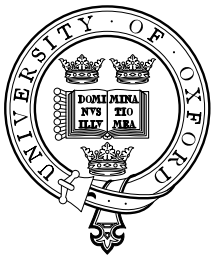
Symmetric Objects in Multiple Affine Views

Torfi Thórhallsson

Balliol College



Robotics Research Group
Department of Engineering Science
University of Oxford

This thesis is submitted to the Department of Engineering Science, University of Oxford, for the degree of Doctor of Philosophy. This thesis is entirely my own work, and, except where otherwise indicated, describes my own research.

Torfi Thórhallsson
Balliol College

Doctor of Philosophy
Trinity Term, 2000

Symmetric Objects in Multiple Affine Views

Abstract

This thesis is concerned with the utilization of object symmety as a cue for segmentation and object recognition. In particular it investigates the problem of detecting 3D bilaterally symmetric objects from affine views.

The first part of the thesis investigates the problem of detecting 3D bilateral symmetry within a scene from known point correspondences across two or more affine views. We begin by extending the notion of skewed symmetry to three dimensions, and give a definition in terms of degenerate structure that applies equally to an affine 3D structure or to point correspondences across two or more affine views. We then consider the effects of measurement errors on symmetry detection, and derive an optimal statistical test of degenerate structure, and thereby of 3D-skewed symmetry. We then move on to the problem of searching for 3D skewed symmetric sets within a larger scene. We discuss two approaches to the problem, both of which we have implemented, and we demonstrate fully automatic detection of 3D skewed symmetry on images of uncluttered scenes. We conclude the first part by investing means of verifying the presence of bilateral rather than skewed symmetry in the Euclidean space, by enforcing mutual consistency between multiple skewed symmetric sets, and by drawing on partial knowledge about the camera calibration.

The second part of the thesis is concerned with the problem of obtaining feature correspondences across multiple affine views, as required for the detection of symmetry. In particular we investigate the geometric matching constraints that exist between affine views. We start by specilizing the four projective multifocal tensors to the affine case, and use these to carry the bulk of all known projective multi-view matching relations to affine views, unearthing some new relations in the process. Having done that, we address the problem of estimating the affine tensors. We provide a minimal set of constraints on the affine trifocal tensor, and search for ways of estimating the affine tensors from point and line correspondences.

Submission date: 6 October 2000
Minor corrections: 11 October 2013

Acknowledgements

First, I would like to express my gratitude to my supervisor David W Murray for his gentle guidance and unfailing support. I would like to extend my gratitude to Andrew Zisserman for his continued interest and constructive comments. The researchers, fellow students, and the numerous visitors in the Robotics Research Group have provided rich inspiration and fruitful discussions. I am particularly indebted to my colleagues in David's Active Vision Laboratory. It has been a privilege to have daily discussions with Phil, Ian, John, Jace, Andy, Lourdes, Eric, and Ben to name but a few.

The Robotics Research Group has been a vastly enjoyable group of people to work with, both in Oxford and away on conferences. The Ultimate Robots team deserves special mention for determination and extreme fun both on and off the pitch. I would also like to thank my wonderful fellow students and the friendly staff at Balliol College for the enjoyable evenings at Holywell Manor. Special thanks go to the BCBC and my fellow oarsmen and coxes for those unforgettable early morning outings.

My deepest thanks go to my beloved Sigga and our children for the life and happiness they have brought. Our friends in Oxford have made our stay all the more enjoyable. Finally, I would like to warmly thank my parents for their continued encouragement.

My studies were made possible through the generous awarding of a Helga Jónsdóttir and Sigurliði Kristjánsson Memorial Grant, a Joint British Chevening / Ístak Scholarship, and an Overseas Research Student Award. The work on the affine matching constraints was supported by Grant GR/L58668 from the Engineering and Physical Sciences Research Council.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Approach 5

1.3 Thesis Overview 8

I Detecting Bilateral Symmetry **10**

2 Detecting 3D Skewed Symmetry **11**

2.1 Overview 11

2.2 Introduction 11

2.3 3D Bilateral Symmetry in a Single Affine View 16

2.4 3D Skewed Symmetry 19

2.4.1 Skewed Symmetry 19

2.4.2 Point Configuration and Degenerate Structure 21

2.4.3 Alternative Formulations 22

2.5 3D Skewed Symmetry in Multiple Affine Views 24

2.5.1 Degenerate Structure in the Affine Joint Image 24

2.5.2 Skewed Symmetry in Multiple Affine Views 26

2.5.3 Degenerate motion 29

2.6 Conclusions 30

3 Optimal Detection of Degenerate Structure from Affine Views **32**

3.1 Overview 32

3.2 Introduction 33

3.3 Measurement Models 36

3.3.1 Parametric Model 36
3.3.2 Normal Form: Linear Restrictions 40
3.4 Optimal Estimation of Affine Structure 41
3.4.1 Error Structure and Non-Existence of ML Estimators 43
3.4.2 MLE for Error Covariance Known up to Scale 44
3.4.3 Optimal Recursive Estimation 49
3.5 Testing for Dimensionality 56
3.5.1 Likelihood Ratio Test for Known Error Variance 57
3.5.2 Applicability to Small Sample Sizes 60
3.5.3 Likelihood Ratio Test for Error Variance Known up to Scale 65
3.5.4 Summary and Discussion 71
3.6 Conclusions 72

4 Searching for Symmetry within a Scene **74**
4.1 Overview 74
4.2 Introduction and Problem Definition 75
4.3 Deterministic Search 77
4.3.1 Sequential Application of Skewed Symmetry Constraints 77
4.3.2 Constraint Satisfaction and Directional Partitioning 79
4.3.3 Choosing the Sharpness of Statistical Tests 81
4.4 Implementation and Experiments 82
4.4.1 Automatic Detection from Interest Points 84
4.4.2 Performance Characterization of Detector in Noise 91
4.5 Searching by Random Sampling 96
4.6 Implementation and Results 101
4.6.1 Implementation 101
4.6.2 Results 103
4.7 Constraining the Search Further 107
4.7.1 Additional Constraints for Pruning or Verification 107
4.7.2 Making Use of Weak Constraints 108
4.8 Conclusions 110

5 Verification of Euclidean Symmetry **111**

5.1 Overview . . . 111
5.2 Introduction . . . 112
5.3 Euclidean Consistency from Three Symmetric Sets . . . 113
5.4 Testing for Orthogonality . . . 114
5.4.1 Distribution of the Inner Product . . . 117
5.4.2 Testing Relations for Consistency . . . 120
5.4.3 Unbiased Estimator of the Variance . . . 121
5.4.4 Computing the Likelihood Ratio . . . 123
5.5 Degenerate Arrangements of Symmetry Planes . . . 124
5.6 Incorporating Partial Camera Calibration . . . 126
5.6.1 Self-Calibration using Assumptions about Camera Intrinsics . . . 127
5.6.2 Constraints from the Weak Perspective Camera . . . 130
5.7 Testing for Orthogonality using Partial Calibration . . . 132
5.7.1 Mixing Constraints on Ω and Ω^* . . . 132
5.7.2 Expressing Constraints in the Dual Conic Ω^* . . . 133
5.7.3 Dual Likelihood Ratio Test . . . 133
5.8 Conclusions . . . 139

II Affine Matching **141**

6 Affine Multilinear Matching Constraints **142**

6.1 Overview . . . 142
6.2 Introduction . . . 143
6.3 The Registered Multifocal Tensors . . . 145
6.4 Affine Specialization of Projective Multifocal Tensors . . . 148
6.4.1 The Unifocal Tensor . . . 149
6.4.2 The Bifocal Tensors . . . 150
6.4.3 The Trifocal Tensors . . . 151
6.4.4 The Quadrifocal Tensor . . . 152
6.5 Transfer Equations and Incidence Relations . . . 156

6.5.1 Bifocal Transfer Equations . . . 157
6.5.2 Bifocal Incidence Relations . . . 158
6.5.3 Trifocal Transfer Equations . . . 158
6.5.4 Trifocal Incidence Relations . . . 160
6.5.5 Quadrifocal Transfer Equations . . . 164
6.5.6 Quadrifocal Incidence Relations . . . 165
6.6 Conclusions . . . 167

7 On the Estimation of the Affine Multifocal Tensors 168
7.1 Overview . . . 168
7.2 Introduction . . . 169
7.3 Over-parameterization of the Affine Tensors . . . 169
7.4 Constraints on the Affine Trifocal Tensor . . . 170
7.5 ML Estimation of Tensors from Point Matches . . . 175
7.5.1 Estimation of the Unifocal Tensor . . . 175
7.5.2 Estimation of the Bifocal Tensor . . . 175
7.5.3 Estimation of the Trifocal Tensor . . . 178
7.5.4 Estimation of the Quadrifocal Tensor . . . 180
7.6 Estimation of a Constrained Trifocal Tensor using Lines . . . 181
7.7 Affine Triangulation . . . 182
7.8 Conclusions . . . 184

8 Conclusions 185

III Appendices 188

A The Affine Camera and Projective Geometry 189
A.1 The Projective Camera . . . 190
A.1.1 Geometric Interpretation of the Projective Camera . . . 191
A.2 Fixed Sets and Calibration . . . 191
A.2.1 The Absolute Disk and its Image . . . 192
A.2.2 The Camera Cone and its Image . . . 193

A.3 The Affine Camera 195
A.3.1 Geometric Interpretation of the Affine Camera 195
A.3.2 The Affine Image of the Absolute Disk 196

B Notes on Affine Multifocal Relations **197**
B.1 Expansion of the Irreducible Tensor Components in Terms of the Registered Tensors 197
B.2 Relations to Known Trilinear Constraints 198
B.2.1 Torr's Affine Trifocal Constraint 198
B.2.2 Quan's Affine Trilinear Form 200

C The Quadratic p-relations of Four Views **201**
C.1 The Quadratic p-relations 201
C.2 The Quadratic p-relations of Four Registered Views 203

1

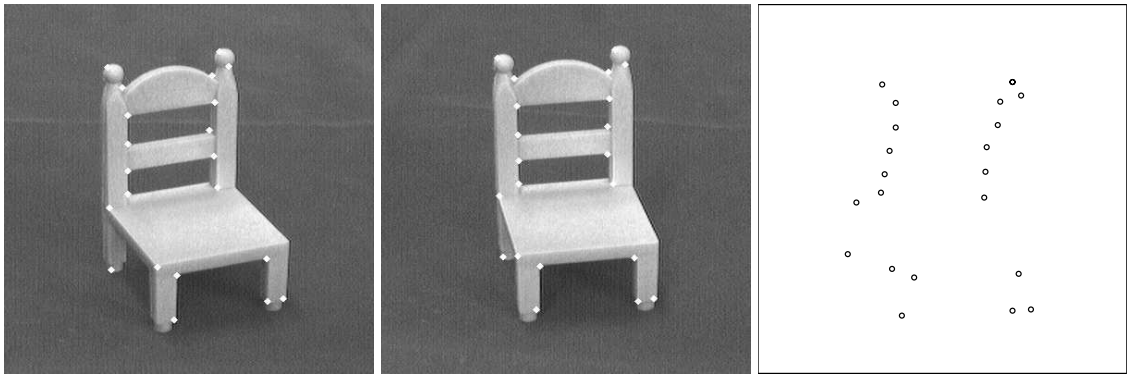
Introduction

A geometry, Klein said, is defined by a group of transformations, and investigates everything that is invariant under the transformations of this given group. Of symmetry one speaks with respect to a subgroup γ of the total group. Finite subgroups deserve special attention. A figure, i.e. any point-set, has the peculiar kind of symmetry defined by the subgroup γ if it goes over into itself by the transformations of γ.

Hermann Weyl: Symmetry (1952)

1.1 Motivation

Recognizing three dimensional objects from images by computer has proven to be one of the hard problems of artificial intelligence. One might think that it could be solved by simply matching an image to a library of stored images. Consider, however, the task of recognizing an object as “a house” or “a chair”. This task is confounded by the enormous variety of objects pertaining to these classes. But even if the intention is only to recognize a particular house or a particular chair, different viewing angles and optics will result in radically different images. Partial occlusion of objects is often observed as well. Even in



(a) Two views with matched corner features (b) Resulting 3D affine structure

Figure 1.1: A typical corner representation of a 3D object from uncalibrated affine views, automatically extracted using well known methods [HS88, BZM95, TK92].

the very special case of a fixed viewpoint, differences in illumination between images taken at different times will render direct matching of image intensity unreliable.

Computational Considerations

To limit the variability in the input data, it is common practice to make do with matching only those image features which are least susceptible to variation. To this end it is common to use only the location of abrupt changes of image intensity, as this information is likely to be largely independent of illumination. This leads to a sparse image representation typically consisting of curved contours, straight edges, or corners. An example is shown in Figure 1.1. The process of obtaining these via differentiation is sensitive to measurement noise, with the result that contours break up and spurious features are introduced. The recognition task then consists in matching these image features to a set of stored object descriptions. Direct matching of image features to a 3D object model means that the pose of the object must be solved for to verify each match, and this must be repeated for each model in the object library [Low85].

To limit the variability due to different viewpoints as well as illumination a different type of feature has been used for matching. These are particular *viewpoint invariant* algebraic functions of the location of a set of features. Consequently, no knowledge of the object pose or the camera calibration is needed in order to retrieve the set of features [MZ92b]. The lower cost of computation achieved by those methods, however, comes at the cost of

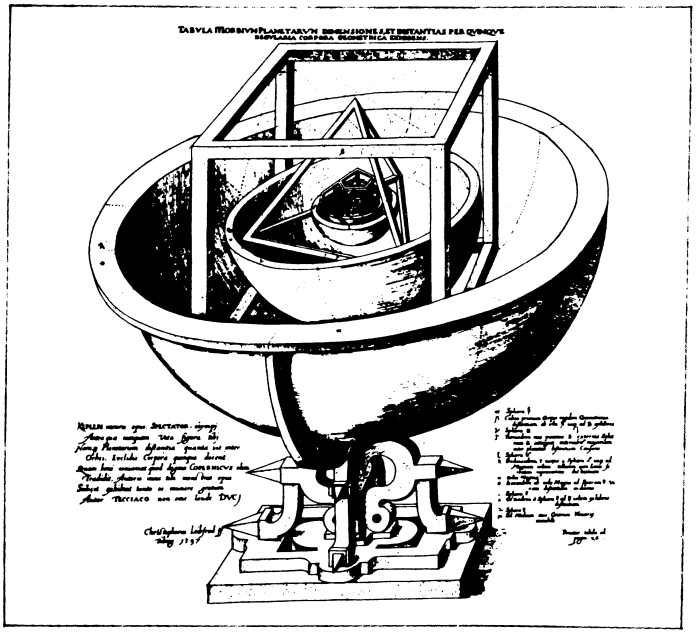
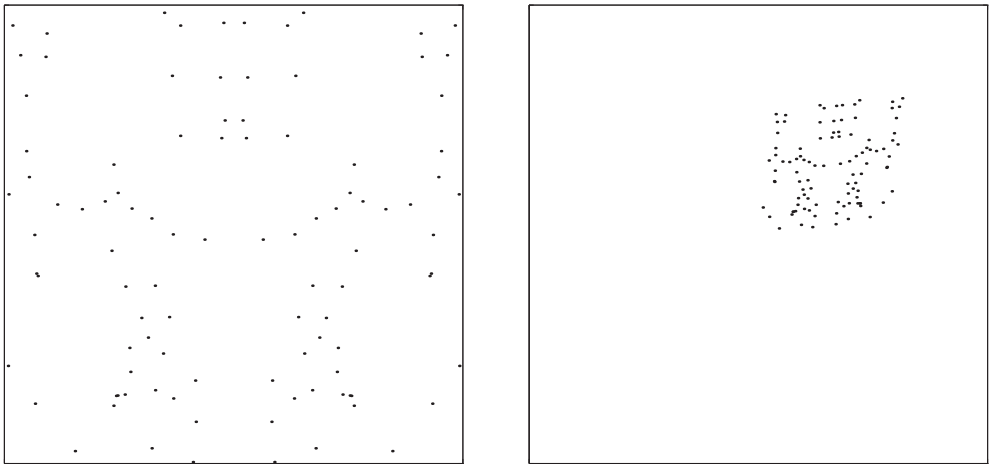


Figure 1.2: Powerful as it may be when appropriately used, symmetry is not the begin all and end all of object representation. Depicted is Kepler's somewhat inappropriate first cosmological model [Kep96] based on the regular solids. Among his many successful creations is the line at infinity, without which modern computer vision would be unthinkable.

lower descriptive power and therefore more ambiguous matching.

To increase the richness of the description, we could look towards using regularities in object structure, which carry to the location of image features. This is not a new idea, but one which has been studied for a long time under the term *grouping*. In its most basic form grouping gives preference to simple arrangement of features in an image such as coincidence and collinearity [Low85]. Taking this idea one step further, one can look for stronger regularities. Noting that collinearity corresponds to invariants of particular infinite subgroups of the Euclidean transformations of space, one is naturally led to the finite subgroups, the *symmetry groups*.

Symmetry is of particular interest for its relation to viewpoint or pose invariance. Depending on the sensor model, viewpoint invariance implies invariance to any transformation in a group of projective, affine, or Euclidean transformations of space. Similarly, symmetry is defined as invariance to a *subgroup*, ie. a subset of the above set of transformations [Wey52], as quoted at the start of the chapter. The connection to viewpoint invariance is obvious, and in a sense we can speak of symmetry as the geometric property closest in gen-



(a) bilateral symmetry (b) skewed symmetry

Figure 1.3: Bilateral symmetry (a) in planar point sets is perceived effortlessly by the human visual system. The same goes for skewed symmetry (b), although the performance drops significantly with increasing skewing angle.

erality to viewpoint invariance. Many man made and natural objects display symmetry to an extent which, if robustly detected, could provide cues for scene segmentation and object recognition. Our desire is to be able to detect 3D object symmetry, bilateral symmetry in particular, with this aim in mind.

Psychophysical Considerations

The ability of the human visual system to *preattentively* detect bilaterally symmetric patterns in the visual field has long intrigued researchers, as the reader can experience from Figure 1.3 (a). Over a century ago Ernst Mach [Mac86] observed how the response is not isotropic, but is strongest for a vertical axis of symmetry. In recent years Barlow and Reeves [BR79] have shown that the detection of bilaterally symmetric point patterns is very robust to noise, both in the form of clutter and perturbation of the pattern, although Jenkins [Jen83] has shown that even the smallest deviation from perfect symmetry does not go unnoticed. As for skewed symmetry, such as would arise in the visual field, when a planar bilateral pattern is presented under an oblique angle (see Figure 1.3 (b)), Wagemans and coworkers [WVd91, WVd92] showed this ability to remain, although performance de-

creases significantly with increasing skewing angle[1]. For an excellent review of empirical evidence and computational models of human perception of bilateral and skewed symmetry see [Wag95].

The importance of *object* symmetry to human viewers has been demonstrated to an extent in experiments on the continuation of partially (self-) occluded objects [Joh57, Bow66, Bow74]. Johansen [Joh57], in particular, showed how human viewers appear to envisage the far side of opaque polyhedral objects such as to maximize the symmetry of the object.

1.2 Approach

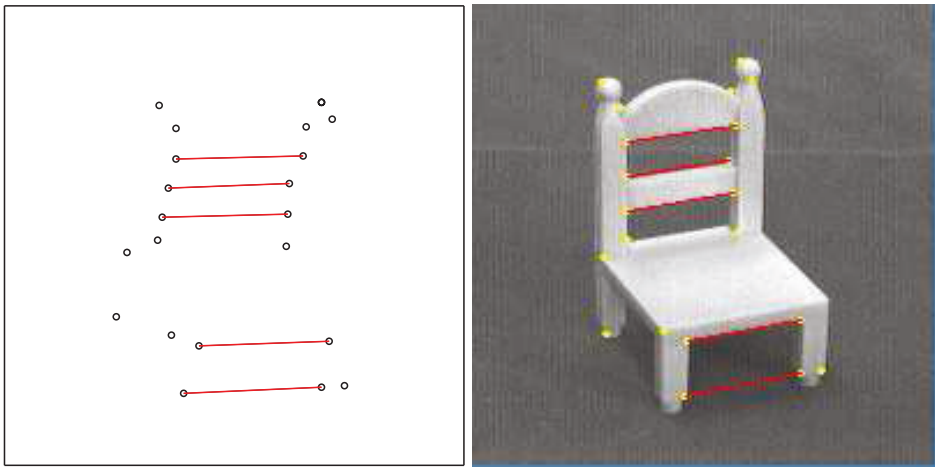
Prior knowledge of bilateral object symmetry has been shown to enable the recovery of 3D structure information from a single view [Kan81, RFZM93, FZB94, VP96]. *Detecting* 3D symmetry from a single view is a very difficult problem, on the other hand, because of the lack of constraints available. Consequently, the bulk of the work published on bilateral symmetry detection falls roughly into two categories: (*i*) planar symmetry from a single view, and (*ii*) 3D symmetry from a (roughly) segmented metric 3D description or CAD model.

Detecting 3D Bilateral Symmetry

We acknowledge the problems involved in detecting 3D symmetry in a single view of unconnected point features, and propose a method for detecting 3D bilateral symmetry from *two or more* affine views, with the aim of providing information to aid scene segmentation and object recognition, rather than reconstruction. We consider only bilateral object symmetry, which arguably is the most common form of symmetry, and also features in quite a few of the larger symmetry groups.

Rather than assuming the existence of a metric 3D reconstruction of the scene, we take as an input the more readily available affine structure, ie. a set of 3D point locations known only up to an affine transformation. We extend the notion of *skewed symmetry* to 3D affine

[1] Among evolutionary explanations of these findings is that they are evidence of a "someone looking at you" detector, essential for survival in the company of predators [Tyl96b].

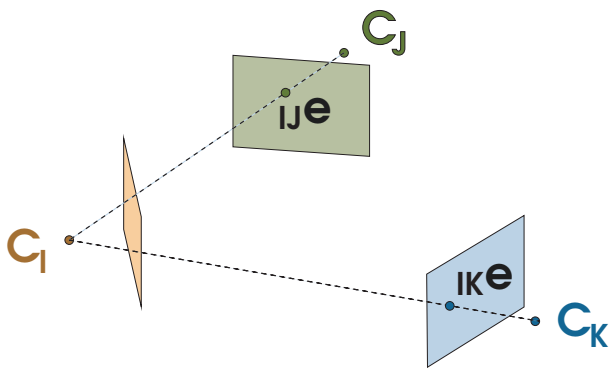


(a) From an affine structure. (b) From two affine views.

Figure 1.4: 3D skewed symmetry automatically detected (a) from an affine structure, and equivalently (b) from two affine views without reconstruction, by the methods described in this thesis.

structure, thereby reducing the problem of symmetry detection to that of detecting geometric degeneracies in sets of 3D points. We develop a method to determine such degeneracies in the presence of non-isotropic uncertainties in point locations. The information contained in a 3D affine structure can be obtained from two affine views of the scene [Kv91], as shown in the example in Figure 1.1. This fact has been used to perform computations on a 3D structure in the image domain, without explicit 3D reconstruction [RM96]. Proceeding in that spirit, we give an alternative method to determine the geometric degeneracies directly from point correspondences in two or more affine views.

