1999 he worked towards a Ph.D. under the supervision of Prof. Luc Van Gool in the VISICS group (part of ESAT-PSI) at the K.U.Leuven. In this period he worked as a research engineer and as a research trainee of the IWT (Flemish Institute for Scientific Research in Industry). His main area of research is computer vision, 3D reconstruction from images and self-calibration.

In january 1998 he received the Marr prize at the International Conference on Computer Vision in Bombay for the paper “Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters” by M. Pollefeys, R. Koch and L. Van Gool.

List of Publications

Articles in International Journals

1. M. Pollefeys, R. Koch and L. Van Gool. *Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters*, accepted for publication in the International Journal of Computer Vision (07/01/1998).

2. M. Pollefeys and L. Van Gool, *Stratified Self-Calibration with the Modulus Constraint*, accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence (01/02/1999).

Articles in Proceedings of International Conferences

3. R. Koch, B. Heigl, M. Pollefeys, L. Van Gool and H. Niemann, *A Geometric Approach to Lightfield Calibration*, accepted for publication in Proceedings International Conference on Analysis of Images and Patterns, september 1999.

4. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, *An Automatic Method for Acquiring 3D models from Photographs: applications to an Archaeological Site*, accepted for Proceedings ISPRS International Workshop on Photogrammetric Measurements, Object Modeling and Documentation in Architecture and Industry, july 1999.

5. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, *Automatic Generation of 3D Models from Photographs*, Proceedings Virtual Systems and MultiMedia, Gifu Japan, 1998.

6. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, *Virtualizing Archaeological Sites*, Proceedings Virtual Systems and MultiMedia, Gifu, Japan, 1998.

7. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, *Metric 3D Surface Reconstruction from Uncalibrated Image Sequences*, Proc. SMILE Workshop (post-ECCV'98), LNCS 1506, pp.138-153, Springer-Verlag, 1998.

8. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, *Flexible acquisition of 3D structure from motion*, Proceedings IEEE workshop on Image and Multidimensional Digital Signal Processing'98, pp.195-198, Alpbach, 1998.

9. R. Koch, M. Pollefeys and L. Van Gool, *Automatic 3D Model Acquisition from Uncalibrated Image Sequences*, Proceedings Computer Graphics International, pp.597-604, Hannover, 1998.

10. L. Van Gool, F. Defoort, R. Koch, M. Pollefeys, M. Proesmans and M. Vergauwen, *3D modeling for communications*, Proceedings Computer Graphics International, pp.482-487, Hannover, 1998.

11. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, *Flexible 3D Acquisition with a Monocular Camera*, Proceedings IEEE Int'l Conf. on Robotics and Automation'98, Vol.4, pp.2771-2776, Leuven, 1998.

12. R. Koch, M. Pollefeys and L. Van Gool, *Multi Viewpoint Stereo from Uncalibrated Video Sequences*, Proc. ECCV'98, LNCS, Springer-Verlag, Freiburg, 1998.

13. M. Pollefeys, R. Koch and L. Van Gool. *Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters*, Proc.ICCV'98 (international Conference on Computer Vision), pp.90-95, Bombay, 1998. joint winner of the David Marr prize (best paper).

14. M. Pollefeys and L. Van Gool. *Self-calibration from the Absolute Conic on the Plane at Infinity*, Proc.CAIP97, LNCS vol.1296, pp.175-182, Kiel, 1997.

15. M. Pollefeys and L. Van Gool. *A stratified approach to metric self-calibration*, Proc.CVPR'97, pp.407-412, Puerto Rico, 1997.

16. M. Pollefeys, L. Van Gool, A. Oosterlinck. *The modulus constraint: a new constraint for self-calibration*, Proc.ICPR'96 (Int'l Conf. on Pattern Recognition), Vol.1, pp.349-353, Vienna, 1996.

17. M. Pollefeys, L. Van Gool, M. Proesmans. *Euclidean 3D reconstruction from image sequences with variable focal lengths*, Proc.ECCV'96, LNCS Vol.1064, Springer-Verlag, pp.31-42, Cambridge(UK), 1996.

18. M. Pollefeys, L. Van Gool, T. Moons. *Euclidean 3D reconstruction from stereo sequences with variable focal lengths*, Proc.ACCV'95 (Asian Conference on Computer Vision), Vol.2, pp.6-10, Singapore, 1995

Parts of Books

19. M. Pollefeys, M. Proesmans, R. Koch, M. Vergauwen and L. Van Gool, *Detailed model acquisition for virtual reality*, in J. Barcelo, M. Forte and D. Sanders, “Virtual Reality in Archaeology”, to appear, ArcheoPress, Oxford.

20. M. Pollefeys, L. Van Gool, A. Oosterlinck. 1997. *Euclidean self-calibration via the modulus constraint*, in F. Dillen, L. Vrancken, L. Verstraelen, and I. Van de Woestijne, ” Geometry and topology of submanifolds, VIII” World Scientific, Singapore, New Jersey, London, Hong Kong, pp.283-291.

21. M. Pollefeys, L. Van Gool, T. Moons. *Euclidean 3D reconstruction from stereo sequences with variable focal lengths*, Recent Developments in Computer Vision, LNCS Vol.1035, pp.405-414, Springer-Verlag, 1996.
22. T. Moons, L. Van Gool, and M. Pollefeys. 1996. *Geometrical structure from perspective image pairs*, in F. Dillen, L. Vrancken, L. Verstraelen, and I. Van de Woestijne, " Geometry and topology of submanifolds, VII" World Scientific, Singapore, New Jersey, London, Hong Kong, pp.305-308

Abstracts

23. M.Pollefeys, M.Proesmans, R.Koch, M.Vergauwen and L. Van Gool. *Flexible 3D reconstruction techniques with applications in archeology*, 26th. Int'l Conf. on Computer Applications in Archeology, Barcelona, 1998.

Internal Reports

24. M. Pollefeys, R. Koch and L. Van Gool. *Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters*, Technical Report Nr. KUL/ESAT/MI2/9707, MI2-ESAT, K.U.Leuven, Belgium, 1997.
25. M. Pollefeys and L. Van Gool. *A stratified approach to metric self-calibration*, Technical Report Nr. KUL/ESAT/MI2/9702, MI2-ESAT, K.U.Leuven, Belgium, 1997.
26. M. Pollefeys, L. Van Gool and Andre Oosterlinck. *Self-calibration with the modulus constraint*, Technical Report Nr. KUL/ESAT/MI2 /9609, MI2-ESAT, K.U.Leuven, Belgium, 1996.
27. M. Pollefeys, L. Van Gool, M. Proesmans. *Euclidean 3D reconstruction from image sequences with variable focal lengths*, Technical Report Nr. KUL/ESAT/MI2/9508, MI2-ESAT, K.U.Leuven, Belgium, 1995.

On-line tutorial

28. Marc Pollefeys, Reinhard Koch, Maarten Vergauwen and Luc Van Gool, *Metric 3D reconstruction from uncalibrated image sequences*, in CVonline: System Models, Calibration and Parameter Estimation:Uncalibrated Vision, 1998.
    `http://www.dai.ed.ac.uk/CVonline/`

Master thesis

29. Marc Pollefeys, *Algoritmen voor Radiale Basis Functie Neurale Netwerken*, Master Thesis Electrical Engineer, ESAT, K.U.Leuven, 1994.