We describe algorithms to detect sets of four or more bilaterally symmetric point pairs, contained within a larger point set. The algorithms employ successive degeneracy testing, and work equally on 3D affine structure and point correspondences in two or more affine views. The detection methods do not require prior segmentation of the scene. They are robust to occlusion and erroneous matching of features, requiring only that at least four symmetric point pairs are visible and correctly matched between the views. By the application of successive testing, the complexity of the search for symmetry is greatly reduced. We present the results of a performance analysis under different noise conditions, as well as examples of applying the algorithms on real views of uncluttered scenes, such as the one shown in Figure 1.4.



$$ {}_{JK}^{I}\mathbf{t}_{jk}^{i}\ {}_{IJ}e^{j}\ {}_{IK}e^{k} = 0^{i}, \qquad i \in \{1, 2\}. $$

Figure 1.5: The necessary and sufficient constraints on the elements of the affine trifocal tensor. Geometrically, the constraint states the fact that the epipoles ${}_{IJ}e$ and ${}_{IK}e$ are the images of the centre of camera I.

Having detected skewed symmetric 3D point sets in an affine frame, we attempt to rectify the affine structure. From multiple 3D skewed symmetric point sets we construct constraints on the transformation between the affine and the Euclidean space, under the assumption of bilateral symmetry in the Euclidean frame. We then use those constraints to detect inconsistencies between the detected 3D skewed point sets, to verify the existence of a bilateral symmetry, and obtain a rectified, scaled Euclidean reconstruction of the entire scene. Optionally, we incorporate commonly available knowledge about the camera intrinsics to further constrain the problem.

Affine Matching Constraints

The above methods assume that the location of corresponding points between two or more affine views is known. This, so called *correspondence problem* is as old as computer vision. With increasing use of uncalibrated cameras in recent years, this problem has received renewed interest, resulting in novel robust algorithms formulated in terms of the multifocal matching tensors, not least the trifocal tensor. The main focus has been on developing algorithms suitable for the most general case of projective cameras. The projective model is not readily applicable to affine cameras, however, where additional constraints should be observed.

In the second part of this thesis we focus on the estimation of multifocal matching constraints between affine views. We specialize the projective multifocal tensors to the affine case. This enables us to obtain a simpler and more stable affine specialization of known projective relations connecting points and lines across two, three, and four views. One problem remains: the trifocal and quadrifocal tensors are an overparametrized representation of the corresponding multiple view geometry, a fact that hampers their estimation. Worse still, the non-linear constraints that exist between the tensor elements are only partly known. As a step towards mending this situation, we derive necessary and sufficient constraints on the affine trifocal tensor, as illustrated in Figure 1.5. We also show how the optimal estimation of the tensors from point correspondences is achieved through factorization and discuss the estimation of the trifocal tensor from line correspondences using the new constraints.

1.3 Thesis Overview

This thesis is divided into two parts. **Part I** investigates the problem of detecting 3D bilateral symmetry within a scene from point correspondences across two or more affine views, while **Part II** is concerned with the problem of obtaining those point correspondences. In particular, the geometric matching constraints existing between affine views are investigated.

The first part is divided into four chapters. **Chapter 2** extends the notion of skewed symmetry to three dimensions, and gives a definition in terms of degenerate structure that applies equally to an affine 3D structure or to point correspondences across two or more affine views. **Chapter 3** considers the effects of measurement errors on symmetry detection, and derives an optimal estimator of affine structure and its error covariance prior to giving statistical tests of degenerate structure. **Chapter 4** considers the problem of searching for 3D skewed symmetry within a scene. Two approaches to the problem are discussed, and automatic detection of 3D skewed symmetry is demonstrated on images of uncluttered scenes. **Chapter 5** considers means of verifying the presence of bilateral rather than skewed symmetry in the Euclidean space, either by common constraints between multiple symmetric sets, or by drawing on partial knowledge about the camera calibration.

The second part of the thesis is divided into two chapters. **Chapter 6** specializes the four projective multifocal tensors to the affine case, and carries the bulk of the known projective multi-view relations to the affine case. **Chapter 7** addresses the problem of matching features between views by estimating the tensors. In particular, the constraints on the affine trifocal tensor due to over-parameterization are given, and the optimal estimation from point correspondences is investigated. The thesis is concluded by a discussion in **Chapter 8**.

Part I

Detecting Bilateral Symmetry

2

Detecting 3D Skewed Symmetry

2.1 Overview

This chapter considers the problem of detecting bilateral object symmetry from point features in uncalibrated affine views or from point structure recovered up to a global affine transformation. Following an introduction to the problem in the next section, Section 2.3 discusses how bilateral symmetry manifests itself in a single affine view. In Section 2.4 the notion of *3D skewed symmetry* is introduced. It is shown how 3D skewed symmetry in a set of points can be described in terms of structure degeneracies expressed by the rank of an affine structure matrix. This makes it possible to carry the definition of 3D skewed symmetry directly to the affine joint image or measurement matrix, in Section 2.5, showing that 3D skewed symmetry is a property of multiple affine views.

2.2 Introduction

This section provides an introduction to the problem of symmetry detection and the main design issues affecting an algorithmic solution: forms of symmetry, viewing conditions, the scope of the detection method, and the choice of features.

Object Symmetry

Symmetry is a particular kind of regularity, which has been the subject of continued mathematical study since Euclid [Wey52, SG92]. In modern mathematics, symmetry is defined in terms of groups of transformations, and the study of symmetry is *group theory*. A particular type of symmetry is thus defined by the group of transformations, which each leave the object under consideration invariant. A geometric figure in Euclidean space thus has a particular type of symmetry if it goes over into itself under a particular sub-group of the Euclidean transformations of space, to which count the rigid motions and reflections. Since the sizes of the groups differ in terms of the number of different transformations they contain, some figures can be said to be more symmetric than others. One group of transformations can also be a sub-group of another, creating a hierarchy of symmetries. The resulting types of symmetry have been extensively catalogued. In computer vision there have been attempts at identifying where in this hierarchy a given shape belongs [JB91, ZPA95]. The larger the symmetry group, however, the more specialized the shape of the figure. The smaller symmetry groups thus apply to a larger class of figures, as well as being found within many of the larger groups. In this thesis, generality is of prime interest, and consequently the focus will be on one of the smallest symmetry groups: that of bilateral (or mirror) symmetry.

3D Bilateral Symmetry

Loosely speaking, bilaterally (or mirror) symmetric 3D objects have points arranged symmetrically about a plane. Algebraically, symmetry can be defined in terms of the transformations, other than the identity, which map the object onto itself. In the case of bilateral symmetry this is a single transformation, the *plane reflection*, mapping symmetrically arranged points onto each other:

> A point set is bilaterally symmetric if for each point $\mathbf{X}_i$ in the set there is also a point $\mathbf{X}'_i$ in the set such that $\mathbf{X}'_i = \mathsf{T}_r\mathbf{X}_i$, where the transformation T_r is in a suitably chosen Euclidean frame represented by the matrix
>
> $$\mathsf{T}_r \stackrel{\text{def}}{=} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.1}$$

In the canonical frame the plane reflection is seen to map the point $\mathbf{X} = (X, Y, Z)$ onto the point $\mathbf{X}' = (-X, Y, Z)$, fixing any point on the plane $X = 0$.

A geometric definition can be given in terms of the line segments or *chords*, connecting symmetric points $\mathbf{X}_i$ and $\mathbf{X}'_i$. A set of chords is thus bilaterally symmetric if

1. The midpoints of all chords are coplanar, and

2. the chords are perpendicular to the plane defined by the midpoints.

A direct consequence of the second statement is:

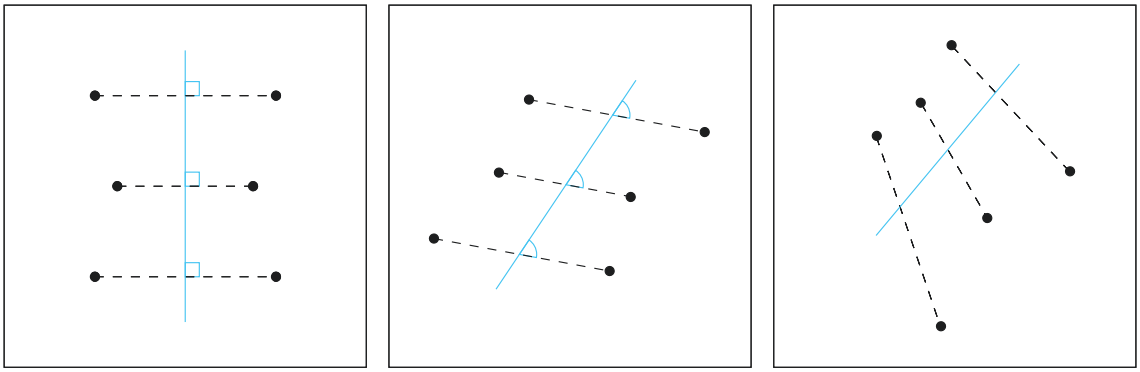
3. All chords are parallel.

Viewing Conditions

The geometric formation of a perspective image in a conventional camera is well described by the *pin-hole camera model*, in which rays of light reach an image plane via a single focal point. This assumes that any imperfections of the camera lens are negligible[1], or that they are known and have been corrected for. The formation of the perspective image in the pin-hole camera is a non-linear mapping between the Cartesian frames of the world and the image plane, which is effectively described by a projective transformation. In mathematics, *projective geometry* [SK52] is the study of those properties which are invariant to every projective transformation. An *uncalibrated* pin-hole camera constitutes an *unknown* projective transformation, known in computer vision as a *projective camera*. As a result the only properties of space known to be preserved in the resulting *projective views* is the projective geometry or the *projective structure* of the scene.

When perspective effects in the image are small, a linear approximation to the pin-hole model may be sufficient or even more appropriate. Some well known imaging conditions commonly described by linear models are *orthography*, *weak perspective* (or *scaled orthography*), and *para-perspective*. These are summarized in [SZB95]. All of these models can be described using affine transformations. In mathematics, *affine geometry* [SK52] is the study of those properties invariant to every affine transformation. An *uncalibrated*

[1] Strictly speaking the pin-hole assumption corresponds to the *thin-lens model* in optics, but can also be used for a camera better described by the *thick-lens model*, if we define the position of a camera in terms of the (primary) focal point.



(a) (b) (c)

Figure 2.1: Images of a bilaterally symmetric *planar* point set under different viewing conditions. The true Euclidean shape of the point set (a) exhibits *chords* connecting symmetric point pairs at right angles to an axis of symmetry. In an affine view (b) the chords remain parallel and the midpoint of each chord lies on the axis of symmetry, but the angle between the chords and the axis is unknown. In a projective view (c) none of the properties of an affine view hold. Instead the chords lie on concurrent lines intersecting in a single point some distance above the top edge of the image. This point and the symmetric point pair are harmonic with the axis of symmetry.

linear camera constitutes an *unknown* affine transformation, known in computer vision as an *affine camera* [MZ92a]. Consequently, the only properties of the scene known to be preserved in the resulting *affine views*, is the affine geometry of the scene or its *affine structure*. This also includes the projective geometry of the scene, since the group of affine transformations is included in the group of projective transformations.

Affine views have proven particularly well suited to the detection of bilateral symmetry in planar figures, depicted in Figure 2.1, a fact borne out by the sheer number of papers published on the subject. It will become apparent from this chapter that affine views are also particularly well suited to the detection of 3D bilateral symmetry.

Detection or Description

By a *symmetry detection method* we mean any method that searches the input data for (possibly small) sets of symmetric features. Each set of symmetric features can be either confined to a small area of *local symmetry* within the data set or, equally, span much of the extent of the input data.

A method which assumes as input a complete shape, and attempts to describe this shape

only in terms of its *global symmetry*, we will refer to as a *shape description method*, a term introduced by Kiryati and Gofman [KG98]. A shape description method can in some cases be used for detection, by exhaustively applying it to the *power set* (the set of all subsets) of the input features. This does not only set aside any considerations of efficiency, but also relies on the shape description method to return a measure a goodness of fit or the symmetricity of the shape, which is not always the case.

Features and Feature Extraction

Symmetry detection is essentially a matching problem, which involves matching symmetrically arranged features. Different types of features provide different information, thereby enabling or requiring different approaches to the matching problem. Features are also inherently different in their invariance to the different viewing conditions. The constraints provided by symmetry are naturally formulated in the language most appropriate for the features.

Of all the different approaches, the direct matching of image intensity has the advantage of doing away with a separate feature extraction stage, but it is very dependent on viewing conditions, and has so far only been reported for planar Euclidean symmetry. See [KG98] for a recent overview. Contour based methods can be divided into methods assuming discrete features, such as corners and line segments, and continuous curves. Point features have the advantage of being fully localized, but they convey the least information in terms of matching constraints. Straight line features provide one additional constraint, in that bilaterally symmetric lines are concurrent, intersecting on the plane of symmetry. Curve features provide the richest description, as encoded in the differential geometry of the curve, but curve matching is still a hard problem that brings complexities of its own [MZB95, VMUO95]. The knowledge of connected features provides information about the *topology* of the object, which can be utilized by graph-matching algorithms [JB91, ZW97]. The local connectivity of individual nodes can also be used to guide geometric matching [ZW97] to reduce the complexity of the matching. Symmetry detection using topology alone, however, has been shown [Man89] to be an NP-complete problem, with the exception of special shapes [JB91], and as such is only solvable by exhaustive methods.

The extraction of features is a separate detection problem, and different types of features

are more or less reliably detected. *Any symmetry detection method is thus only as accurate as its underlying feature detection method.* For a method spanning multiple views as is considered in this thesis, this is doubly true, since it relies on the features being consistently extracted and matched across multiple views, prior to the detection of symmetry. The only other method proposed to date, ie. that reported Zabrodsky and Weinshall [ZW94, ZW97], was presented using features extracted and matched by hand.

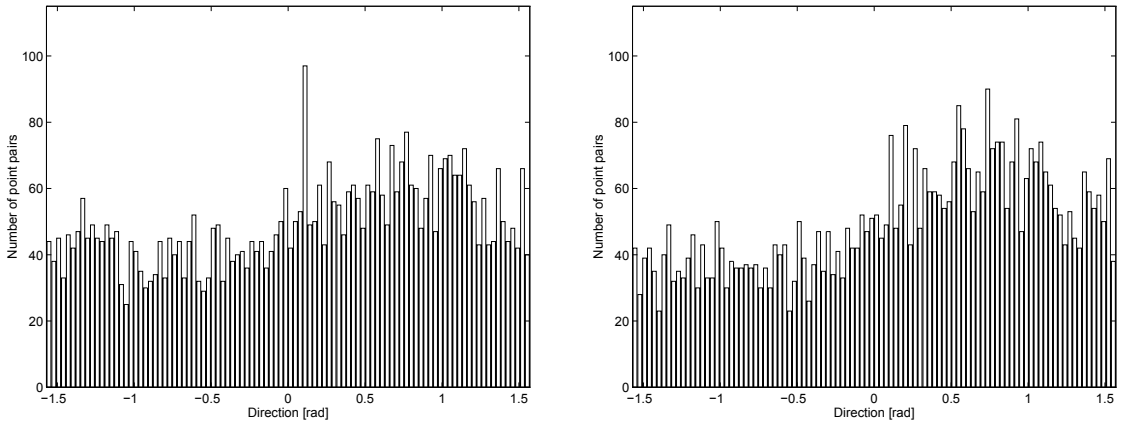
Considering the above, it is not surprising that existing approaches and proposed solutions to the detection of symmetry differ considerably, and that each method is generally inseparable from the type of features it assumes. The work in this thesis is limited to point features. The main reason for this is simplicity, but a secondary consideration is that by this choice one avoids making use of symmetry constraints specific to more descriptive features like lines or curves. The hope is then that any success with point features will indicate that even better results might be achieved using more descriptive features.

2.3 3D Bilateral Symmetry in a Single Affine View

Before proceeding to the main subject of this chapter we discuss the feasibility of detecting 3D bilateral symmetry from the location of point features in a *single* affine view.

In an affine view, the parallel chords on the object are by definition parallel in the image. Whether the midpoints are coplanar or not, cannot be determined from a single view however. Consequently, there is only one constraint available from a single affine view, that of parallel chords. This constraint was termed the *projected mirror-symmetry constraint* by Zabrodsky and Weinshall [ZW94, ZW97], who used it as a pruning constraint in an otherwise 3D Euclidean method. By itself, this constraint is not very powerful, as we shall now see.

Consider an affine view of a 3D point set consisting of n points, a certain proportion a ($0 \leq a \leq 1$) of which form bilaterally symmetric pairs, thereby defining $m = \frac{an}{2}$ chords. The total number of different point pairs that can be generated from the whole set is $\binom{n}{2} \approx \frac{n^2}{2}$. Each of these point pairs defines a line segment, a hypothetical chord, that has a particular direction in the image. Given that we are able to distinguish between at most k discrete line directions, due to finite measurement accuracy, then the average number of



(a) (b)

Figure 2.2: Histograms of the direction of line segments connecting each pair of points in two synthetically generated sets of 100 random points, distinguishing between $k = 100$ different directions. The first point set (a) is totally symmetric, that is every point has a symmetric counterpart ($a = 1$), and the direction of the chords stands out as a single large peak. In the second set (b) 50 of the points do not have symmetric counterparts ($a = 0.5$), and the peak due to the chords is reduced beyond distinction.

point pairs sharing any given direction i is $\overline{m}_i \approx \frac{n^2}{2k}$. With $a \cdot n$ truly symmetric points in the set, forming $m = \frac{an}{2}$ parallel chords, the expected ratio of symmetric chords to accidentally aligned points sharing the same direction is

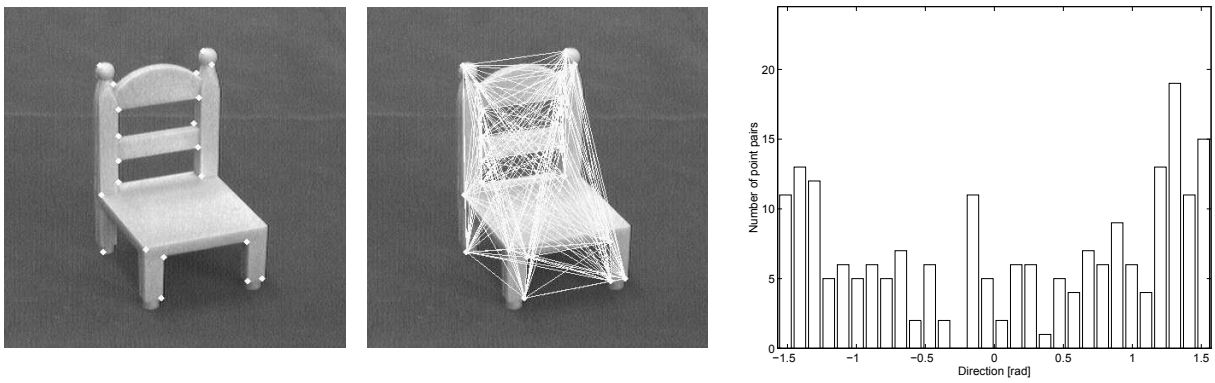
$$\frac{m}{\overline{m}_i} \approx \frac{ak}{n}. \tag{2.2}$$

From this equation and from Fig. 2.2, we observe the following:

- With k limited by measurement accuracy, (2.2) shows that the direction of symmetric chords will stand out less well as n increases, even if the set is fully symmetric ($a = 1$).

- Sharing the direction of the symmetric chords is an average number of $\overline{m}_i$ incorrect pairs, which are not resolved by the single constraint.

The actual number of accidentally aligned point pairs with a given direction is strongly dependent on the overall geometry of the point set. With the exception some highly uniform point sets, large deviations from the average number $\overline{m}_i$ just given are to be expected. Consequently, in the general case:



(a) (b) (c)

Figure 2.3: An example of a single view of a chair, under affine viewing conditions. Shown superimposed are (a) the 21 automatically detected corner features, twelve of which form six true chords, and (b) the 210 different line segments obtained by connecting any two features in (a). Finally (c) shows a direction histogram of the line segments in (b). The six symmetric chords are contained in the peak of eleven segments near the centre of the histogram.

- there do not exist any values of a, k, and n, for which the direction of symmetric chords is guaranteed to stand out as the most populated direction of point pairs in the image.

A Real Example

We demonstrate the problems associated with detecting bilateral symmetry of 3D objects on a single view of a chair shown in Fig. 2.3(a). The image is one of a pair of views used in later experiments (Chapter 4). To facilitate comparison between experiments we have used the same 21 corner features (shown in Fig. 2.3(a)) automatically detected using Harris' corner detector [HS88] and matched between the two views using the algorithm of Beardsley et al. [BZM95]. The viewing distance is approximately ten times the size of the object, making the assumption of an affine view a reasonable one.

The chair has a single global plane of bilateral symmetry, with 12 of the 21 points detected forming six symmetric chords. The proportion of symmetric points in the set is thus $a = 12/21 = 0.57$. The total number of different line segments connecting any two points in the set is $\binom{21}{2} = 210$. These are shown superimposed in Fig. 2.3(b), and a histogram of the direction of the segments is shown in Fig. 2.3(c). As can be expected from the high proportion of symmetric points, the direction of the chords is among the

higher peaks of the histogram, but it is still only the eighth highest, containing eleven line segments. Of these eleven, only six segments are true chords, and the identity of these is not resolved by the histogram.

From the discussion in this section we conclude that in the absence of additional constraints, such as local attributes or connectivity, the bilateral symmetry of an arbitrary 3D point set should not be expected to be determined with any degree of accuracy from a single affine view.

A Projective Note

Parallelism is not a projective property, and is therefore not preserved in projective views. There remains a constraint, however, as we shall now see. If parallel lines are regarded as intersecting at infinity [SK52], what remains in a projective frame, where the location of infinity is unknown, is the fact that the lines do intersect, or are *concurrent*. In a single projective view the symmetry chords therefore lie on concurrent lines. As every two lines on the plane intersect, this constraint is only meaningful for three or more chords, making the affine constraint computationally more efficient. The affine constraint of parallelism is also stronger in the sense that only those chords whose supporting lines intersect *at infinity* need to be considered. As a result the conclusions of this section apply even more strongly to single projective views of bilaterally symmetric point sets.

2.4 3D Skewed Symmetry

Having pointed out the lack of constraints available for the detection of 3D bilateral symmetry from a single view, we now turn to the problem of detecting 3D bilateral symmetry in a set of *3D points*, whose structure is known only up to an unknown affine transformation.

2.4.1 Skewed Symmetry

Looking at the geometric properties involved in the three statements about the chords in Section 2.2, we note that coplanarity is a projective property, while parallelism and the midpoint on a line segment are affine properties. Perpendicularity, on the other hand is a (scaled) Euclidean property. The affine structure of a bilaterally symmetric 3D object thus

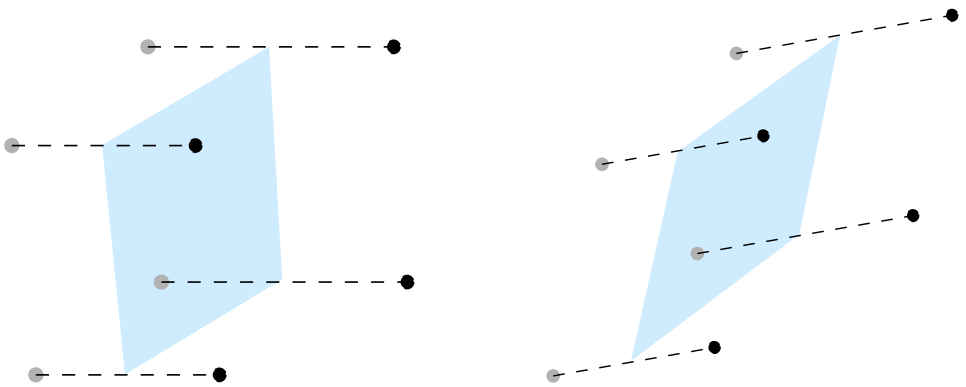


Figure 2.4: A bilaterally symmetric 3D point set (a) which undergoes an arbitrary nonsingular affine transformation (b) is termed *3D skewed symmetric*. The chords which connected symmetrically arranged point pairs remain parallel after the transformation, and the midpoints of the chords remain coplanar, but the chords are not orthogonal to this plane.

retains both the parallelism of chords and the coplanarity of the midpoints, and we have the two statements:

1. The chords are parallel and

2. the set of midpoints of the chords is coplanar.

This is illustrated in Figure 2.4. Compared to the single affine view, we have two properties to test for, in the place of one. Furthermore we have a stronger constraint on parallelism, as the chords must be parallel in 3D, rather than just in the single view. The only constraint lost in a reconstruction up to an unknown affine transformation is the perpendicularity of chords to the plane of symmetry. We will term a 3D point set satisfying the first two properties above *3D skewed symmetric*, in analogy to Kanade's [Kan81] definition of the affine view of a bilaterally symmetric planar figure.[2]

The main strength of the above definition of 3D skewed symmetry lies in the fact that the two properties can be tested for separately. A test of parallelism is particularly appealing, since it only involves comparing two hypothetical chords, and further efficiency can be expected from partitioning the set of possible chords according to spatial direction (more on this in Chapter 4).

[2]Kanade [Kan81] defined planar skewed symmetry assuming orthography, but his definition is equally valid under the weaker assumption of affine viewing conditions.

2.4.2 Point Configuration and Degenerate Structure

In order to get an algebraic handle on parallelism and coplanarity, we now provide a definition using the concept of degenerate structure. This particular definition proves to be a key enabling factor that allows us to carry 3D skewed symmetry into the image domain.

Given a number of 3D points $\mathbf{X}_p = \left[X_p, Y_p, Z_p\right]^T, p = 1, \ldots, N$, we can form the $3 \times N$ *structure matrix* (or *shape matrix*) consisting of the column vectors $\mathbf{X}_p - \mathbf{X}_0$

$$\Delta\mathsf{X} \stackrel{\text{def}}{=} \left[\begin{array}{cccc} \mathbf{X}_1 - \mathbf{X}_0 & \mathbf{X}_2 - \mathbf{X}_0 & \ldots & \mathbf{X}_N - \mathbf{X}_0 \end{array} \right], \tag{2.3}$$

where $\mathbf{X}_0$ is any linear combination of the N points, be it a single point or the centroid

$$\overline{\mathbf{X}} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{p=1}^{N} \mathbf{X}_p. \tag{2.4}$$

For convenience of notation, we treat the symbol Δ as an operator that can be applied to any matrix expression. We then have the relation between the rank of the structure matrix and the geometric configuration of the N points as given in Table 2.1.

$\operatorname{rank} \Delta\mathsf{X}$	configuration
3	general
2	coplanar
1	collinear
0	coincident

Table 2.1: The relation between structure degeneracy and the rank of the structure matrix.

The rank of a matrix can be determined using Gaussian elimination, or by counting the number of nonzero singular values. However if the matrix is corrupted by measurement error, it will in general be of full rank. The test of degenerate shape, as presented above, will thus not reveal the true configuration of the structure in the presence of noise. This problem is deferred to the next chapter.

Another Projective Note

In a 3D projective frame bilateral symmetry takes the following form:

1. The chords, or line segments, connecting symmetric points $\mathbf{X}_i$ and $\mathbf{X}'_i$ lie on concurrent lines, intersecting in a single point $\mathbf{D}$.

2. Each chord intersects the plane of symmetry in a point $\bar{\mathbf{X}}_i$ which is *harmonic* with respect to $\mathbf{X}_i$, $\mathbf{X}'_i$, and $\mathbf{D}$, meaning that the cross-ratio of the four points is -1.

As in the case of a single projective view, we have a concurrency relation between chords, in the place of parallelism. This is a stronger constraint in 3D projective space than in a single projective view, however, since two space lines do not necessarily intersect whereas planar ones always do, so there is already a constraint for two chords. The affine constraint is stronger nevertheless, since only those line segments need to be considered whose supporting lines intersect *at infinity*. The coplanarity constraint is the same as in the affine case, the only difference being the way in which the intersection with the plane of symmetry is computed.

2.4.3 Alternative Formulations

We will now discuss two alternative formulations of bilateral symmetry in affine structure, based on its definition as invariant to a plane reflection.

Conjugacy to a Reflection

The coordinate transformation that maps symmetrically related points in an affine frame is a *skewed reflection* T_s. In an affine frame having its origin on the plane of symmetry, as is accomplished by choosing the centroid of the symmetric point set as the origin, the skewed reflection can be decomposed into

$$\mathsf{T}_s = \mathsf{A}^{-1}\mathsf{T}_r\mathsf{A}, \tag{2.5}$$

where T_r is the plane reflection (2.1) and A is an invertible linear transformation that maps point coordinates from the affine frame into the canonical Euclidean frame. From the form of (2.5) linear algebra tells us that the skewed reflection T_s is *conjugate* to the reflection T_r [Blo79]. As a result the two transformations have the same eigenvalues, ie. -1, and a repeated 1.

The *skewed reflection* T_s, mapping symmetric points in an affine structure, has eigenvalues 1, 1, and -1.

This provides three constraints on the 3×3 matrix T_s, which is uniquely determined by three symmetric point pairs. To apply the above constraint as a test of skewed reflection, one would therefore need a minimum of three point pairs, as opposed to two pairs in the case of the parallelism constraint of skewed symmetry. A further argument in favour of skewed symmetry, is the ease with which the skewed symmetry constraints are expressed in terms of measurement errors, as is made evident in the next chapter.

3D Affine Harmonic Homology

A skewed reflection is known in projective geometry as a *3D affine harmonic homology* [SK52]. A 3D harmonic homology is completely characterized by (1) the fixed *vertex* $\mathbf{D}$, where chords intersect, which in the affine case is confined to the plane at infinity, (2) the *axial* or symmetry plane $\mathbf{\Pi}$ of fixed points, and (3) the *modulus* -1, the fixed harmonic cross-ratio of a symmetric point pair $\mathbf{X}_i$, $\mathbf{X}'_i$ with $\mathbf{D}$ and the intersection $\bar{\mathbf{X}}_i$ of the chord with the symmetry plane. The 3D affine harmonic homology thus has five degrees of freedom, two in the choice of $\mathbf{D}$ and three in the choice of $\mathbf{\Pi}$.

Equiaffinity and Involution

As an alternative to first forming a set of hypothetical chords, and then testing the set for compliance with skewed symmetry, it may also be possible to compute an invariant to any skewed reflection (2.5) from local features, and to compare its value across the structure in search of symmetric counterparts. In the context of skewed symmetric planar contours, Van Gool et al. [VMUO95] derived the following equivalent definition of 2D skewed reflections. The two properties also hold for 3D skewed reflections:

1. Skewed reflection is *equiaffine*, ie. an affine transformation with $\left\|\det \mathsf{T}_s\right\| = 1$.
2. Skewed reflection is an *involution*, ie. $\mathsf{T}_s\mathsf{T}_s = \mathsf{I}_3$.

For the case of planar contours, Van Gool et al. were able to compute local invariants from the differential geometry of the curve in a single point. In the case of unconnected 3D

points, considered here, equiaffinity means that volume is preserved, with the effect that the volume of the tetrahedron defined by any four point features is the same as that of the symmetrically opposite tetrahedron. Since four points are involved in the computation of the invariant, computational complexity is higher than in the case of planar curves, although additional efficiency can be gained by limiting the spatial extent of the tetrahedra considered.

2.5 3D Skewed Symmetry in Multiple Affine Views

We will now give a method to infer 3D bilateral symmetry of a set of points *directly from two or more affine views*. Using the definition of skewed symmetry given in the previous section, this only requires provision of a method to infer degenerate structure from two or more affine views.

2.5.1 Degenerate Structure in the Affine Joint Image

Whereas the effect of a single affine camera is to *project* the structure of three dimensional space onto a lower dimensional image space, with inevitable loss of information, the effect of multiple affine cameras is to *embed* the structure of three dimensional space in the higher dimensional measurement space of the joint image. Provided the motion between the viewpoints is non-degenerate this embedding incurs no loss of information, in the sense that the affine structure of the scene is completely contained in the joint image.

Given $F \geq 2$ affine views of a set of N scene points, whose correspondence between the views is known, we denote the image of the scene point $\mathbf{X}_p$ in view f by $\mathbf{x}_{fp} = \left[x_{fp}, y_{fp}\right]^T$. We form the form the $2F \times N$ *registered measurement matrix*

$$\Delta \mathsf{W} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{x}_{1,1} - \overline{\mathbf{x}}_1 & \dots & \mathbf{x}_{1,N} - \overline{\mathbf{x}}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{x}_{F,1} - \overline{\mathbf{x}}_F & \dots & \mathbf{x}_{F,N} - \overline{\mathbf{x}}_F \end{bmatrix},$$

where $\overline{\mathbf{x}}_f \stackrel{\text{def}}{=} \frac{1}{N} \sum_{p=1}^{N} \mathbf{x}_{fp}$ is the centroid of the image points in view f. Tomasi and Kanade [TK92] showed[3] that the measurement matrix $\Delta \mathsf{W}$ can be factorized into two ma-

[3]The measurement matrix, as defined here, is equal to the registered measurement matrix defined in [TK92] up to row ordering. This difference does not affect the cited results.

trices M' and $\Delta\mathsf{X}'$,

$$\Delta\mathsf{W} = \mathsf{M}'\Delta\mathsf{X}', \tag{2.6}$$

where M' is a $2F \times 3$ matrix representing the camera orientation in each view and $\Delta\mathsf{X}'$ is a $3 \times N$ matrix representing the structure matrix (2.3). Tomasi and Kanade showed by this that the matrix $\Delta\mathsf{W}$ can at most have rank three. Using (2.6) as a model of the imaging process

$$\Delta\mathsf{W} = \mathsf{M}\Delta\mathsf{X} \tag{2.7}$$

with the particular choice (2.4), it is easily shown that if one of the matrices on the right hand side, e.g. the matrix $\Delta\mathsf{X}$, is rank deficient then the rank of $\Delta\mathsf{W}$ is reduced accordingly. More precisely:

> **Extended Rank Theorem** *If the number of views $F \geq 2$ and the joint projection matrix M is of full rank, then the rank of the measurement matrix $\Delta\mathsf{W}$ equals the rank of the matrix $\Delta\mathsf{X}$.*

Proof. The matrix $\Delta\mathsf{X}$ has N columns. If its rank is r then there must exist $N - r$ independent vectors $\mathbf{k}_i \neq \mathbf{0}, (i = 1, \ldots, N - r)$ such that

$$\Delta\mathsf{X}\mathbf{k}_i = \mathbf{0}. \tag{2.8}$$

Multiplying (2.8) from the left by M and using (2.7) we similarly have that

$$\Delta\mathsf{W}\mathbf{k}_i = \mathsf{M}\Delta\mathsf{X}\mathbf{k}_i = \mathbf{0}$$

and thus $\Delta\mathsf{W}$ can have at most rank r.

If $F \geq 2$ and the $2F \times 3$ matrix M is of full rank, then its rank is three, and hence its $2F$ row vectors span a three dimensional subspace. The transformation defined by M thus preserves the dimensionality of the 3D structure contained in the columns of the structure matrix $\Delta\mathsf{X}$. The rank of $\Delta\mathsf{W}$ thus cannot be less than the rank of $\Delta\mathsf{X}$. □

It follows from the extended rank theorem that the relation between the rank of $\Delta\mathsf{X}$ and the spatial configuration of 3D points, as given earlier in Table 2.1, also holds when $\Delta\mathsf{X}$ is replaced by the matrix $\Delta\mathsf{W}$ as in Table 2.2. It should be noted that if the joint projection matrix M is rank deficient, then $\Delta\mathsf{W}$ and $\Delta\mathsf{X}$ will not necessarily be of the same rank. As in the case of 3D affine structure, this relation only holds for noise free measurement. In the next chapter we address the case of noisy measurement.

$\operatorname{rank}\Delta\mathsf{W}$	3D structure
3	general
2	coplanar
1	collinear
0	coincident

Table 2.2: The relation between degeneracy in a set of 3D scene points and the rank of the measurement matrix obtained from two or more views.

2.5.2 Skewed Symmetry in Multiple Affine Views

Using results from the preceding sections we will now provide an algebraic definition of skewed symmetry in terms of degenerate structure. Without loss of information, we can represent a pair of bilaterally symmetric points $\mathbf{X}$, $\mathbf{X}'$ in terms of the difference and the sum of their coordinates

$$\mathbf{X}_- \stackrel{\text{def}}{=} \mathbf{X} - \mathbf{X}' = \mathbf{X} - \mathsf{T}_r\mathbf{X} = (\mathsf{I}_3 - \mathsf{T}_r)\mathbf{X} \tag{2.9}$$

$$\mathbf{X}_+ \stackrel{\text{def}}{=} \mathbf{X} + \mathbf{X}' = \mathbf{X} + \mathsf{T}_r\mathbf{X} = (\mathsf{I}_3 + \mathsf{T}_r)\mathbf{X}, \tag{2.10}$$

where T_r is the plane reflection (2.1) and the frame is the canonical Euclidean frame with plane of symmetry ($X = 0$). Using (2.1) can write the transformations on the right hand side as

$$\mathsf{T}_- \stackrel{\text{def}}{=} \mathsf{I}_3 - \mathsf{T}_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.11}$$

$$\mathsf{T}_+ \stackrel{\text{def}}{=} \mathsf{I}_3 + \mathsf{T}_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{2.12}$$

Counting the linearly independent rows of the two matrices on the right hand side we have that the rank of T_- is one while the rank of T_+ is two. From (2.9) and (2.11) and similarly from (2.10) and (2.12) we then have

$$\operatorname{rank}\mathsf{X}_- = \operatorname{rank}(\mathsf{X} - \mathsf{X}') = \operatorname{rank}(\mathsf{T}_-\,\mathsf{X}) \leq \operatorname{rank}\mathsf{T}_- = 1 \tag{2.13}$$

$$\operatorname{rank}\mathsf{X}_+ = \operatorname{rank}(\mathsf{X} + \mathsf{X}') = \operatorname{rank}(\mathsf{T}_+\,\mathsf{X}) \leq \operatorname{rank}\mathsf{T}_+ = 2, \tag{2.14}$$

chord configuration	$\operatorname{rank}(\Delta\mathsf{X}_+ + \mathsf{X}_-)$	$\operatorname{rank}\Delta\mathsf{X}_+$	$\operatorname{rank}\mathsf{X}_-$
parallel asymmetric			
non-coplanar	3	3	1
coplanar	2	2	1
concurrent	1	1	1
parallel symmetric			
3D skewed	3	2	1
2D skewed	2	1	1
1D bilateral	1	0	1
repeated points			
non-coplanar	3	2	0
coplanar	2	1	0
collinear	1	0	0
coincident	0	0	0

Table 2.3: Classification of the spatial arrangement of a set of 3D affine point pairs based on the dimensionality of the set and the rank of the sum and difference matrices.

if X is not required to be full rank. As the rank of a matrix is preserved under any invertible linear transformation, including the conjugation (2.5), the above rank constraints equally apply in every affine frame with origin on the plane of symmetry. One way of guaranteeing that the origin of the affine frame lies on the plane of symmetry is by selecting the origin as the centroid (2.4) of the skewed symmetric point set.

The two constraints (2.13) and (2.14) state algebraically the geometric collinearity and coplanarity constraints presented on page 20. The inequalities in the constraints reflect the fact that the rank of X and X' is reduced if the point pairs under consideration lie in a plane or a smaller dimensional subspace. The inequalities can be replaced by an equality if the set of point pairs is non-coplanar. The resulting stronger constraint, however, excludes some valid 3D symmetric point configuration, such as planar symmetry. The constraint (2.13) and (2.14), includes these, but also some asymmetric point configurations, as seen in Table 2.3. To distinguish between symmetric and asymmetric configurations of lower dimensions, one needs to take into account the dimensionality of the point set. In

an arbitrary affine frame, this can be inferred from the rank of the $3 \times 2N$ structure matrix $\Delta[\mathsf{X}\ \mathsf{X}']$ containing the N point pairs and registered to the centroid. Equivalently, one can consider only the rank of the smaller $3 \times N$ matrix $\Delta\mathsf{X}_+ + \mathsf{X}_-$, which can be shown to be registered to the same point. From the classification in Table 2.3, we infer that the two constraints

$$\operatorname{rank} \mathsf{X}_- \leq 1 \tag{2.15}$$

$$\operatorname{rank} \Delta\mathsf{X}_+ = \operatorname{rank}(\Delta\mathsf{X}_+ + \mathsf{X}_-) - 1 \tag{2.16}$$

include all point sets that map setwise onto themselves under a 3D skewed reflection, while the stronger constraints

$$\operatorname{rank} \mathsf{X}_- = 1 \tag{2.17}$$

$$\operatorname{rank} \Delta\mathsf{X}_+ = \operatorname{rank}(\Delta\mathsf{X}_+ + \mathsf{X}_-) - 1 \tag{2.18}$$

exclude point sets that have no points off the plane of symmetry. A 3D skewed symmetric figure of a particular dimension d can be obtained by requiring

$$\operatorname{rank} \mathsf{X}_- = 1 \tag{2.19}$$

$$\operatorname{rank} \Delta\mathsf{X}_+ = d - 1 \tag{2.20}$$

$$\operatorname{rank}(\Delta\mathsf{X}_+ + \mathsf{X}_-) = d. \tag{2.21}$$

By the Extended Rank Theorem from the previous section, all the above rank constraints translate directly into two or more affine views, e.g.

$$\operatorname{rank} \mathsf{W}_- \stackrel{\text{def}}{=} \operatorname{rank}(\mathsf{W} - \mathsf{W}') = 1 \tag{2.22}$$

$$\operatorname{rank} \Delta\mathsf{W}_+ \stackrel{\text{def}}{=} \operatorname{rank} \Delta(\mathsf{W} + \mathsf{W}') = d - 1 \tag{2.23}$$

$$\operatorname{rank}(\Delta\mathsf{W}_+ + \mathsf{W}_-) = d, \tag{2.24}$$

where W and W' are measurement matrices containing the location of corresponding symmetric points in the affine joint image. Again, the origin of each image frame can be chosen as the centroid of the symmetric point set in order for the skewed symmetric rank relations to hold, as indicated by the registration operator Δ. It should be noted that the difference matrices X_- and W_- are by definition unaffected by a uniform translation of the point pairs, so that the constraints involving these matrices hold irrespective of the choice of origin.

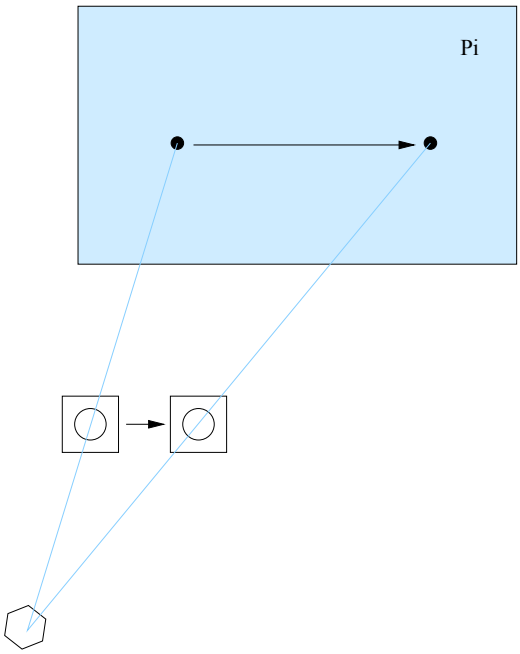


Figure 2.5: The only camera motion which is non-degenerate with respect to affine structure recovery is one that moves the affine centre of projection on the plane at infinity. In other words a motion that changes the aspect on the object.

2.5.3 Degenerate motion

The Extended Rank Theorem makes in possible to infer the degeneracy of a 3D structure directly from the joint image on the explicit condition that the joint projection matrix M is of full rank. In this sense, the motion between views is required to be non-degenerate. The conditions under which the motion of an affine camera is degenerate with respect to structure recovery have not been sufficiently clearly stated in the literature (see eg. [SZB95]).

The affine camera is a specialization of the projective camera model (see Appendix A for an overview of the releation between the two models). As such, one would expect that the only degenerate camera motion with respect to the affine recovery of point structure is one *which does not move the centre of projection of the affine camera in a reference frame fixed on the object*. But where is the camera centre of an affine camera located, and how does it move when we move the real camera being modelled?

We should start by noting that the affine camera model is commonly used as a (piecewise) approximation of central projection. This means that a bundle of optical rays intersecting at the camera centre $\mathbf{C}$ is approximated by a parallel bundle of rays intersecting in

some point $\mathbf{C}_a$ on the plane at infinity π_∞. This defines the image projection as affine, and the point $\mathbf{C}_a$ is the centre of projection (A.3) of the affine camera.

It now follows that any motion of the real camera that does not move $\mathbf{C}_a$ is degenerate with respect to affine structure recovery. This includes any rotation about $\mathbf{C}$ as well as any translation of $\mathbf{C}$ in the direction of the affine camera center $\mathbf{C}_a$, or any combination of the two. Since a point at infinity defines a direction, we have that the only motion of a camera under affine approximation which is non-degenerate with respect to point structure recovery is a motion which *changes the aspect on the object.* This is illustrated in Figure 2.5.

That this condition does indeed guarantee that the affine joint projection matrix M is of full rank, can be seen by an algebraic argument. This argument draws on the geometric interpretation of the affine camera given in Section A.3.1 in the Appendicies. Each affine camera contributes two rows to the joint projection matrix. The two rows consist of the 2D homogeneous coordinates of the two lines on π_∞ where the first two reference planes of the affine camera intersect with π_∞ - the third reference plane of every affine camera. The point of intersection of the three reference planes is the centre of projection $\mathbf{C}_a$. Now for M to be of full rank it must contain three linearly independent rows. This means that the three corresponding lines on π_∞ cannot be concurrent. In other words, they cannot intersect in a common point. It follows that the corresponding affine cameras cannot have a common centre of projection.

2.6 Conclusions

This chapter considered the problem of detecting bilateral symmetry in a set of 3D points whose location is known only up to a global affine transformation. In particular it has brought us the following:

1. The problem of detecting bilateral symmetry from a single affine view of a set of unconnected points is poorly constrained.

2. There are more constraints available in a 3D affine structure, and these can be described collectively by *3D skewed symmetry*.

3. By expressing 3D skewed symmetry in terms of degenerate structure, it was shown that 3D skewed symmetry is carried into two or more uncalibrated affine views *without loss of information*.

4. In the case of point features, the only degenerate motion of the affine camera with respect to the determination of 3D skewed symmetry is one which leaves its aspect on the object the same in all views.

The definition of degenerate structure in terms of the rank of a measurement matrix or a recovered structure matrix, which lies at the core of the proposed approach, assumes perfect measurements. The next chapter addresses the problem when the measurements are perturbed by noise.

3

Optimal Detection of Degenerate Structure from Affine Views

3.1 Overview

This chapter considers the problem of detecting a particular type of degeneracy in a set of 3D scene points, such as coplanarity or collinearity, when the points are known either in an affine 3D reconstruction or in two or more uncalibrated affine views. The exact location of the points is assumed to be concealed by measurement errors, whose variance may be unknown. Following an introduction to the problem in Section 3.2 and a discussion of suitable methodologies for solving it, a formalization of the problem is given in Section 3.3 in the form of a stochastic measurement model. Section 3.4 then derives optimal statistical estimators of the parameters of this model. The resulting ML estimators of affine structure and motion provide an estimate of the likelihood of the observations, which is required for subsequent testing. We also derive a second estimator for the case when an a priori structure estimate is available. In both cases, the solutions turn out to be familiar methods for affine structure recovery. Section 3.5 then presents the actual statistical tests of degeneracy. These are given for the two cases when the error covariance of the measurement errors is completely known and when it is known only up to scale.

3.2 Introduction

In the last chapter we proposed a method for determining whether an affine 3D point structure is skewed symmetric, based on the concept of degenerate structure as expressed in the rank of the structure matrix. It was also shown that the same method can also be applied directly to image locations in affine views of the points set, without prior reconstruction. It was noted that the dimensionality of the 3D structure is only reflected in the rank of the structure in the absence of measurement errors. For a practical method this will not do.

Aim: Testing for Rank in the Presence of Noise

Measurement errors will in general cause the measurement matrix and the recovered structure matrix to be of full rank, instead of the rank faithfully portraying the dimensionality of the imaged structure. In terms of the singular values of these matrices, this means that singular values are non-zero where they should vanish.

In an attempt to cope with this problem, one might consider choosing a small positive threshold value and taking any singular values smaller than this threshold as zero. This may prove sufficient for the odd test, but in the detection method proposed here (see Chapter 4), tests for collinearity and coplanarity are applied to point sets of different sizes and with varying measurement accuracy. This would require a string of different thresholds, whose values need determining, leading to a large number of twiddle-factors. A consistent, if not optimal, way of deciding on the dimensionality of a recovered structure is needed instead, and as one would expect the answer lies in statistics.

Method: Hypotheses Testing or Model Selection?

There are two questions about the dimensionality of an observed structure that may be of interest. The first question is of the type:

> Is the structure coplanar?

and the second question is:

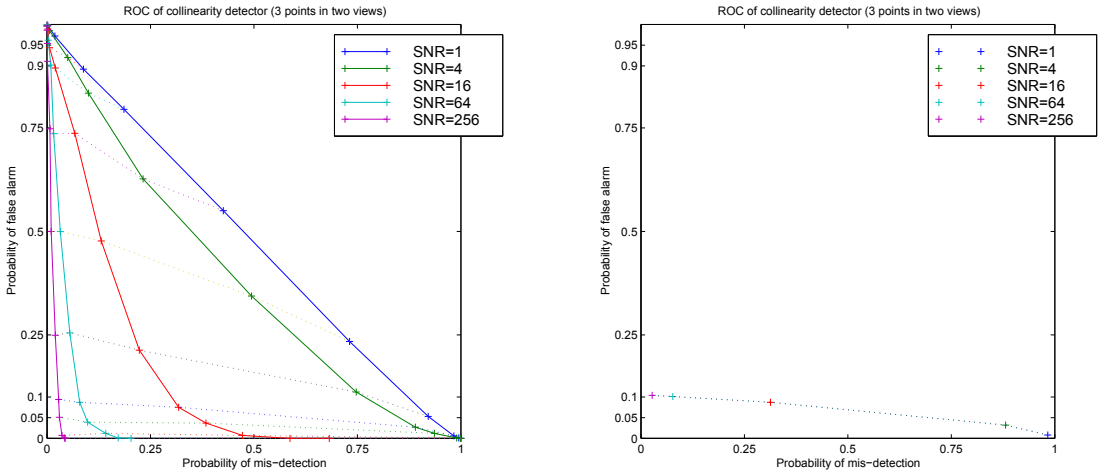
> What is the dimensionality of the structure?

Given the observed point locations and the probability distribution of the measurement errors the answer to each of these questions can be inferred using statistical methods. The means of doing so are different for each question, however, although the answers to both are based on fitting a model or multiple models of the appropriate dimensions to the data, and observing whether the residual error of the fit can be reasonably explained by the measurement errors.

The first question is answered by testing the hypothesis that a model of a specified dimension (two in the case of coplanarity) fits the observations, against the alternative hypothesis that a higher dimensional model is required. There is no single correct way of making that decision, but the answer that minimizes the probability of *mis-detection* at a set probability of *false alarm* is provided by the *likelihood ratio test* as given by the *Neyman-Pearson lemma* (see e.g. [BNB96]).

Turning to the second question, the problem is confounded by the well known fact (see [Tor98] for an in-depth discussion) that a higher dimensional model will always fit the data better than a lower dimensional model, regardless of the dimensionality of the structure. This problem was first addressed by Akaike [Aka74], whose model selection method is based on *an information criterion* (*AIC*), which has been shown [Boz87] to be an unbiased estimator of the mean expected log likelihood of each model. This provides the means to select the model (ie. the dimensionality of the structure) associated with the highest mean expected likelihood. Akaike's original method assumes that the models are *nested*, ie. that each model is obtained from the next more general model by setting some of its parameters to zero. In situations where the competing models are of a very different nature, more general methods, such as Kanatani's [Kan96] *geometric information criterion* have been shown [Tor98] to be of crucial importance. The models under consideration here are nested, so Akaike's model selection method applies without modification. An altogether different scoring function, based on *minimum description length*, is given in [TZM98] for use with robust estimators such as RANSAC [FB81].

The simple relation between the models may also allow us to perform a likelihood ratio test on each of the hypothesized models, and armed with Occam's razor to slice away all but the lowest dimensional model surviving the likelihood ratio test. The statistical foundations of this procedure, first suggested by Tintner [Tin46] and described in Section 3.5.1 below,



(a) likelihood ratio tests (b) AIC model selection

Figure 3.1: In the detector based on likelihood ratio tests (a) the probability of the two types of detection errors, *false alarm* and *mis-detection*, can be traded against each other. The significance level α chosen for the hypothesis tests sets the probability of false alarm to approximately $1 - \alpha$ (dotted lines), the probability of mis-detection varying with the signal to noise ratio. In the model selection method using Akaike's information criterion (b), there is no such choice of operating point. The probability of false alarm for this particular example is seen to be roughly fixed at 0.1, giving comparable results to likelihood ratio tests (a) at significance level 0.9.

are unclear.

Akaike's method differs from the likelihood ratio test in two ways. Firstly, its derivation relies rather heavily on the assumption of a large number of observations by equating the observed residual with the expected mean residual. This assumption is not obviously warranted when the tests are performed on a minimal set of points. Secondly, it is a parameter free method. This can be an advantage, but it also deprives us of the possibility of trading off the rates of mis-detection and false alarm to select an operating point for the detector which suits the application. The similarity and difference between the two methods is illustrated by the example in Figure 3.1.

Prerequisite: Probability Distribution of the Observations

In order to infer the dimensionality of the structure from the observed images by statistical means, we need a model describing the probability distribution of the observations for any given structure and cameras. Since the true structure and the true cameras are not known

except from these same observations, it is further necessary to estimate the parameters of the cameras and the structure, either implicitly or explicitly. This is done by determining the parameter values that maximize the likelihood of the observations. How to go about doing that is the subject of the next two sections.

3.3 Measurement Models

We will assume that the observed location $\mathbf{w}_p$ of point p in the joint image is related to its true location $\mathbf{x}_p$ by

$$\mathbf{w}_p = \mathbf{x}_p + \mathbf{n}_p, \tag{3.1}$$

where the random additive measurement errors $\mathbf{n}_p$ are assumed to be independent of $\mathbf{x}_p$ with zero mean and covariance Ψ, i.e.

$$\mathrm{E}\{\mathbf{n}\}_p = \mathbf{0}_M, \quad \mathrm{E}\left\{\mathbf{n}_p\mathbf{x}_p^T\right\} = \mathsf{0}_{M\times M}, \quad \mathrm{E}\left\{\mathbf{n}_p\mathbf{n}_p^T\right\} = \Psi. \tag{3.2}$$

The structure of the $M \times M$ error covariance matrix Ψ dictates whether the observation errors of point p in each of the images constituting the joint image are isotropic or anisotropic, whether they vary between the images, and whether they are correlated between images.

In what follows, the error covariance matrix Ψ is *not necessarily assumed to be known*, although it is shown that certain assumptions are needed in order for a hypothesis test to exist. The fact that there is only one Ψ for a set of points under consideration, however, means that all point features are assumed to share the same distribution of errors. The converse case of a known (e.g. anisotropic) error distribution that varies from point to point but remains largely the same between images is thus not encompassed by the current model. For that case, the reader is referred to recent work by Irani and Anandan [IA00] on affine structure recovery from strongly directional or linear image features.

3.3.1 Parametric Model

The model of the joint affine projection of the p-th scene point $\mathbf{X}_p$

$$\mathbf{x}_p = \mathsf{M}_r\mathbf{X}_p + \mathbf{t} \tag{3.3}$$

leads to the measurement model

$$\mathbf{w}_p = \mathsf{M}_r\mathbf{X}_p + \mathbf{t} + \mathbf{n}_p \tag{3.4}$$

where

$\mathbf{w}_p$ is an M-vector of the measured coordinates of the p-th point in the joint image. The vector $\mathbf{w}_p$ could be either the p-th column of the joint measurement matrix W of F views, in which case $M = 2F$, or a point in a given 3D affine structure, in which case $M = 3$.

$\mathbf{X}_p$ is an unobserved non-stochastic r-vector containing the coordinates of the p-th scene point in an r-dimensional space spanned by the measurements. If the measurements are of a general configuration of space points, then $r = 3$, and the $r \times N$ matrix $\mathsf{X} \stackrel{\text{def}}{=} \{\mathbf{X}_p\}$ is the $3 \times N$ structure matrix (2.3). If the structure is degenerate, however, then r will be less than 3. Without loss of generality we will assume that the structure is registered to its centroid, ie. that it has mean zero

$$\sum_{p=1}^{N} \mathbf{X}_p = \mathbf{0}, \tag{3.5}$$

since a non-zero mean can be absorbed by $\mathbf{t}$. By this we effectively choose the centroid of the scene points as the origin of the world coordinate system.

M_r is a fixed unobserved $M \times r$ mapping. If W is the joint measurement matrix, and the point configuration is general, then M_3 is the $2F \times 3$ inhomogeneous joint projection matrix, modulo a 3×3 transformation of space. If $r < 3$ then M_r is accordingly $M \times r$. If $\mathbf{w}_p$ are points in an affine structure then M_r is a $3 \times r$ mapping from the r-dimensional space onto the affine structure.

$\mathbf{t}$ is a fixed M-vector, and

$\mathbf{n}_p$ is the random M-vector of image measurement errors defined above.

A Note on Statistical Methods

The following points may help to place this model in a statistical context:

- A model of the type (3.3), where M_r and $\mathbf{X}_p$ are unobserved non-stochastic variables is known within statistics as a *linear functional relationship* [KS73, And84a]

or a *non-stochastic factor model* [And84b]. With the exception of an excellent tutorial paper by Anderson [And84a], the statistical theory of the linear functional relationship is mostly scattered around the statistical literature [And84b, AR56, KS73, LM63, Tin45, Tin46]. For the related non-central Wishart distribution see [Mui82].

- The above model differs from the better known *linear structural relationship* [KS73, And84a], in which the matrix M_r is stochastic, with the effect that the correlation observed in $\mathbf{w}_p$ is only due to the *factors* $\mathbf{X}_p$ [And84b]. The latter model provides the basis for classical *factor analysis*, a field of statistics which originated in experimental psychology through the work of Spearman [Spe04] and Thurstone [Thu31], prior to being put on a firm statistical foundation by Lawley [Law40, LM71].

- Further assuming that there is no systematic part in the observations leads to the field of *principal component analysis* [Hot33], where the covariance matrix of the observations is the only meaningful entity. This last case gives rise to the well known *Karhunen-Loève expansion* [Kar46, Loè55] often accredited to Pearson [Pea01] who proposed it as a geometric method. Applied to the analysis of covariance, the *Karhunen-Loève* or *Hotelling Transform* entered the pattern recognition literature in 1965 [Wat65].

Factor analysis makes special provision for the case when the noise is large, and hence is absorbed into the principal components. For small noise variance we expect principal component analysis and factor analysis to give similar results [MKB79, p.276]. The number of parameters in the three models differs greatly, as do the circumstances under which optimal methods of statistical inference exist.

Ambiguity of the Parametric Model

In the parametric model (3.3) *both* M_r and $\mathbf{X}_p$ are unknown. Consequently there is an intrinsic ambiguity in the model, apparent from

$$\mathsf{M}_r\mathbf{X}_p = \mathsf{M}_r(\mathsf{A}\mathsf{A}^{-1})\mathbf{X}_p = (\mathsf{M}_r\mathsf{A})(\mathsf{A}^{-1}\mathbf{X}_p) = \mathsf{M}'_r\mathbf{X}'_p, \tag{3.6}$$

where A is any $r \times r$ nonsingular matrix. This is the cause of the well known affine indeterminacy of the world frame in 3D reconstruction from uncalibrated affine views. In classical factor analysis [And84b], this ambiguity is eliminated by imposing two constraints

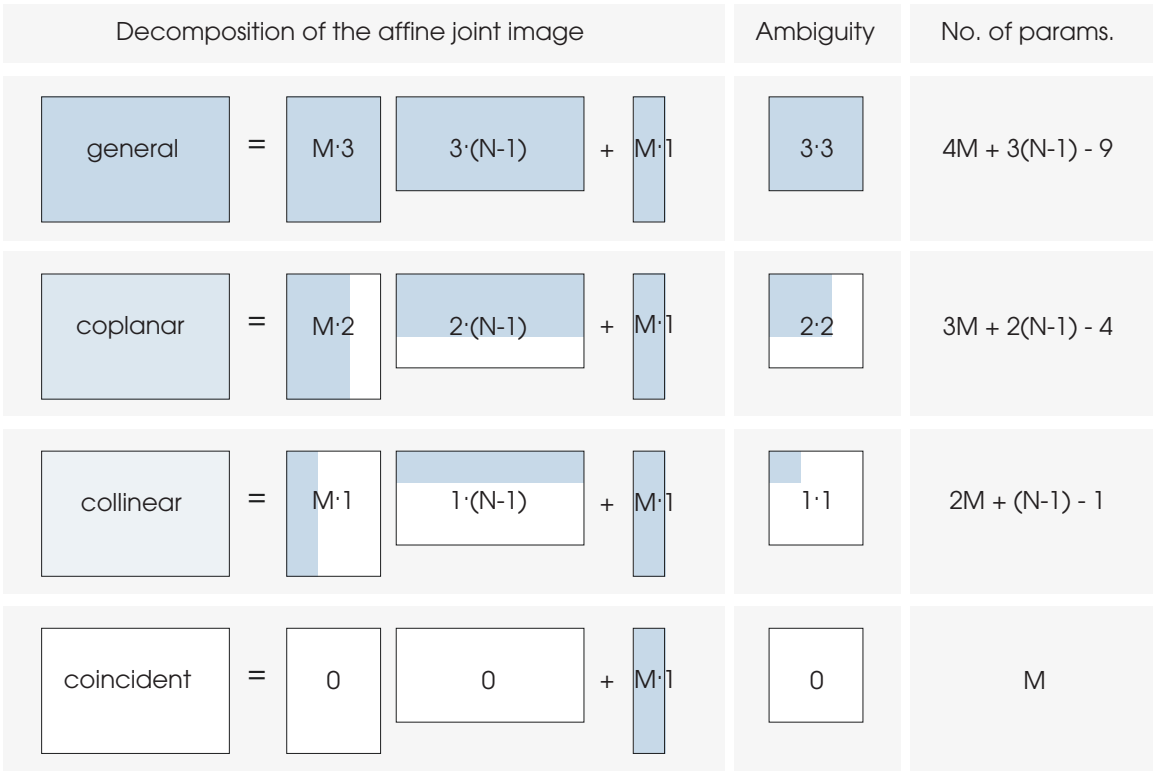


Table 3.1: The number of independent parameters in the parametric model (3.3) of the affine joint image for different rank of the depicted structure, taking into account the inherent ambiguity (3.6) in the model.

on the factors $\mathbf{X}_p$, with the understanding that the factors of the model are related to the true factors by an unknown linear transformation A. The constraints are arbitrary, but chosen such as to simplify the analysis of the model. The first constraint

$$\sum_{p=1}^{N} \mathbf{X}_p \mathbf{X}_p^T = N \cdot \mathsf{I}_{3\times 3}, \tag{3.7}$$

imposes *orthonormality* onto the factors. The remaining rotational ambiguity is commonly resolved by requiring that the matrix

$$\mathsf{M}_r^T \Psi^{-1} \mathsf{M}_r \tag{3.8}$$

be diagonal, with ordered and distinct diagonal elements.

Number of Independent Parameters in the Model of the Joint Image

The number of independent model parameters in the affine joint image (3.3) of N scene points is given in Table 3.1. The joint image of a non-degenerate (rank $r = 3$) or a degen-

Decomposition of 3D affine structure				Ambiguity	No. of params.
general	=	3·3	3·(N-1)	3·3	9 + 3(N-1) - 9
coplanar	=	3·2	2·(N-1)	2·2	6 + 2(N-1) - 4
collinear	=	3·1	1·(N-1)	1·1	3 + (N-1) - 1
coincident	=	0	0	0	0

Table 3.2: A degenerate $3 \times N$ structure matrix can be decomposed in the same way as a degenerate measurement matrix (compare Table 3.1). It is assumed that the structure is registered to its centroid as in (3.5).

erate (rank $r < 3$) point structure can be decomposed into the product of two matrices MX, with the dimensions $M \times r$ and $r \times N$ respectively. The number of independent parameters in each of these models equals the number of parameters in M and X, less the parameters in the $r \times r$ matrix A, due to the ambiguity (3.6). The number of independent columns of X is $N - 1$, rather than N, because of the registration to the centroid (3.5).

A degenerate $3 \times N$ structure matrix can be decomposed in the same way as a degenerate measurement matrix, as shown in Table 3.2. It is assumed that the structure is registered to its centroid as in (3.5).

3.3.2 Normal Form: Linear Restrictions

There exists a dual form of the parametric model, which is also of interest. The parametric model (3.3) describes an M dimensional measurement space in which the true part $\mathbf{x}_p$ of

the measurement varies in a linear subspace of dimension r. By using the *normal form* [Nom66] of the hyperplane, the variables $\mathbf{x}_p$ can be equivalently represented by the dual model [And84a]

$$\mathsf{B}_q \mathbf{x}_p = \mathbf{b}, \tag{3.9}$$

where the $q \times M$ matrix B_q has rank

$$q \stackrel{\text{def}}{=} M - r. \tag{3.10}$$

Each row of B_q and $\mathbf{b}$ then defines a hyperplane in the measurement space. The intersection of the q hyperplanes cuts out the r dimensional linear subspace of $\mathbf{x}_p$ in the measurement space. It is noted [And84a] that the matrix B_q can be left multiplied by an arbitrary $q \times q$ matrix without affecting the constraint (3.9) imposed by the model. The two models (3.3) and (3.9) are connected by the equations [And84a]

$$\mathsf{B}_q \mathsf{M}_r = \mathsf{0}_{q \times r}, \tag{3.11}$$

$$\mathsf{B}_q \mathbf{t} = \mathbf{b}. \tag{3.12}$$

The maximum likelihood estimators of the parameters of the model (3.9) were derived by Tintner [Tin45] (see Section 3.4.2). This model will play a role in the estimation of the multifocal affine tensors, presented in Chapter 7.

3.4 Optimal Estimation of Affine Structure

This section provides optimal statistical estimators of the parameters of the measurement models given in the previous section, assuming normal measurement errors. The resulting estimators of affine structure and motion provide an estimate of the likelihood of the observations, which is required for the hypothesis tests derived in Section 3.5.

Criteria of Optimality for Batch and Recursive Estimation

By the *optimal* estimator of the affine structure (and motion) we mean the *MAP* estimator [BSF88], i.e. the set of model parameters $\widehat{\theta} = \{\widehat{\Psi}, \widehat{\mathsf{M}}_r, \widehat{\mathsf{X}}_r, \widehat{\mathbf{t}}\}$ maximizing the a-posteriori probability density function $p(\theta|\mathsf{W})$, which by *Bayes' rule* can be written as

$$p(\theta|\mathsf{W}) = \frac{p(\mathsf{W}|\theta)p(\theta)}{p(\mathsf{W})}. \tag{3.13}$$

The measurement model (3.4) provides the conditional probability density function $p(\mathsf{W}|\theta)$. In order to maximize the right hand side of (3.13), one needs in addition the a-priori p.d.f. $p(\theta)$, which is not always known. If $p(\theta)$ is unknown, θ can be taken as uniformly distributed, with $p(\theta)$ constant. In this case the problem reduces to maximizing the conditional p.d.f., also known as the *likelihood function*

$$L_\theta(\mathsf{W}) \stackrel{\text{def}}{=} p(\mathsf{W}|\theta). \tag{3.14}$$

For the problem of estimating structure from multiple views, one can distinguish between two approaches, *batch* estimation and *recursive* estimation. In the former approach, the data from all views are combined to produce an estimate of the structure in a single step. In the absence of an a-priori estimate from an *external* source, the optimal estimate is thus provided by the *maximum likelihood* (or *ML*) estimator. In recursive estimation, views are introduced sequentially, and a running structure estimate is updated in each step. Hence an a-priori estimate is available from the previous step, and the optimal update is provided by the MAP estimator. An exception is the first step of the recursion, in which an internal a-priori estimate is not available and the optimal estimate is provided by the ML estimator.

The Likelihood Function

In order to define the likelihood function (3.14), required for the batch estimation of the structure and camera parameters, we need to complete the specification of the the probability distribution of the measurement errors given in (3.2). Adding the assumption of normal errors $\mathsf{n}_p \sim N(\mathsf{0}, \Psi)$, the likelihood function of the parametric model can be written down as

$$L = (2\pi)^{-\frac{1}{2}MN} \, |\Psi|^{-\frac{1}{2}N} \exp\left\{ -\tfrac{1}{2} \sum_{p=1}^{N} (\mathbf{w}_p - \mathsf{M}_r \mathbf{X}_{r,p} - \mathbf{t})^T \Psi^{-1} (\mathbf{w}_p - \mathsf{M}_r \mathbf{X}_{r,p} - \mathbf{t}) \right\}. \tag{3.15}$$

The maximum likelihood estimators are the values of the parameters Ψ, M_r, $\mathbf{X}_{r,p}$, and $\mathbf{t}$ that maximize the likelihood function (3.15), for a given rank r. Starting with $\mathbf{t}$, we have

$$\frac{\partial \log L}{\partial \mathbf{t}} = -\sum_{p=1}^{N} \Psi^{-1} \left(\mathbf{w}_p - \mathsf{M}_r \mathbf{X}_{r,p} - \mathbf{t} \right) = \Psi^{-1} \left(N\mathbf{t} - \sum_{p=1}^{N} \mathbf{w}_p \right) \stackrel{!}{=} 0,$$

since the structure $\mathbf{X}_{r,p}$ is defined to have zero mean, giving the maximum likelihood estimator

$$\widehat{\mathbf{t}} = \tfrac{1}{N} \sum_{p=1}^{N} \mathbf{w}_p \stackrel{\text{def}}{=} \overline{\mathbf{w}}. \tag{3.16}$$

Replacing $\mathbf{t}$ by its ML estimator (3.16) we can write (3.15) more simply as

$$L = (2\pi)^{-\frac{1}{2}MN} \, |\Psi|^{-\frac{1}{2}N} \exp \operatorname{tr} \left\{ -\tfrac{1}{2} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r)^T \Psi^{-1} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r) \right\}, \tag{3.17}$$

where

$$\Delta \mathsf{W} \stackrel{\text{def}}{=} \{ \mathbf{w}_p - \overline{\mathbf{w}} \} \tag{3.18}$$

is the *registered measurement matrix* known from [TK92].

3.4.1 Error Structure and Non-Existence of ML Estimators

The existence of a maximum value of the function (3.17), and hence the existence of maximum likelihood estimators for the remaining parameters of the model, Ψ, M_r, and $\mathbf{X}_{r,p}$, depends critically on the assumptions made about the error covariance Ψ defined in (3.2). Anderson [And84a] considers the following four different assumptions:

1. Ψ completely known,
2. Ψ known up to scale,
3. Ψ an unknown diagonal matrix, implying uncorrelated errors, and
4. Ψ unknown unrestricted.

The three restricted forms of the covariance matrix (1-3 above), frequently crop up in practical situations, e.g.

1. when the accuracy of the feature detector is either known or can be measured,
2. when the the error variance can be assumed isotropic and equal in all images, but its scale is unknown,
3. when the images stem from different sources, each with its own (unknown) error variance.

We will now discuss each of the assumptions (1-4) in turn, starting with the last one, and give the maximum likelihood estimators where they exist in a practical setting.

Unknown Unrestricted Error Covariance

If the error covariance is unknown and unrestricted, replicated observations are needed in order to estimate Ψ. Otherwise the parameters are not identified [And84a]. This requires a special kind of imagery: sets of repeated images, such that within each set there is zero motion between the camera and the scene. We will not consider this case further here, but remark that the maximum likelihood estimators for the parameters of the linear restriction model are given in [And84a], provided that replicated observations are available.

Unknown Uncorrelated Measurement Errors

This is the case of an unknown diagonal covariance matrix Ψ, corresponding to the non-stochastic factor model [And84a]. In contrast to the widely used stochastic factor model [LM71], it has been shown [AR56, And84b] that the likelihood function (3.17) for the non-stochastic model does not have an upper bound, and it follows that maximum likelihood estimators do not exist. To see this, we rewrite (3.17) as

$$L = \frac{1}{\left[(2\pi)^M \prod_{i=1}^M \psi_{ii}\right]^{N/2}} \prod_{i=1}^{M} \exp\left\{ -\tfrac{1}{2} \sum_{\alpha=1}^{N} \frac{(w_{i\alpha} - \bar{w}_i - \sum_{j=1}^{r} m_{ij} X_{j\alpha})^2}{\psi_{ii}} \right\}$$

and note that we are free to choose any one row of M_r to contain only one non-zero element. Let this element be $m_{i'j'} = 1$, then we can estimate the j'-th row of the structure matrix by the i'-th row of the centred measurement matrix, resulting in a perfect fit to the i'-th row of measurements. Accompanied with a zero residual, the exponent containing the unknown variance $\psi_{i'i'}$ disappears from the likelihood function, and $\psi_{i'i'}$ remains only in the denominator of the normalizing constant. Hence the likelihood function will increase without bound as $\psi_{i'i'} \to 0$.

3.4.2 MLE for Error Covariance Known up to Scale

If the error variance is known up to scale, i.e. if $\Psi = \sigma^2 \Psi_0$ with σ unknown, the maximum likelihood estimators of the structure (and motion) exist. The ML estimators were originally derived by Tintner [Tin45] in terms of the restriction model (3.9). Here we give an alternative derivation in terms of the parametric model.

The likelihood function, with $\mathbf{t}$ replaced by its ML estimator (3.17), becomes

$$L = (2\pi)^{-\frac{1}{2}MN} \left|\sigma^2 \Psi_0\right|^{-\frac{1}{2}N} \exp \operatorname{tr} \left\{ -\tfrac{1}{2} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r)^T (\sigma^2 \Psi_0)^{-1} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r) \right\}. \tag{3.19}$$

The ML estimators are the values of σ, M_r, and X_r that maximize (3.19) for a given rank r. Reasoning along the lines of [AR56, And84b], we find[1] that the ML estimators only exist for $r < \operatorname{rank} \Delta \mathsf{W}$. This is not a problem in practice, since the presence of noise increases the rank of $\Delta \mathsf{W}$ above that of the underlying structure. By the definition of the Frobenius norm

$$\left\| \mathsf{A} \right\|_F \stackrel{\text{def}}{=} \sqrt{\operatorname{tr} \mathsf{A}^\mathsf{T} \mathsf{A}}$$

we can replace the trace in (3.19) by the squared Frobenius norm of the matrix $\Psi^{-\frac{1}{2}}(\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r)$, and take the logarithm to obtain the criterion

$$\log L = -\tfrac{1}{2} MN \log(2\pi) - \tfrac{1}{2} N \log |\Psi_0| - \tfrac{1}{2} MN \log \sigma^2 - \tfrac{1}{2\sigma^2} \left\| \Psi_0^{-\frac{1}{2}} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r) \right\|_F^2 \stackrel{!}{=} \max.$$

The first two terms are constant, so the criterion can be reduced to

$$MN \log \sigma^2 + \tfrac{1}{\sigma^2} \left\| \Psi_0^{-\frac{1}{2}} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r) \right\|_F^2 \stackrel{!}{=} \min. \tag{3.20}$$

Differentiating with respect to σ^2, we find that for a given value $F > 0$ of the Frobenius norm

$$F \stackrel{\text{def}}{=} \left\| \Psi_0^{-\frac{1}{2}} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r) \right\|_F$$

the criterion (3.20) is minimized when

$$\sigma^2 = \frac{F^2}{MN}. \tag{3.21}$$

Substituting (3.21) back into (3.20) we obtain

$$MN \log F^2 - MN \log MN + MN \stackrel{!}{=} \min,$$

which is satisfied at the minimum Frobenius norm

$$F^2 = \left\| \Psi_0^{-\frac{1}{2}} (\Delta \mathsf{W} - \mathsf{M}_r \mathsf{X}_r) \right\|_F^2 \stackrel{!}{=} \min. \tag{3.22}$$

[1] For $r \geq \operatorname{rank} \Delta \mathsf{W}$ and unknown σ^2 the likelihood function does not have an upper bound. This is because the exponent can be made to vanish altogether, by choosing $\mathsf{M}_r \mathsf{X}_r = \Delta \mathsf{W}$, leaving σ^2 only in the denominator of the normalizing constant. Hence $L \to \infty$ as $\sigma^2 \to 0$.

Hence the criterion (3.20) is equivalent to the two criteria (3.21) and (3.22). Substituting the Mahalanobis[2] transformed variables

$$\Delta \mathsf{W}' \stackrel{\text{def}}{=} \Psi_0^{-\frac{1}{2}} \Delta \mathsf{W} \quad \text{and} \quad \mathsf{M}'_r \stackrel{\text{def}}{=} \Psi_0^{-\frac{1}{2}} \mathsf{M}_r \tag{3.23}$$

into (3.22) we obtain the specific form

$$F^2 = \left\| \Delta \mathsf{W}' - \mathsf{M}'_r \mathsf{X}_r \right\|_F^2 \stackrel{!}{=} \min . \tag{3.24}$$

This equation has a known solution [HY38] (see also [RM96]). For a given rank r of the matrix $\mathsf{M}'_r \mathsf{X}_r$ the norm (3.24) is minimized by $\mathsf{M}'_r \mathsf{X}_r = \mathsf{U}'_r \mathsf{D}'_r \mathsf{V}'^T_r$, where the matrices U'_r, D'_r, and V'_r contain the r largest singular values and the corresponding vectors of the singular value decomposition $\Delta \mathsf{W}' = \mathsf{U}' \mathsf{D}' \mathsf{V}'^T$. The minimum value of the Frobenius norm (3.24) is given by

$$\min F^2 = \operatorname{tr} \mathsf{D}'^2 - \operatorname{tr} \mathsf{D}'^2_r = \sum_{i=r+1}^{M} d'^2_i . \tag{3.25}$$

It follows that the maximum likelihood estimators of M_r and X_r, when the error covariance is known up to scale, are any two matrices $\widehat{\mathsf{M}}_r$ and $\widehat{\mathsf{X}}_r$ satisfying

$$\Psi_0^{-\frac{1}{2}} \widehat{\mathsf{M}}_r \widehat{\mathsf{X}}_r = \mathsf{U}'_r \mathsf{D}'_r \mathsf{V}'^T_r . \tag{3.26}$$

Applying the particular disambiguating constraints (3.7)-(3.8), we obtain the estimators

$$\boxed{\begin{aligned} \widehat{\mathsf{M}}_r &= N^{-\frac{1}{2}} \cdot \Psi_0^{\frac{1}{2}} \mathsf{U}'_r \mathsf{D}'_r , \\ \widehat{\mathsf{X}}_r &= N^{\frac{1}{2}} \cdot \mathsf{V}'^T_r . \end{aligned}} \tag{3.27}$$

To see this, we substitute the above estimators into (3.7)-(3.8) to obtain

$$\widehat{\mathsf{X}}_r \widehat{\mathsf{X}}_r^T = N \cdot \mathsf{V}'^T \mathsf{V}' = N \cdot \mathsf{I}_{3\times 3}$$

in compliance with the first constraint, and

$$\widehat{\mathsf{M}}_r^T (\widehat{\sigma}^2 \Psi_0)^{-1} \widehat{\mathsf{M}}_r = \frac{1}{N\widehat{\sigma}^2} \cdot \mathsf{D}'^T_r \mathsf{U}'^T_r \mathsf{U}'_r \mathsf{D}'_r = \frac{1}{N\widehat{\sigma}^2} \cdot \mathsf{D}'^2_r , \tag{3.28}$$

[2]Strictly speaking the *Mahalanobis transformation* [MKB79] standardizes a multi-normal variable to zero mean and unit variance. As defined in (3.23) the transformed measurement matrix $\Delta \mathsf{W}'$ has zero mean and unit variance up to unknown scale. The term is applied loosely to the transformed projection matrix M'_r.

which is diagonal by the definition of D'_r and the SVD, and generally has distinct diagonal elements.

In the case of unknown variance σ^2, we obtain the maximum likelihood estimator $\widehat{\sigma}^2$ by substituting the the minimum (3.25) into (3.21) yielding

$$\boxed{\widehat{\sigma}^2 = \frac{\min F^2}{MN} = \frac{1}{MN}\sum_{i=r+1}^{M} {d'}_i^2.} \tag{3.29}$$

Error Covariance of the ML Structure Estimate

Having obtained the ML estimators for the motion and structure, we are in a position to give the error covariance of the estimated structure. From

$$\Sigma^{-1}_{\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i} = \widehat{\mathsf{M}}_r^T \Sigma^{-1}_{\mathbf{w}_i\mathbf{w}_i} \widehat{\mathsf{M}}_r = \widehat{\mathsf{M}}_r^T \Psi^{-1} \widehat{\mathsf{M}}_r$$

we obtain by (3.28)

$$\boxed{\Sigma_{\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i} = N\widehat{\sigma}^2 \cdot {\mathsf{D}'}_r^{-2}.} \tag{3.30}$$

This particular expression depends on the choice of the camera matrix $\widehat{\mathsf{M}}_r$ in the product $\widehat{\mathsf{M}}_r\widehat{\mathsf{X}}_r$ as effected by the disambiguating constraints (3.7)-(3.8).

Isotropic Factorization

A different factorization of the ML estimate $\widehat{\mathsf{M}}_r\widehat{\mathsf{X}}_r$ is given by

$$\boxed{\begin{aligned} \widehat{\mathsf{M}}_r &\stackrel{\text{def}}{=} \Psi_0^{\frac{1}{2}} \mathsf{U}'_r \\ \widehat{\mathsf{X}}_r &\stackrel{\text{def}}{=} {\mathsf{D}'}_r {\mathsf{V}'}_r^T. \end{aligned}} \tag{3.31}$$

This particular choice leads to *isotropic errors* in the structure $\widehat{\mathsf{X}}_r$, with

$$\boxed{\Sigma_{\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i} = \widehat{\sigma}^2 \cdot \mathsf{I}_r.} \tag{3.32}$$

This form is particularly useful when fitting a model to the recovered structure, since the Gaussian maximum likelihood error function becomes equivalent to the sum of squared distances.

Factorization	$\widehat{\mathsf{M}}_r$	$\widehat{\mathsf{X}}_r$	$\Sigma_{\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i}$
Classical Factor Analysis (3.27)	$\frac{1}{\sqrt{N}}\Psi_0^{\frac{1}{2}}\mathsf{U}'_r\mathsf{D}'_r$	$\sqrt{N}\mathsf{V}'^T_r$	$N\widehat{\sigma}^2\mathsf{D}'^{-2}_r$
Isotropic (3.31)	$\Psi_0^{\frac{1}{2}}\mathsf{U}'_r$	$\mathsf{D}'_r\mathsf{V}'^T_r$	$\widehat{\sigma}^2\mathsf{I}_r$
Tomasi and Kanade [TK92]	$\mathsf{U}_3\mathsf{D}_3^{\frac{1}{2}}$	$\mathsf{D}_3^{\frac{1}{2}}\mathsf{V}_3^T$	$\widehat{\sigma}^2\mathsf{D}_3^{-1}$

Table 3.3: A comparison of the different choices of factorization of the ML estimate $\widehat{\mathsf{M}}_r\widehat{\mathsf{X}}_r$, and the effect on the error covariance matrix of the resulting structure $\widehat{\mathsf{X}}_r$.

Tomasi and Kanade Factorization

A third form of the factorization was used by Tomasi and Kanade in their original factorization paper [TK92]. They assumed isotropic image errors, which in the current notation means $\Psi = \sigma^2\mathsf{I}_M$ or $\Psi_0 = \mathsf{I}_M$ implying $\Delta\mathsf{W}' = \Delta\mathsf{W}$. Consequently (3.26) can be written in terms the singular value decomposition of $\Delta\mathsf{W}$ as

$$\widehat{\mathsf{M}}_r\widehat{\mathsf{X}}_r = \mathsf{U}_r\mathsf{D}_r\mathsf{V}_r^T,$$

or in the case of rank $r = 3$

$$\widehat{\mathsf{M}}_3\widehat{\mathsf{X}}_3 = \mathsf{U}_3\mathsf{D}_3\mathsf{V}_3^T,$$

which is the expression obtained by Tomasi and Kanade. It follows that the maximum likelihood estimators $\widehat{\mathsf{M}}_3$ and $\widehat{\mathsf{X}}_3$ derived here are up to an affine transformation equal to the *rotation* and *shape* matrices obtained by Tomasi and Kanade when the errors in the images are isotropic normal, and of equal size in all images. Consequently, their original factorization method is optimal in the maximum likelihood sense under the same conditions. This is to be expected, for it has been shown since [RM96] that Tomasi and Kanade's method minimizes SSD reprojection error.

Comparison to (3.27) and (3.30) indicates that Tomasi and Kanade's factorization

$$\boxed{\begin{aligned}\widehat{\mathsf{M}}_3 &\stackrel{\text{def}}{=} \mathsf{U}_3\mathsf{D}_3^{\frac{1}{2}}\\ \widehat{\mathsf{X}}_3 &\stackrel{\text{def}}{=} \mathsf{D}_3^{\frac{1}{2}}\mathsf{V}_3^T\end{aligned}} \tag{3.33}$$

has the associated error covariance

$$\boxed{\Sigma_{\widehat{\mathbf{x}}_i\widehat{\mathbf{x}}_i} = \widehat{\sigma}^2 \cdot \mathsf{D}_3^{-1}.} \tag{3.34}$$

For comparison between the three choices of factorization and the associated error covariances see Table 3.3.

MLE for Known Error Covariance

The case of known error variance $\Psi = \Psi_0$ equals the case of variance known up to scale $\Psi = \sigma^2\Psi_0$ with the additional information $\sigma^2 = 1$. Substituting $\sigma^2 = 1$ into the criterion (3.20) yields the criterion (3.22) so the maximum likelihood estimators $\widehat{\mathsf{M}}_r$ and $\widehat{\mathsf{X}}_r$ are the same as in the case of variance known up to scale.

3.4.3 Optimal Recursive Estimation

This section applies the optimality criteria laid down at the beginning of Section 3.4 to obtain an optimal recursive estimator for affine structure, effectively providing a common framework for both batch and recursive structure estimation. This section also provides the error covariance of the recovered structure as is required for the current task.

A couple of very similar recursive estimators were derived by McLauchlan et al. [MRM94, MM95] within the general framework of the *Variable State Dimension Filter* (VSDF). The VSDF encompasses recursive estimation of structure in the form of points, lines, and surfaces, coupled with the estimation of the motion and the calibration of the camera under various imaging conditions [MM96, MRM94, MM95, McL99]. Due to its broad scope and the emphasis on fast real-time implementation the VSDF can appear complex and obscure. It is hoped that the present derivation exposes the extreme simplicity of the optimal recursive structure estimator, in terms of implementation as well as computational complexity, in the simple case of recovering affine point structure from affine views.

Recursive MAP Estimation

Let us assume that at each time step we are supplied with one or more additional rows to the measurement matrix, as depicted in Figure 3.2. Given the most recent estimate of the structure $\widehat{\mathsf{X}}^0$, and the new measurements W, we seek an updated estimate of the structure X along with an estimate of the additional rows to the projection matrix M and the translation t corresponding to the new measurements. Under the assumption of normal measurement errors and of normal errors in the structure estimate, we maximize the a posteriori proba-

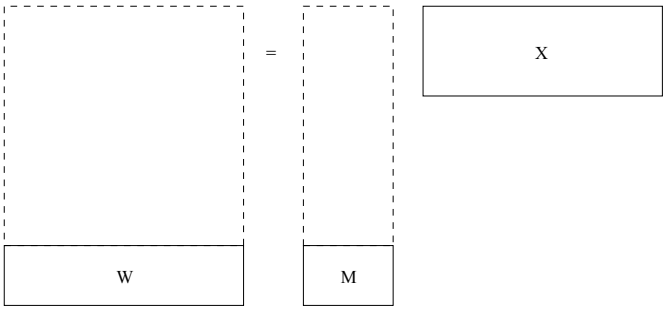


Figure 3.2: Recursive estimation of affine structure. Given a prior structure estimate $\widehat{\mathsf{X}}^0$ and additional rows $\Delta\mathsf{W}$ to the registered measurement matrix, we want to obtain an updated structure estimate $\widehat{\mathsf{X}}$, as well as estimating the current projection matrix M.

bility $p(\mathsf{M}, \mathbf{t}, \mathsf{X}|\mathsf{W}, \widehat{\mathsf{X}}^0)$ to obtain the *maximum a posteriori* (MAP) estimator (see [BSF88, sec. 3.5]).

Three cases will be considered. The first and simplest case is when all the measured image features can be assumed to have identical error covariance and every feature is visible in all views. This is the assumption of the optimal batch method of the previous section. For extended sequences, for which recursive methods are eminently suitable because of their low computational complexity, these assumptions may not be realistic. Two modifications to the basic method will therefore be given, one to allow the disappearance and reappearance of features, and the second to allow the measured image features to have individually different error covariance.

Uniform Accuracy of Features and Structure

Assuming that all measurements at the current time step share a common error covariance matrix $\Psi = \Psi_0$, we again obtain the estimator (3.16) for $\mathbf{t}$. To estimate the remaining parameters M and X we are left with maximizing the exponent

$$g(\mathsf{M},\mathsf{X}) \stackrel{\text{def}}{=} \operatorname{tr}\left\{-\tfrac{1}{2}(\Delta\mathsf{W}-\mathsf{M}\mathsf{X})^T\Psi_0^{-1}(\Delta\mathsf{W}-\mathsf{M}\mathsf{X}) - \tfrac{1}{2}(\widehat{\mathsf{X}}^0-\mathsf{X})^T\Sigma^{-1}_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}(\widehat{\mathsf{X}}^0-\mathsf{X})\right\} \stackrel{!}{=} \max, \tag{3.35}$$

assuming a common covariance matrix $\Sigma_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}$ of the errors $\tilde{\mathsf{X}}^0 \stackrel{\text{def}}{=} \mathsf{X} - \widehat{\mathsf{X}}^0$ in the prior structure estimate. In (3.35) the covariance Ψ_0 of the measurement errors is assumed fixed throughout the sequence, whereas we expect the error covariance $\Sigma_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}$ of the estimated structure to become smaller as more data becomes available.

Starting with M, we obtain

$$\frac{\partial g}{\partial \mathsf{M}} = \Psi_0^{-1}\Delta\mathsf{W}\mathsf{X}^T - \Psi_0^{-1}\mathsf{M}\mathsf{X}\mathsf{X}^T \stackrel{!}{=} 0_{2\times 3},$$

from which we can write the MAP estimator $\widehat{\mathsf{M}}$ as a function of the estimator $\widehat{\mathsf{X}}$, as follows

$$\boxed{\widehat{\mathsf{M}} = \mathcal{M}(\widehat{\mathsf{X}}) \stackrel{\text{def}}{=} \Delta\mathsf{W}\widehat{\mathsf{X}}^T(\widehat{\mathsf{X}}\widehat{\mathsf{X}}^T)^{-1} = \Delta\mathsf{W}\widehat{\mathsf{X}}^\dagger,} \tag{3.36}$$

where $\widehat{\mathsf{X}}^\dagger$ denotes the pseudo-inverse of $\widehat{\mathsf{X}}$. Substituting $\mathcal{M}(\mathsf{X})$ for M in (3.35), we find that the resulting objective function $g(\mathcal{M}(\mathsf{X}), \mathsf{X})$ contains terms higher than second order in X. Thus we are left with a non-linear optimization problem, the solution to which is the MAP estimate $\mathsf{X} = \widehat{\mathsf{X}}$, that maximizes $g(\mathcal{M}(\mathsf{X}), \mathsf{X})$. We could solve this problem using any suitable numerical non-linear optimization method. This implies an iterative algorithm at each time step.

A Decoupled Recursive MAP Estimator

Assuming we have a reasonably accurate prior estimate of the structure $\widehat{\mathsf{X}}^0$ from previous time steps, but no prior knowledge of M which is considered free to change between time steps, a simpler optimization scheme is obtained by alternating between estimating the projection M by $\mathcal{M}(\widehat{\mathsf{X}}^i)$ in (3.36) and estimating the structure X by solving the *linear* problem

$$g(\mathcal{M}(\widehat{\mathsf{X}}^i), \mathsf{X}) \stackrel{!}{=} \max.$$

Setting the derivative w.r.t. X equal to zero

$$\frac{\partial g(\mathcal{M}(\widehat{\mathsf{X}}^i), \mathsf{X})}{\partial \mathsf{X}} = \mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}(\Delta\mathsf{W} - \mathcal{M}(\widehat{\mathsf{X}}^i)\mathsf{X}) + \Sigma^{-1}_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}(\widehat{\mathsf{X}}^0 - \mathsf{X}) \stackrel{!}{=} 0$$

gives the $(i+1)$-st structure estimate

$$\widehat{\mathsf{X}}^{i+1} = \left(\mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}\mathcal{M}(\widehat{\mathsf{X}}^i) + \Sigma^{-1}_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}\right)^{-1}\left(\mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}\Delta\mathsf{W} + \Sigma^{-1}_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}\widehat{\mathsf{X}}^0\right).$$

This equation can be rewritten as [BSF88, p. 34]

$$\boxed{\widehat{\mathsf{X}}^{i+1} = \widehat{\mathsf{X}}^i + \Sigma_{\tilde{\mathbf{X}}^i\tilde{\mathbf{X}}^i}\mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}\left(\Delta\mathsf{W} - \mathcal{M}(\widehat{\mathsf{X}}^i)\widehat{\mathsf{X}}^i\right),} \tag{3.37}$$

where

$$\begin{aligned}\Sigma_{\tilde{\mathbf{X}}^i\tilde{\mathbf{X}}^i} &\stackrel{\text{def}}{=} \left(\mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}\mathcal{M}(\widehat{\mathsf{X}}^i) + \Sigma^{-1}_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}\right)^{-1} \\ &= \Sigma_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0} - \Sigma_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}\mathcal{M}(\widehat{\mathsf{X}}^i)^T\left(\mathcal{M}(\widehat{\mathsf{X}}^i)\Sigma_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}\mathcal{M}(\widehat{\mathsf{X}}^i)^T + \Psi_0\right)^{-1}\mathcal{M}(\widehat{\mathsf{X}}^i)\Sigma_{\tilde{\mathbf{X}}^0\tilde{\mathbf{X}}^0}\end{aligned} \tag{3.38}$$

is the covariance matrix of the estimation error $\tilde{\mathsf{X}}^i \stackrel{\text{def}}{=} \mathsf{X} - \widehat{\mathsf{X}}^i$ at the end of the i-th iteration. This *decoupled recursive MAP estimator* of structure and motion is still iterative within the time step, but provides a closed form for both the structure and the motion at each step of the iteration.

We note that the structure update (3.37) is performed in a time proportional to N, the number of points in the structure, independent of the number of frames processed so far. The covariance update (3.38) is performed in constant time, involving only the inversion of an $r \times r$ matrix (top) or an $m \times m$ matrix (bottom), whichever is smaller. The time complexity at each time step is therefore $O(N)$, so that a complete recursive structure estimate from F frames has time complexity $O(FN)$, compared to $O\left((F+N)^3\right)$ using the SVD.

New and Missing Correspondences

If features cannot be tracked throughout the whole sequence, but rather appear and disappear at different time steps, the individual points in the estimated structure may have a different error covariance $\Sigma_{\tilde{\mathbf{X}}^0_p\tilde{\mathbf{X}}^0_p}$. Hence (3.35) becomes

$$\begin{aligned}g(\mathsf{M},\mathsf{X}) \stackrel{\text{def}}{=} &-\tfrac{1}{2}\operatorname{tr}\left\{-\tfrac{1}{2}(\Delta\mathsf{W}-\mathsf{M}\mathsf{X})^T\Psi_0^{-1}(\Delta\mathsf{W}-\mathsf{M}\mathsf{X})\right\} \\ &-\tfrac{1}{2}\sum_{p=1}^{N}(\widehat{\mathbf{X}}^0_p-\mathbf{X}_p)^T\Sigma^{-1}_{\tilde{\mathbf{X}}^0_p\tilde{\mathbf{X}}^0_p}(\widehat{\mathbf{X}}^0_p-\mathbf{X}_p) \stackrel{!}{=} \max.\end{aligned} \tag{3.39}$$

We note that the first part of (3.39) is unchanged, so the MAP estimator $\widehat{\mathsf{M}}$ is still given by (3.36). The structure update must now be performed point-wise. Differentiating (3.39) w.r.t. a single structure point, we obtain

$$\frac{\partial g(\mathcal{M}(\widehat{\mathsf{X}}^i),\mathsf{X})}{\partial\mathbf{X}_\mathsf{p}} = \mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}(\Delta\mathbf{w}_p - \mathcal{M}(\widehat{\mathsf{X}}^i)\mathbf{X}_p) + \Sigma^{-1}_{\tilde{\mathbf{X}}^0_p\tilde{\mathbf{X}}^0_p}(\widehat{\mathbf{X}}^0_p-\mathbf{X}_p) \stackrel{!}{=} 0.$$

Hence we replace (3.37) by the pointwise update equation

$$\widehat{\mathbf{X}}^{i+1}_p = \widehat{\mathbf{X}}^i_p + \Sigma_{\tilde{\mathbf{X}}^i_p\tilde{\mathbf{X}}^i_p}\mathcal{M}(\widehat{\mathsf{X}}^i)^T\Psi_0^{-1}\left(\Delta\mathbf{w}_p - \mathcal{M}(\widehat{\mathsf{X}}^i)\widehat{\mathbf{X}}^i_p\right). \tag{3.40}$$

The structure covariance $\Sigma_{\tilde{\mathbf{X}}_p^i \tilde{\mathbf{X}}_p^i}$ is maintained separately for each point, and updated by applying (3.38) to the covariance matrices of currently observed correspondences only.

Variable Feature Accuracy within a Time Step

If point features are obtained by a corner detector, it may be reasonable to assign the same error covariance Ψ_0 to every feature. This may not always be appropriate, however. When, for example, point features are obtained by intersecting lines, each feature can have a different error covariance $\Psi_{0,p}$, depending on the accuracy of the particular lines and their angle of incidence. In this case the MAP estimate of the translation becomes

$$\widehat{\mathbf{t}} = \left(\sum_{p=1}^{N} \Psi_{0,p}^{-1} \right)^{-1} \left(\sum_{p=1}^{N} \Psi_{0,p}^{-1} \mathbf{w}_p \right).$$

Assuming, as above, that features are not necessarily tracked throughout the sequence, (3.35) takes the most general form

$$\begin{aligned} g(\mathsf{M}, \mathsf{X}) \stackrel{\text{def}}{=} & -\tfrac{1}{2} \sum_{p=1}^{N} (\Delta \mathbf{w}_p - \mathsf{M}\mathbf{X}_p)^T \Psi_{0,p}^{-1} (\Delta \mathbf{w}_p - \mathsf{M}\mathbf{X}_p) \\ & -\tfrac{1}{2} \sum_{p=1}^{N} (\widehat{\mathbf{X}}_p^0 - \mathbf{X}_p)^T \Sigma_{\tilde{\mathbf{X}}_p^0 \tilde{\mathbf{X}}_p^0}^{-1} (\widehat{\mathbf{X}}_p^0 - \mathbf{X}_p) \stackrel{!}{=} \max. \end{aligned} \tag{3.41}$$

This is the objective function maximized in [MRM94, MM95], with $\mathbf{t}$ replaced by its MAP estimate. Differentiating (3.41) w.r.t. M, we obtain

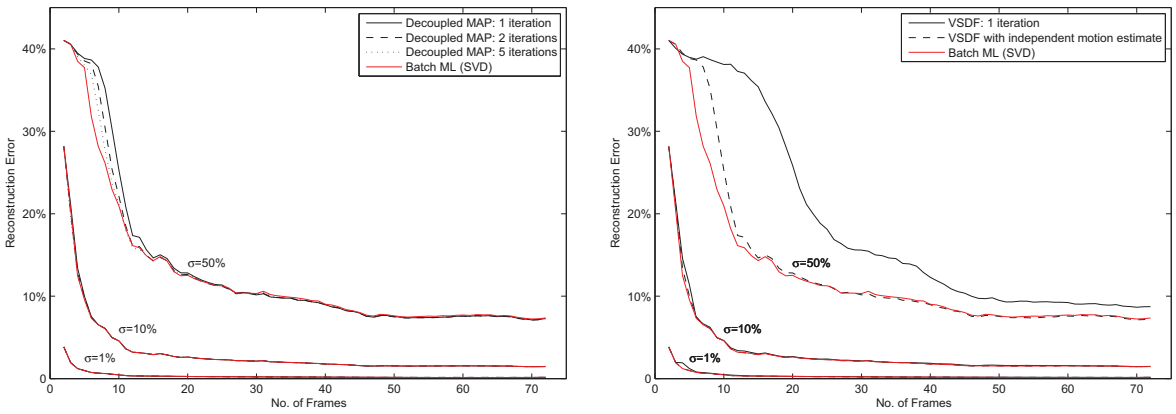
$$\frac{\partial g}{\partial \mathsf{M}} = \sum_{p=1}^{N} \Psi_{0,p}^{-1} \Delta \mathbf{w}_p \mathbf{X}_p^T - \sum_{p=1}^{N} \Psi_{0,p}^{-1} \mathsf{M} \mathbf{X}_p \mathbf{X}_p^T \stackrel{!}{=} \mathsf{0}_{2\times 3},$$

a set of six linear equations in the six unknown components of M. We can write the solution in terms of the column vector $\mathbf{m}$, consisting of the two rows of M, obtaining the estimator [3]

$$\widehat{\mathbf{m}} = \boldsymbol{m}(\widehat{\mathbf{X}}_p) \stackrel{\text{def}}{=} \left(\sum_{p=1}^{N} \Psi_{0,p}^{-1} \otimes \widehat{\mathbf{X}}_p \widehat{\mathbf{X}}_p^T \right)^{-1} \left(\sum_{p=1}^{N} \Psi_{0,p}^{-1} \Delta \mathbf{w}_p \otimes \widehat{\mathbf{X}}_p \right). \tag{3.42}$$

[3] With the notation $\mathsf{A} \otimes \mathsf{B}$ we refer to the matrix

$$\begin{bmatrix} a_{11}\mathsf{B} & a_{12}\mathsf{B} & \dots \\ a_{21}\mathsf{B} & a_{22}\mathsf{B} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$



(a) (b)

Figure 3.3: Comparison of the reconstruction accuracy of two recursive estimators (black) against that of the optimal ML batch estimator (red). On the left (a) is the decoupled recursive MAP estimator (3.37) and on the right (b) the VSDF estimator [MRM94] with and without an independent motion estimate.

We note that the estimation of M now involves inverting a larger $mr \times mr$ matrix, where m is the number of new rows added to the measurement matrix at each time step. Typically we have $mr = 2 \cdot 3 = 6$. The structure update (3.40) remains the same, apart from Ψ_0 being replaced by the individual feature covariance $\Psi_{0,p}$,

$$\widehat{\mathbf{X}}_p^{i+1} = \widehat{\mathbf{X}}_p^i + \Sigma_{\tilde{\mathbf{X}}_p^i \tilde{\mathbf{X}}_p^i} \mathcal{M}(\widehat{\mathsf{X}}^i)^T \Psi_{0,p}^{-1} \left(\Delta \mathbf{w}_p - \mathcal{M}(\widehat{\mathsf{X}}^i) \widehat{\mathbf{X}}_p^i \right). \tag{3.43}$$

The update of $\Sigma_{\tilde{\mathbf{X}}_p^i \tilde{\mathbf{X}}_p^i}$ is modified accordingly.

Comparison with the Variable State Dimension Filter

The recursive structure estimation method using the *variable state dimension filter (VSDF)* is a general MAP estimation method applicable to point and line features in projective, affine, or Euclidean views. For affine point features the VSDF method [MRM94] and the last form of the decoupled method just given, are two algorithms maximizing the same objective function (3.41). Rather than exploiting the structure of this particular problem in the way of (3.36), the VSDF computes and makes use of the covariance between the motion and structure parameters. In a real time setting, where typically only one or at most a few iterations can be computed within each step of the recursion, one can expect different results from the two methods if their rates of convergence differ. This is indeed observed

in Figure 3.3, where the decoupled MAP estimator (a) displays better convergence than the basic VSDF (b) at high levels of image noise. At moderate noise levels the accuracy of the two estimators is comparable. The figure shows results for a synthetically generated image sequence of a random set of 30 points rotating about an oblique axis at the rate of 15° per frame, replicating an experiment in [MRM94]. The reconstruction was repeated with three different levels of normal image noise with a standard deviation σ given as a percentage of the average apparent (RMS) size of the object in the images. The reconstruction error was measured by finding the 3D affine transformation A that most closely maps the known true structure X to the reconstruction $\widehat{\mathsf{X}}^i$. The residual RMS error e obtained from

$$e^2 \stackrel{\text{def}}{=} \min_{\mathsf{A}} \operatorname{tr} \left\{ (\widehat{\mathsf{X}}^i - \mathsf{A}\mathsf{X})^T (\widehat{\mathsf{X}}^i - \mathsf{A}\mathsf{X}) \right\}$$

is given in the figure as a percentage of the (RMS) size of the true structure. Shown in Figure 3.3 (a) is the reconstruction error of the decoupled recursive MAP estimators $\widehat{\mathsf{X}}^1$, $\widehat{\mathsf{X}}^2$, $\widehat{\mathsf{X}}^5$ of (3.37) and in Figure 3.3 (b) the reconstruction error of the VSDF estimator [MRM94], initialized by the previous motion estimate (solid black line). The reconstruction error of the optimal batch ML estimator (3.27) is given for comparison.

There is a twist, however. In order to remove the linearization error of the VSDF seen in Figure 3.3 (b) McLauchlan et al. [MRM94] proposed to first estimate the motion parameters independently of the structure to produce an initial estimate of the motion which for affine point features is essentially[4] the decoupled motion estimate (3.42). This estimate is then used as an initial value for the VSDF algorithm in place of the motion parameters from the previous time step. It can be shown that this particular initial value keeps the VSDF from updating the motion estimate, leaving the motion estimate at its initial value. It then follows directly that the VSDF does not make use of the computed covariance between motion and structure parameters so the structure estimate reduces to the decoupled structure estimate (3.43), as is apparent from the dotted line in Figure 3.3 (b).

Consequently, there is no advantage in using the VSDF for the estimation of affine point structure over the decoupled MAP estimator, since the VSDF displays poorer convergence unless initialized with an independent motion estimate, in which case it produces the same results as the decoupled MAP estimator but with redundant computation.

[4] [MRM94] bundle the translation parameters $\mathbf{t}$ with the parameters of M, typically estimating 8 instead of 6 parameters. Hence their equivalent to (3.42) involves inverting a larger 8×8 matrix.

Comparison with McLauchlan and Murray's Simple Recursive Structure Update

The decoupled structure update (3.43) has the same form as the *simple recursive structure update*, given by McLauchlan and Murray [MM95] as an approximation to the VSDF. There is an important difference however. In (3.43) the motion estimate $\widehat{\mathsf{M}} = \mathcal{M}(\widehat{\mathsf{X}}^i)$ is updated along with the estimated structure $\widehat{\mathsf{X}}^i$ using (3.42) in every step of the iteration such as to maximize the MAP criterion (3.35). In the *simple recursive structure update*, on the other hand, the motion estimate is kept at the initial estimate $\mathcal{M}(\widehat{\mathsf{X}}^0)$. With a fixed motion estimate the structure estimate is reached in a single step, since that problem is linear, and further iteration does not bring any improvement. Hence the *simple recursive structure update* can be seen as providing only the first iteration $\widehat{\mathbf{X}}_p^1$ of the decoupled structure estimate (3.43).

3.5 Testing for Dimensionality

We now proceed to the main subject of this chapter, deriving optimal statistical tests of the dimensionality of affine structure. The aim is to obtain a set of tests that are optimal in the sense of the Neyman-Pearson lemma irrespective of the rank r being tested for, the number of points N included in the tests, and the number of views F, in the case of direct application to unreconstructed views. To be able to devise effective search algorithms, we would like the tests to apply to the smallest non-trivial point set that can be tested for a particular rank r. Where tests are derived using the asymptotic or large sample distribution, the validity of the last property is verified.

The tests are based on the likelihood ratio as provided by the maximum likelihood estimators from the previous section. The limitations of the ML estimators discussed in Section 3.4.1 thus apply here as well. In particular, the likelihood ratio is undefined for unrestricted error covariance[5] as well as for an unknown diagonal covariance matrix. Two cases for which it is defined are: (1) when the error covariance is known completely, and (2) when the error covariance is known up to an unknown scale. We will derive separate tests for the two cases.

[5]When the error covariance is completely unspecified, replicated observations are needed in order to estimate it. The likelihood ratio criterion for testing the hypothesis of $q = q_0$ linear relationships against $q > q_0$, when replicated observations are available, is given in [And84a]

3.5.1 Likelihood Ratio Test for Known Error Variance

This section provides tests for the presence of a certain degeneracy in an affine structure, such as coplanarity or collinearity of points. Two forms of the tests are given, one for testing a recovered affine structure and a second form for testing directly the image locations in an affine view, without requiring prior reconstruction. The tests are formulated for the two measurement models discussed in Section 3.3, the parametric model and the restriction model. We further give a related method for the estimation of the dimensionality of the structure, as discussed in the introduction to this chapter.

A Test of Degeneracy

Based on the parametric model (3.4) this section provides a likelihood ratio test of the hypothesis whether the dimensionality r of the true structure is r_0 against the alternative hypothesis that it is greater than r_0

$$H_0 : r = r_0, \quad \text{against} \quad H_1 : r > r_0. \tag{3.44}$$

The likelihood function (3.15) under the alternative hypothesis is maximized at $r = M$, where an exact fit to the measurements is possible $\widehat{\Delta \mathsf{W}} = \mathsf{M}_M \mathsf{X}_M = \Delta \mathsf{W}$ with the effect that the exponent in (3.15) vanishes. Consequently the likelihood ratio

$$\Lambda_{r_0} \stackrel{\text{def}}{=} \frac{\max L_{r=r_0}}{\max L_{r>r_0}} \tag{3.45}$$

is

$$\Lambda_{r_0} = \frac{\max_{\mathsf{\Psi}, \mathsf{M}_{r_0}\mathsf{X}_{r_0}} \left\{ |\mathsf{\Psi}|^{-\frac{1}{2}N} \exp \operatorname{tr} \left\{ -\frac{1}{2} (\Delta \mathsf{W} - \mathsf{M}_{r_0}\mathsf{X}_{r_0})^T \mathsf{\Psi}^{-1} (\Delta \mathsf{W} - \mathsf{M}_{r_0}\mathsf{X}_{r_0}) \right\} \right\}}{\max_{\mathsf{\Psi}} \left\{ |\mathsf{\Psi}|^{-\frac{1}{2}N} \right\}}.$$

The denominator of this expression is only bounded for fixed (known) $\mathsf{\Psi} = \mathsf{\Psi}_0$. In this case the constants $|\mathsf{\Psi}_0|^{-\frac{1}{2}N}$ cancel leaving

$$\Lambda_{r_0} = \max_{\mathsf{M}_{r_0}\mathsf{X}_{r_0}} \left\{ \exp \operatorname{tr} \left\{ -\tfrac{1}{2} (\Delta \mathsf{W} - \mathsf{M}_{r_0}\mathsf{X}_{r_0})^T \mathsf{\Psi}_0^{-1} (\Delta \mathsf{W} - \mathsf{M}_{r_0}\mathsf{X}_{r_0}) \right\} \right\},$$

and consequently

$$\begin{aligned} -2 \log \Lambda_{r_0} &= \min \left\{ (\Delta \mathsf{W} - \mathsf{M}_{r_0}\mathsf{X}_{r_0})^T \mathsf{\Psi}_0^{-1} (\Delta \mathsf{W} - \mathsf{M}_{r_0}\mathsf{X}_{r_0}) \right\} \\ &= \min \left\| \Delta \mathsf{W}' - \mathsf{M}'_{r_0}\mathsf{X}_{r_0} \right\|_F^2, \end{aligned}$$

which by (3.25) is

$$\boxed{-2\log\Lambda_{r_0} = \sum_{i=r_0+1}^{M} {d'}_i^2,} \tag{3.46}$$

the squared sum of the $M - r_0$ smallest singular values of the Mahalanobis transformed measurement matrix $\Delta\mathsf{W}'$. It is easily shown[6] that the right hand side of (3.46) is in fact the test statistic suggested by Fisher [Fis38], with an asymptotic null distribution given by Hsu [Hsu41]. Equation (3.46) shows that this statistic is -2 times the logarithm of the likelihood ratio. Two things follow from this observation. Firstly, it is a general property of the likelihood ratio [MKB79], that the null distribution of -2 times the log likelihood ratio approaches the χ^2-distribution as $N \to \infty$. In the case of (3.46) this distribution has

$$\begin{aligned}\nu_r &= M(N-1) - [Mr + r(N-1) - r^2] \\ &= (M-r)(N-1-r)\end{aligned}$$

degrees of freedom. This provides an alternative to Hsu's proof of the asymptotic distribution of the statistic (3.46). Secondly, it follows from the *Neyman-Pearson lemma* [BNB96] that a test based on the likelihood ratio or equally its logarithm (3.46) is the *most powerful* test of the hypothesis (3.44), and any other statistic would lead to a suboptimal test.

From (3.46) follows the Neyman-Pearson test of the hypotheses (3.44): The null hypothesis is rejected at a chosen significance level α if

$$-2\log\Lambda_{r_0} > \chi^2_{(M-r_0)(N-1-r_0),\,\alpha}. \tag{3.47}$$

This being the asymptotic distribution, the propriety of the test in the small sample case cannot be guaranteed. For this reason we verify the distribution of the test statistic experimentally in Section 3.5.2.

We have obtained the following tests of degenerate structure, when the error covariance of the measurements is known. Table 3.4 gives the tests on a set of N points in an affine structure. Table 3.5 gives the tests on the joint measurement matrix of N points in F views.

[6] Fisher [Fis38] suggested the test statistic $(N-1)\sum_{i=r_0+1}^{M}\lambda_i$, where λ_i is the i-th largest root of the determinantal equation $|\mathsf{C} - \lambda\Psi_0| = 0$, involving the sample covariance matrix of the measurements $\mathsf{C} \stackrel{\text{def}}{=} \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{w}_i - \overline{\mathbf{w}})(\mathbf{w}_i - \overline{\mathbf{w}})^T$. It is easily verified that $d_i^2 = (N-1)\lambda_i$ so that the two statistics are identical.

H_0	H_1	N_{min}	Test
coplanar	non-coplanar	4	$-2\log\Lambda_2 \overset{?}{>} \chi^2_{N-3,\,\alpha}$
collinear	non-collinear	3	$-2\log\Lambda_1 \overset{?}{>} \chi^2_{2N-4,\,\alpha}$
coincident	non-coincident	2	$-2\log\Lambda_0 \overset{?}{>} \chi^2_{3N-3,\,\alpha}$

Table 3.4: Tests with which to reject a hypothesis of a degenerate affine structure, at a given probability α of false rejection, when the error covariance of the structure is known.

H_0	H_1	N_{min}	Test
coplanar	non-coplanar	4	$-2\log\Lambda_2 \overset{?}{>} \chi^2_{(2F-2)(N-3),\,\alpha}$
collinear	non-collinear	3	$-2\log\Lambda_1 \overset{?}{>} \chi^2_{(2F-1)(N-2),\,\alpha}$
coincident	non-coincident	2	$-2\log\Lambda_0 \overset{?}{>} \chi^2_{2F(N-1),\,\alpha}$

Table 3.5: Tests with which to reject a hypothesis of a degenerate scene structure from point correspondences between F views, at a given probability α of false rejection, when the error covariance of the image measurements is known.

A Test on the Number of Linear Restrictions

Tintner [Tin45] (see also [Tin46]) considered the dual problem of testing the number q of linear restrictions existing on the space of true measurements. To test the hypothesis

$$H_0 : q = q_0, \quad \text{against} \quad H_1 : q < q_0, \tag{3.48}$$

whether there are q_0 or less than q_0 independent relationships between the systematic variables $\mathbf{x}_p$, i.e. that the restriction matrix B_q in (3.9) has rank q_0 or less than q_0, he uses the dual form of Fisher's test function, which by (3.10) and (3.46) can be written as

$$-2\log\Lambda^*_{q_0} \overset{\text{def}}{=} -2\log\Lambda_{M-q_0}$$

Hence $-2\log\Lambda^*_q$ has the distribution $\chi^2_{q(N-M-1+q)}$. From (3.47) follows that the null hypothesis H_0 (3.48) is rejected at a chosen significance level α if

$$-2\log\Lambda^*_{q_0} > \chi^2_{q_0(N-M-1+q_0),\,\alpha}. \tag{3.49}$$

Estimation of the Number of Linear Restrictions

To estimate the number of linear restrictions q existing between the true part $\mathbf{x}_p$ of the measurements, i.e. the rank of the restriction matrix B_q, Tintner proposed the following test procedure in a later paper [Tin46]:

> Apply the test (3.49) for each hypothesized number q_0 in the sequence $1, 2, \ldots, M$ until the null hypothesis is rejected at the specified significance level. Take the estimated number $\hat{q}$ as the last value of q_0 for which H_0 was not rejected.

The test (3.49) is only available if the number of degrees of freedom is greater than zero

$$q_0(N - M - 1 + q_0) > 0. \tag{3.50}$$

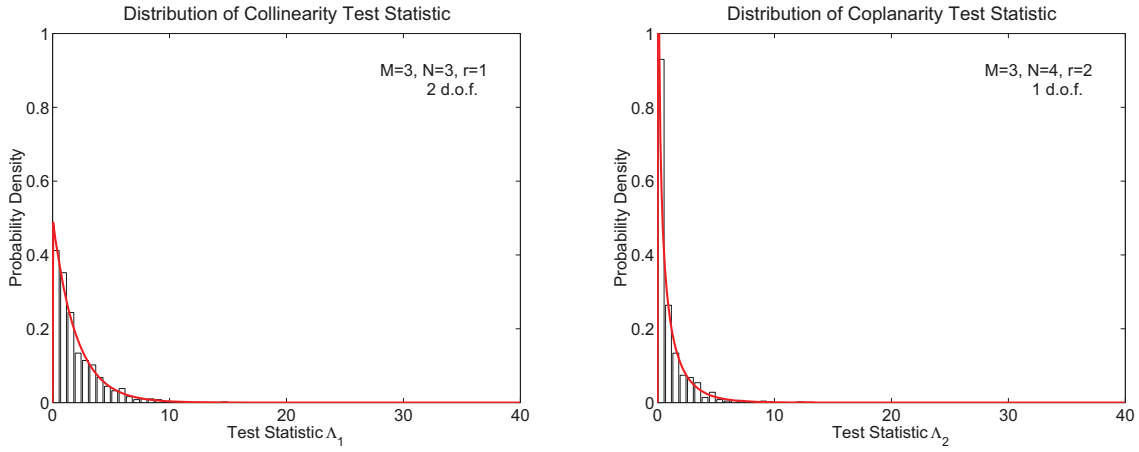
Depending on the values of N and M a test may therefore not exist for small q_0. In Tintner's procedure, a test is assumed to be available for all $q_0 \geq 1$, requiring $N > M$. This requirement is easily removed as follows: The registered measurement matrix has at most $N-1$ independent columns, each consisting of M coordinates. If $N \leq M$ then the columns span at most an $N - 1$ dimensional subspace in M-space. Hence there must exist at least $M - (N - 1)$ linear restrictions on the measurement matrix, so we only need to commence our testing sequence at $q_0 = M - (N - 1) + 1$, at which (3.50) is satisfied.

Estimation of Rank

The dual test procedure is obtained using the substitution (3.10). For r_0 in the sequence $M - 1, M - 2, \ldots, 0$ test the dual hypothesis (3.44) until H_0 is rejected at the specified significance level α, by the test (3.47). Take the estimated rank $\hat{r}$ as the last value of r_0 for which H_0 was not rejected. For $N \leq M$ we know that $r \leq N-1$, and hence we commence the testing sequence at $r_0 = N - 2$.

3.5.2 Applicability to Small Sample Sizes

The statistical tests given in Table 3.4 and Table 3.5 provide us with a single mechanism to test for rank deficiency in the presence of noise, for different rank r_0, number of points N, and number of views F, in the case of direct application to unreconstructed views. Each test is parametrized only in the acceptable false-rejection rate α. The null distribution of the



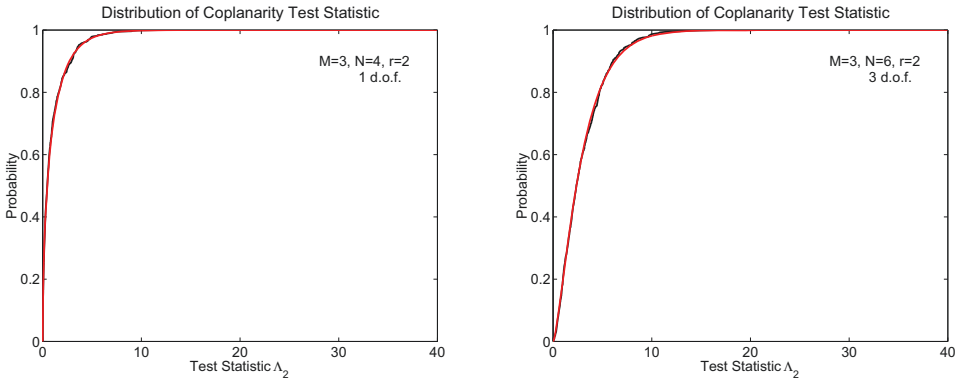
(a) (b)

Figure 3.4: The null distribution of the degeneracy test statistic for a 3D affine structure with known variance for the minimal sets of (a) three collinear points, and (b) four coplanar points. The histogram of the empirical statistic is shown in black and the theoretical probability density function (derived for large sample sizes) in red.

test statistic was derived in the asymptotic case, i.e. assuming large sample sizes. In this thesis small sample sizes are of particular interest for use in effective search algorithms. We therefore test whether the asymptotic null distribution also applies to small samples. For this we need the empirical null distribution for small samples.

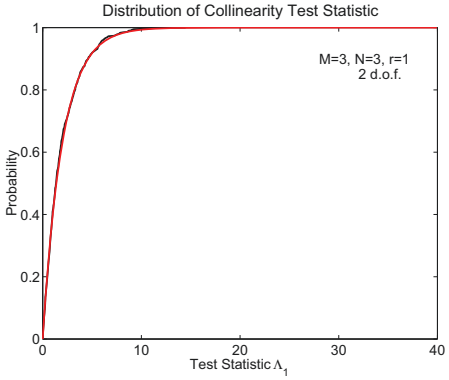
Empirical Small Sample Distribution

To create the empirical small sample null distribution, we generated a set of synthetic affine structures and measurement matrices of rank r, and added normally distributed noise of known variance [KH95]. For the tests on 3D structure, we repeatedly computed the test statistic (3.46) for the minimal non-trivial set size $N = N_{\min}$ as well as for $N = N_{\min} + 2$. For the tests on the joint measurement matrix, we computed the test statistic for the minimal non-trivial set size $N = N_{\min}$ in $F = 2$ views as well as for $N = N_{\min} + 2$ in $F = 3$ views. Figure 3.4 shows histograms of the outcome of 800 trials for the smallest non-trivially degenerate 3D point sets of (a) three collinear points and (b) four coplanar points along with the theoretical probability density function from Table 3.4. It is apparent from the figure that the theoretical and empirical distributions match closely for the two minimal sets.

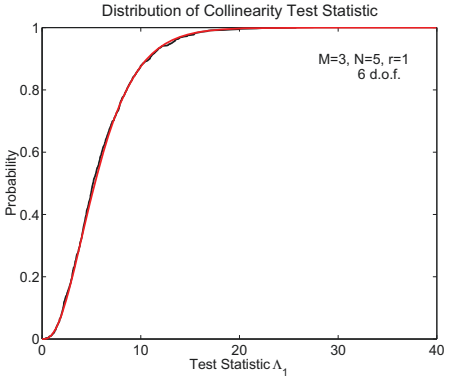


(a) Four coplanar points.

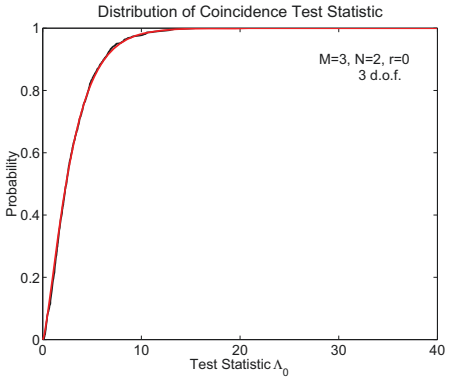
(b) Six coplanar points.



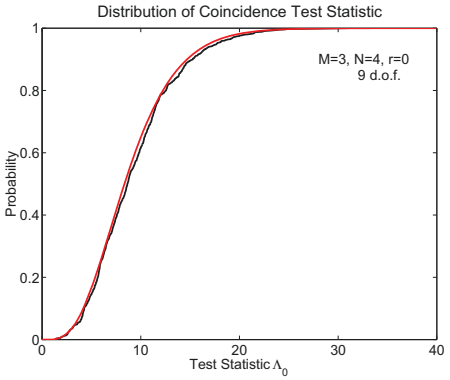
(c) Three collinear points.



(d) Five collinear points.



(e) Two coincident points.

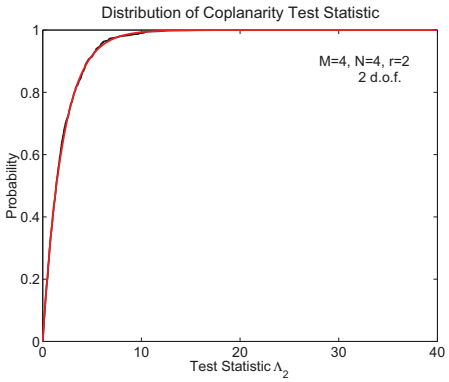


(f) Four coincident points.

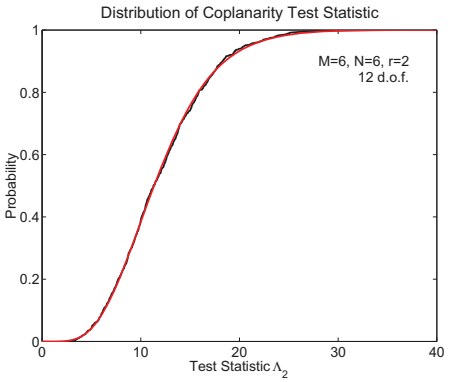
Figure 3.5: The empirical small sample distribution (black) and the theoretical asymptotic distribution (red) of the degeneracy test statistic for a *3D affine structure with known variance*. The figures on the left hand side show the smallest non-trivially degenerate sets of points, while the figures on the right hand side have two additional points.

Testing the Applicability of the Asymptotic Distribution

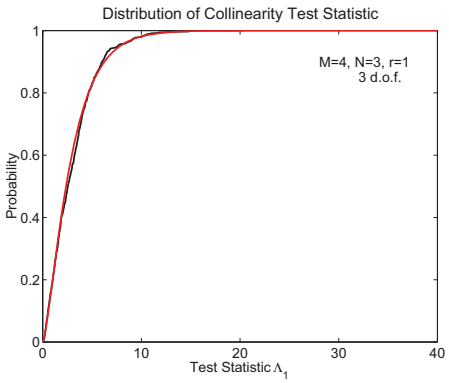
A statistically meaningful comparison of the theoretical and empirical distribution is facilitated by the well known *Kolmogorov-Smirnov test* (see e.g. [BNB96]). It tests the hypoth-



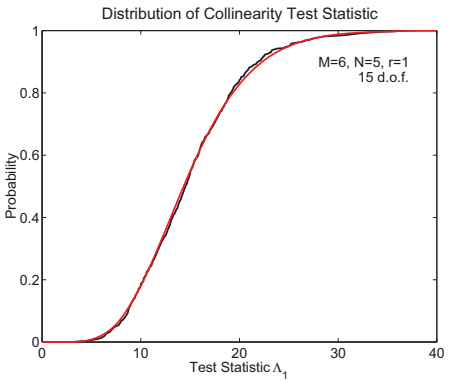
(a) Four coplanar points in two views.



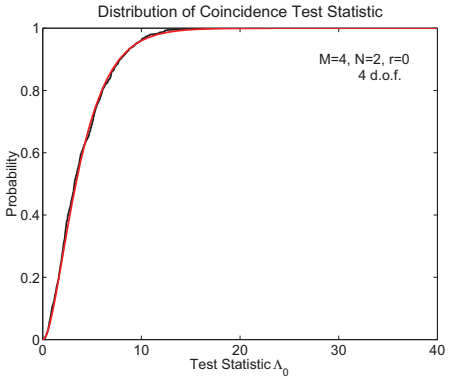
(b) Six coplanar points in three views.



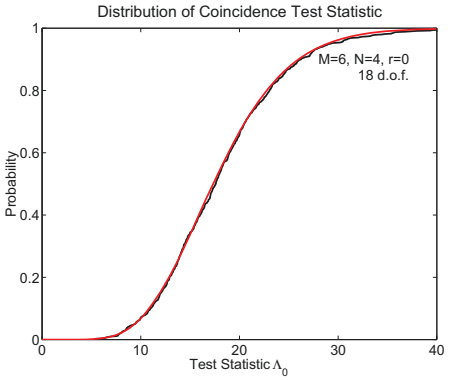
(c) Three collinear points in two views.



(d) Five collinear points in three views.



(e) Two coincident points in two views.



(f) Four coincident points in three views.

Figure 3.6: The empirical small sample distribution (black) and the theoretical asymptotic distribution (red) of the degeneracy test statistic from *point correspondences in two or more views with known error covariance*. The figures on the left hand side show the smallest non-trivially degenerate sets of point correspondences from two views. The figures on the right hand side show the same test statistics for three views and two additional correspondences.

esis that the samples making up the empirical distribution are drawn from the theoretical distribution. The test is applied to the *cumulative distribution functions*. The empirical and theoretical cumulative distribution functions are shown in Figure 3.5 for 3D structure and in Figure 3.6 for point correspondences in two and three views. The results of the Kolmogorov-Smirnov tests are given in Tables 3.6 and 3.7. In all cases the empirical distribution is accepted at the significance level $\alpha = 0.01$.

M	N	r	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
3	4	2	0.650	0.791	pass
3	3	1	0.689	0.728	pass
3	2	0	0.804	0.537	pass
3	6	2	0.698	0.714	pass
3	5	1	0.757	0.614	pass
3	4	0	1.085	0.189	pass

Table 3.6: Results of the Kolmogorov-Smirnov test on the null distributions of the test statistics on the $3 \times N$ structure matrices shown in Figure 3.5.

M	N	r	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
4	4	2	0.547	0.924	pass
4	3	1	1.413	0.036	pass
4	2	0	1.056	0.214	pass
6	6	2	0.495	0.966	pass
6	5	1	0.652	0.788	pass
6	4	0	0.659	0.776	pass

Table 3.7: Results of the Kolmogorov-Smirnov test on the null distributions of the test statistics on the $M \times N$ joint measurement matrices shown in Figure 3.6.

3.5.3 Likelihood Ratio Test for Error Variance Known up to Scale

In this section we evaluate a likelihood ratio test on the dimensionality of an affine structure that requires the error variance to be known only up to an unknown scale. This is the *sphericity test* due to Anderson [And63], which is based on testing whether the $M - r$ smallest eigenvalues of the covariance matrix Σ are equal in size, as one would expect if they contained only measurement noise. This method makes no assumption about the size of the image noise, the error variance being known up to scale only.

The Sphericity Test

Given a sample of size N from the $N_M(\mathbf{t}, \Sigma)$ distribution, the likelihood ratio statistic for testing the null hypothesis

$$H_r : \lambda_{r+1} = \cdots = \lambda_M \quad (= \lambda, \text{unknown}) \tag{3.51}$$

is $\Lambda_r = V_r^{N/2}$, where V_r is the *ellipticity statistic* [And63][Mui82, p. 407]

$$V_r \stackrel{\text{def}}{=} \frac{\prod_{i=r+1}^{M} l_i}{\left(\frac{1}{M-r}\sum_{i=r+1}^{M} l_i\right)^{M-r}}.$$

The exact distribution of this statistic is unknown, but the asymptotic (large sample) distribution was derived by Anderson [And63]. The general theory of likelihood ratio tests shows that asymptotically (as $N \to \infty$)

$$P_r^A = -2\log\Lambda_r = -N\log V_r \quad \text{is} \quad \chi^2_{(q+2)(q-1)/2} \tag{3.52}$$

when H_r is true [And63][And84b, p. 477][SK79, p. 293][Mui82, p. 408]. Using heuristic arguments Bartlett suggested a different multiplying factor [Mui82, p. 408]

$$P_r^B = -\left[N - 1 - r - \frac{2q^2 + q + 2}{6q}\right]\log V_r, \tag{3.53}$$

and Lawley added a term to eliminate the effects of the r largest latent roots [And63][Mui82, p. 409]

$$P_r^L = -\left[N - 1 - r - \frac{2q^2 + q + 2}{6q} + \sum_{i=1}^{r}\frac{\bar{l}_q^2}{(l_i - \bar{l}_q)^2}\right]\log V_r. \tag{3.54}$$

The sphericity test requires that $M > r + 1$ and $N > M$. In order to be able test for all three types of degeneracy ($r = 0, 1, 2$), we must therefore have $M > 3$. This rules out tests

H_0	H_1	N_{min}	Test
coplanar	non-coplanar	$M+1 \geq 5$	$P_2 \overset{?}{>} \chi^2_{M(M-3)/2,\alpha}$
collinear	non-collinear	$M+1 \geq 5$	$P_1 \overset{?}{>} \chi^2_{(M+1)(M-2)/2,\alpha}$
coincident	non-coincident	$M+1 \geq 5$	$P_0 \overset{?}{>} \chi^2_{(M+2)(M-1)/2,\alpha}$

Table 3.8: Tests with which to reject a hypothesis of a degenerate scene structure from point correspondences between F views, at a given probability α of false rejection, when the error covariance is known up to scale. There are three versions of each test, depending on whether P_r is the test statistic of Anderson (3.52), Bartlett (3.53), or Lawley (3.54).

on recovered 3D structure (where $M = 3$), but allows tests to be conducted directly on two or more affine views ($M = 2F \geq 4$), as given in Table 3.8. The requirement $N > M$ means that at least five correspondences are needed for the minimum number of two views.

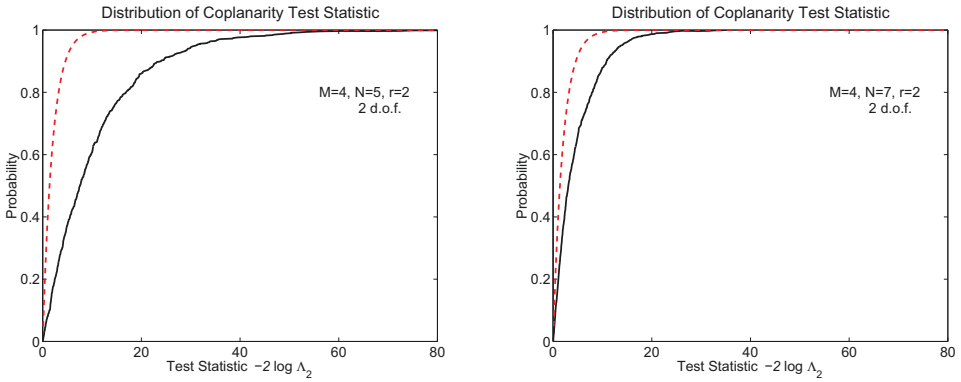
Applicabilty of the Sphericity Test to Small Sample Sizes

The chi-square distribution given in (3.52) is the asymptotic distribution of the of the sphericity test statistic. To test whether this distribution also applies to small samples, as in the case of the tests of known variance, we apply the methods used in Section 3.5.2.

The empirical distribution of the three variants of the sphericity test statistic (3.52)-(3.54)

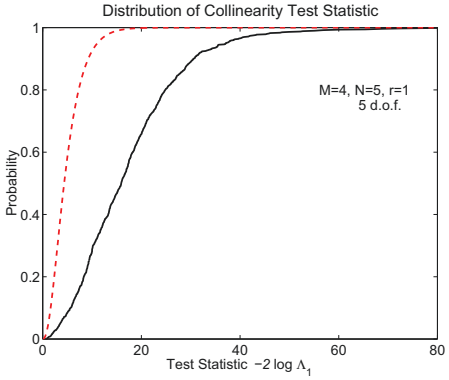
M	N	r	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
4	5	2	15.95	10^{-221}	reject
4	5	1	18.93	10^{-312}	reject
4	5	0	18.71	10^{-304}	reject
4	7	2	8.214	10^{-59}	reject
4	7	1	11.56	10^{-116}	reject
4	7	0	11.94	10^{-124}	reject
6	7	2	22.49	10^{-440}	reject
6	7	1	22.62	10^{-445}	reject
6	7	0	21.99	10^{-420}	reject

Table 3.9: Results of the Kolmogorov-Smirnov test on the null distributions of Anderson's test statistic (3.52) in Figure 3.7.

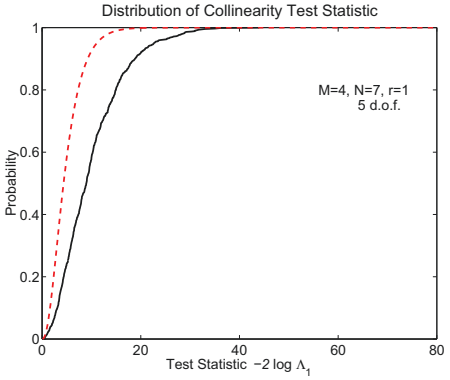


(a) Five coplanar points.

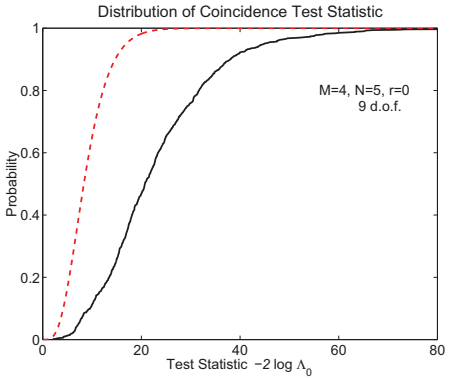
(b) Seven coplanar points.



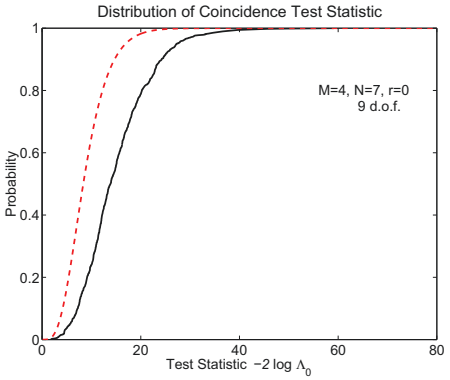
(c) Five collinear points.



(d) Seven collinear points.



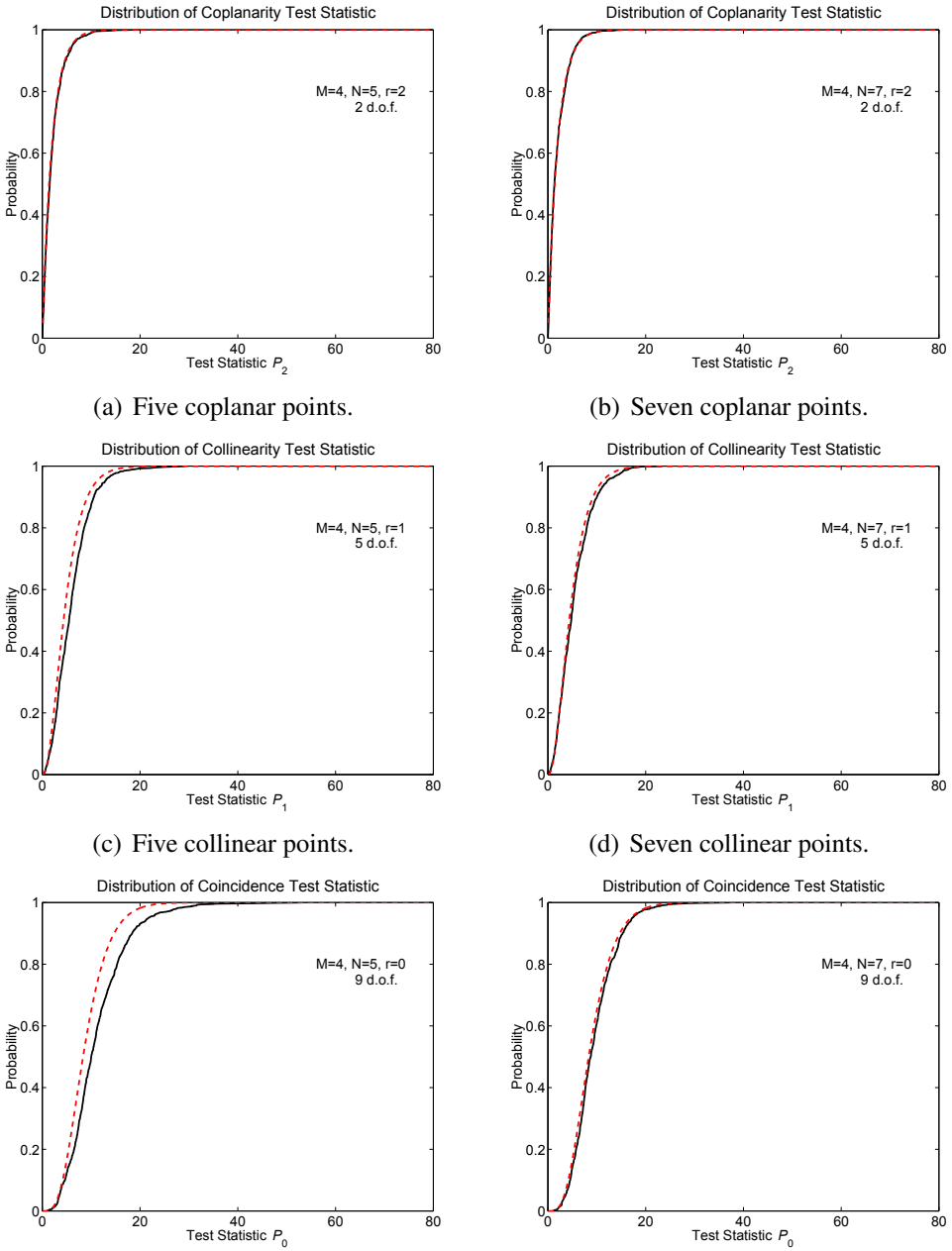
(e) Five coincident points.



(f) Seven coincident points.

Figure 3.7: The empirical small sample distribution (black) and the theoretical asymptotic distribution (red) of *Anderson's sphericity test* (3.52) applied to point correspondences in two views with variance known up to scale.

was generated by repeated trials on synthetic measurement matrices. The corresponding empirical cumulative distributon function is shown in Figures 3.7 to 3.9, respectively, together with the asymptotic distributon. Comparing the three figures, it is apparent that the



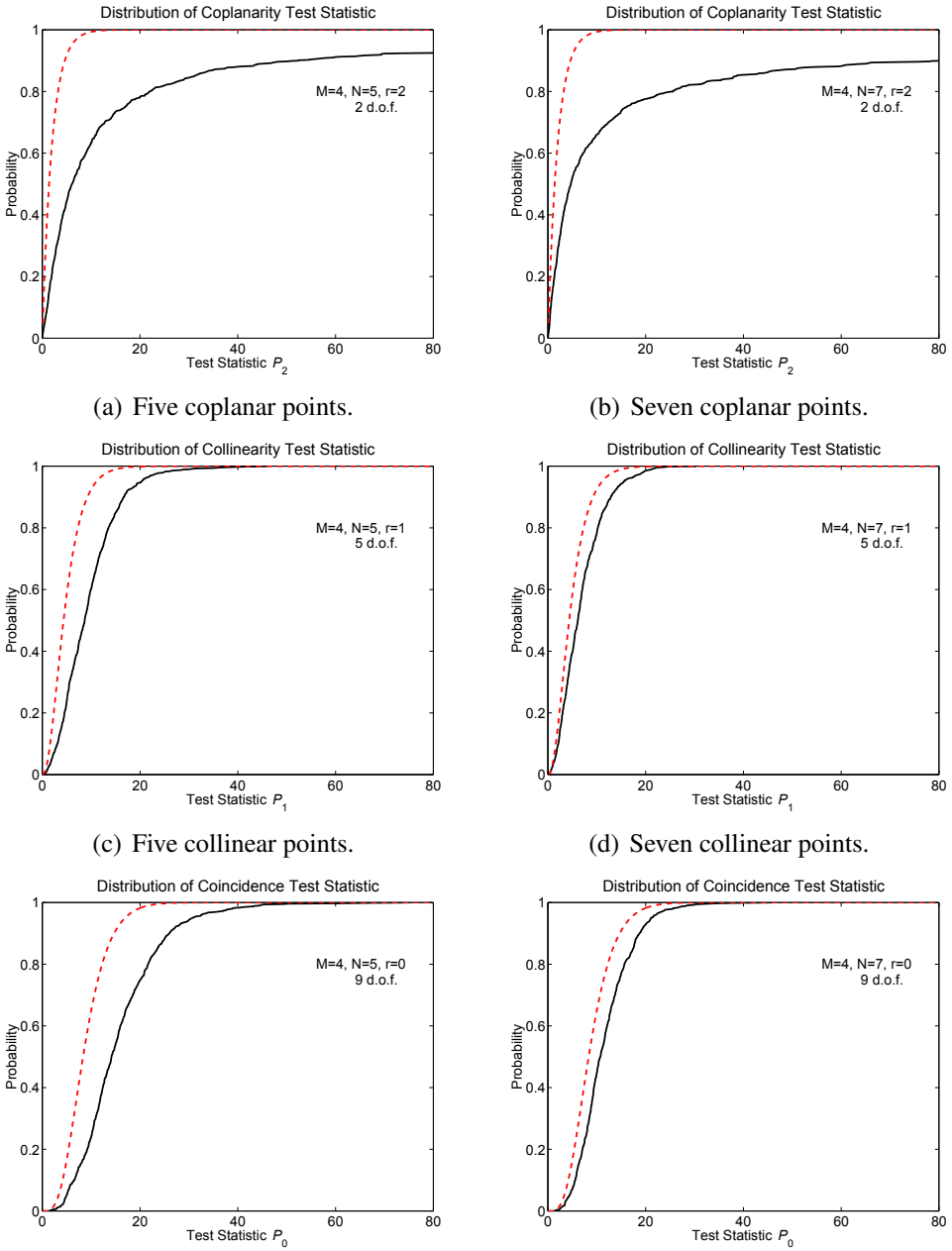
(a) Five coplanar points. (b) Seven coplanar points.

(c) Five collinear points. (d) Seven collinear points.

(e) Five coincident points. (f) Seven coincident points.

Figure 3.8: The empirical small sample distribution (black) and the theoretical asymptotic distribution (red) of *Bartlett's variant* of the sphericity test (3.53) applied to point correspondences in two views with variance known up to scale.

Bartlett variant of the sphericity test (Figure 3.8) is best described by the chi-square distribution. The Kolmogorov-Smirnov test results for the three variants are given in Tables 3.9 to 3.11. Three small sample cases are tabulated: five points in two views, and seven points



(a) Five coplanar points. (b) Seven coplanar points.

(c) Five collinear points. (d) Seven collinear points.

(e) Five coincident points. (f) Seven coincident points.

Figure 3.9: The empirical small sample distribution (black) and the theoretical asymptotic distribution (red) of tests of *Lawley's variant* of the sphericity test (3.54) applied to point correspondences in two views with variance known up to scale.

in two or three views. Only Bartlett's variant (3.53) passes the Kolmogorov-Smirnov test at significance level $\alpha = 0.01$, and then only in the case of two views ($M = 4$), but not in the case of three views ($M = 6$). It may seem counter-intuitive that the fit worsens as more

M	N	r	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
4	5	2	1.091	0.184	pass
4	5	1	4.101	10^{-15}	reject
4	5	0	4.512	10^{-18}	reject
4	7	2	0.711	0.692	pass
4	7	1	1.605	0.0115	pass
4	7	0	1.620	0.0105	pass
6	7	2	4.797	10^{-20}	reject
6	7	1	7.061	10^{-44}	reject
6	7	0	7.430	10^{-48}	reject

Table 3.10: Results of the Kolmogorov-Smirnov test on the null distributions of Bartlett's test statistic (3.53) in Figure 3.8.

M	N	r	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
4	5	2	13.82	10^{-166}	reject
4	5	1	11.25	10^{-110}	reject
4	5	0	12.02	10^{-126}	reject
4	7	2	11.52	10^{-116}	reject
4	7	1	5.893	10^{-30}	reject
4	7	0	6.121	10^{-33}	reject
6	7	2	12.53	10^{-137}	reject
6	7	1	13.51	10^{-159}	reject
6	7	0	13.93	10^{-169}	reject

Table 3.11: Results of the Kolmogorov-Smirnov test on the null distributions of Lawley's test statistic (3.54) in Figure 3.9.

views become available, but this is understandable when one considers that the number of latent roots tested for equality goes up with M without a corresponding increase in the sample size N.

Degeneracy Test with Known Error Variance

Hypothesis (3.44)

$$H_0 : r = r_0, \quad \text{against} \quad H_1 : r > r_0$$

Test statistic (3.46)

$$-2 \log \Lambda_{r_0} = \sum_{i=r_0+1}^{M} d_i'^2$$

Reject H_0 at chosen significance level α if (3.47)

$$-2 \log \Lambda_{r_0} > \chi^2_{(M-r_0)(N-1-r_0),\,\alpha}$$

Table 3.12: A summary of tests of degeneracy with known error variance.

3.5.4 Summary and Discussion

We have obtained two likelihood ratio tests for degenerate scene point configurations. They can be applied when measurements are perturbed by noise and the singular values do not vanish as the Extended Rank Theorem from Section 2.5.1 would predict for noise-free measurements. The tests are summarized in Table 3.12 for the case of known error variance and in Table 3.13 for the case of error variance known up to scale. The theoretical distributions provide us with a decision threshold for a given rank r_0, number of points N, and number of views $F = M/2$. The threshold value is parametrized only in a single design parameter: the significance level α of the hypothesis test, which describes the *acceptable probability of false rejection*. The accuracy with which α describes the actual false rejection rate is determined by the fit between the empirical and the theoretical distribution. For the test with known error variance the cumulative density functions fit closely as seen in Figures 3.5 and 3.6. For Bartlett's variant of the test with error variance known up to scale (Figure 3.8) the fit is not as good, the emprirical cumulative probability being slightly lower than predicted by the chi-square distribution. Consequently, the acceptable false rejection rate $\alpha = 1 - p$ is reached at a higher threshold value than predicted by the theoretical

Degeneracy Test with Error Variance Known up to Scale

Hypothesis (3.51)

$$H_0 : {d'}^2_{r_0+1} = \cdots = {d'}^2_M$$

Test statistic (3.53)

$$P^B_{r_0} = -\left[N - 1 - r_0 - \frac{2(M-r_0)^2 + (M-r_0) + 2}{6(M-r_0)}\right] \log \frac{\prod_{i=r_0+1}^{M} {d'}_i}{\left(\frac{1}{M-r_0}\sum_{i=r_0+1}^{M} {d'}_i\right)^{M-r_0}}$$

Reject H_0 at chosen significance level α if

$$P^B_{r_0} > \chi^2_{(M-r_0+2)(M-r_0-1)/2,\,\alpha}$$

Table 3.13: A summary of tests of degeneracy with error variance known up to scale.

distribution, as can be read off the figure. For the type of degeneracy of greatest interest in this thesis: the coplanarity and the collinearity statistic from point correspondences in two views, however, this difference is minimal.

3.6 Conclusions

This chapter has presented optimal statistical methods for detecting degeneracy in a set of 3D scene points, either from a recovered affine structure or directly from two or more affine views. In so doing it has drawn extensively on the statistical literature and brought together in a statistical framework two of the best known methods of affine structure from motion: factorization and recursive VSDF. Among the particular results in this chapter we have:

1. The question whether a particular degeneracy, such as collinearity or coplanarity, is present in the scene is more naturally posed as a hypothesis test than a model selection problem. The optimal statistical solution to the former has the added practical advantage that the ratio of the two errors: false alarm and mis-detection can be varied to suit the application.

2. The optimal ML estimator of affine structure under normal measurement errors does not exist when the error covariance is completely unknown, even when the error covariance is restricted to diagonal form.

3. When the error covariance is known either completely or up to scale and a prior structure estimate is not available, then the optimal ML estimator of affine structure of any dimension is provided by factorization using the SVD. It is shown how to define the factorization such as to achieve isotropic errors on the recovered structure.

4. When a prior structure estimate is available, or when the measurements are presented incrementally for recursive estimation, then the optimal MAP estimator is the variable state dimension filter or VSDF, with factorization providing the bootstrap to the recursion. A very simple and effective form of this method, the *decoupled recursive MAP estimator,* can be used in the case of affine point structure from affine views.

5. The optimal statistical method for detecting a particular degeneracy, when the error covariance is completely known, is provided by Fisher's test statistic or dually by Tintner's statistic, both of which are shown to be a log likelihood ratio. The Fisher-Tintner test is shown to be well suited to small sets of scene points.

6. The sphericity test statistic given by Anderson for error covariance known up to scale is not suited to test a *small* set of scene points, since the small sample distribution of the statistic is inadequately described by the known asymptotic distribution. This problem is fixed in limited cases by Bartlett's variant of the sphericity test. The minimum number of points required for the sphericity test is also significantly higher than for the Fisher-Tintner test.

4

Searching for Symmetry within a Scene

4.1 Overview

This chapter considers the problem of detecting skewed symmetric point sets embedded within a larger set of scene points, given either as a 3D affine structure or as a set of point correspondences across two or more affine views. Following an introduction to the problem in Section 4.2, a deterministic search algorithm is given in Section 4.3. The implementation of the algorithm is described in Section 4.4 and fully automatic detection of bilateral symmetry is demonstrated and evaluated on real images of uncluttered scenes. An alternative approach to the search problem based on random sampling is discussed in Section 4.5. A randomized search algorithm is given in Section 4.6 along with experimental results on moderately cluttered scenes. Before concluding the topic, Section 4.7 briefly discusses the possibilities of accommodating further constraints or appearance based information to speed up the search. An early version of this work was presented in [Thó96].

4.2 Introduction and Problem Definition

The symmetry detection method proposed in the previous two chapters allows us to test a set of hypothetical chords, known either in an affine structure or from affine image correspondences, for compliance with a set-wide skewed symmetry. We now turn to the following problem: Given a (possibly large) set of point correspondences across two or more affine views or an affine reconstruction, we wish to detect (possibly small) *subsets* of these features which are skew symmetric. The motivation behind this problem is to provide:

- a *grouping* or *interest operator* for object recognition, based on the assumption that a symmetric group of 3D features recovered from a scene may be more likely to belong to one object (or a particular arrangement of objects) than does a set of features which does not display this kind of regularity,
- the means to *rectify* an affine structure, ie. upgrade it to a scaled Euclidean structure, using the methods developed in the next chapter of this thesis. This is effectively a detector of *bilateral* symmetry as opposed to skewed symmetry.

The two proposed application areas are linked in that a detector of bilateral rather than skewed symmetry is a more selective grouping operator. The knowledge of the Euclidean structure also provides tighter constraints for object recognition.

Before continuing, let us look at the combinatorics of such an undertaking. Let $\mathcal{C}$ be the set of all distinct unordered pairs of N point features. This set has cardinality $|\mathcal{C}| = \binom{N}{2} = \frac{N(N-1)}{2} \approx \frac{N^2}{2}$. Consequently, the total number of distinct subsets that can be drawn from this set (amounting to the *powerset* of $\mathcal{C}$) is $|\mathrm{Pow}(\mathcal{C})| = 2^{|\mathcal{C}|} = 2^{\binom{N}{2}} \approx 2^{\frac{N^2}{2}}$. To help bring out the enormity of the search problem, these numbers for an original set of $N = 50$ detected point features are $|\mathcal{C}| = 1225$ and a staggering $|\mathrm{Pow}(\mathcal{C})| = 10^{369}$, respectively. It is therefore obvious that attention must be given to the way in which these possibilities are examined. Contemplating different potential applications of a symmetry detection algorithm leads us to the following three problems, whose solution may differ both in tractability and method:

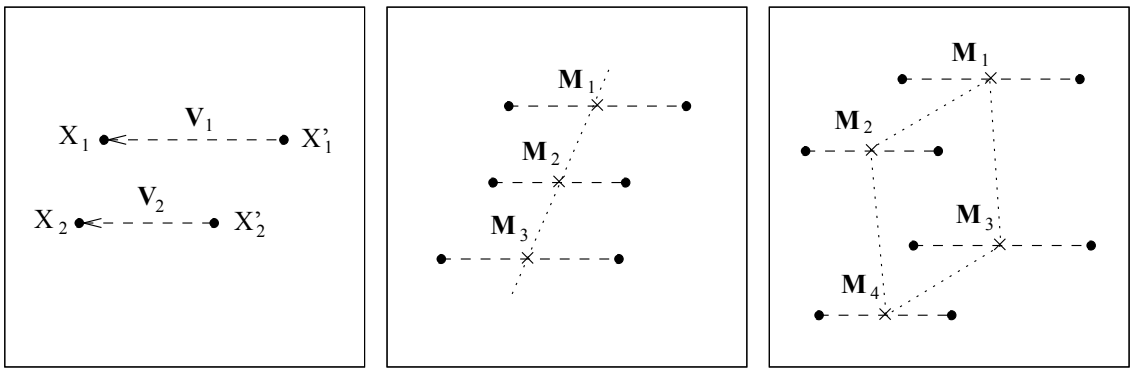
1. Partition the complete powerset $\mathrm{Pow}(\mathcal{C})$ into two partitions, one containing all skew symmetric sets, and another containing the rest.

2. Find the in some sense *most prominent* symmetry sets in $\mathrm{Pow}(\mathcal{C})$, where the prominence J of a set $\mathcal{S} \in \mathrm{Pow}(\mathcal{C})$ could be given by its cardinality $J(\mathcal{S}) \stackrel{\mathrm{def}}{=} |\mathcal{S}|$. This problem could either be defined as finding all sets with prominence $J > J_{\min}$ or the n most prominent sets, for a specified number n.

3. In the special case of $n = 1$, problem 2 consists in finding only the *single* most prominent symmetric set. This could then be used to successively recover the n most prominent *disjoint* symmetric sets.

At a first glance, Problem 1 above may look like the holy grail of symmetry detection. It should be noted however, that this problem implies that each m chord set $\mathcal{S}$ in the symmetric partition of $\mathrm{Pow}(\mathcal{C})$ is accompanied by a total of $2^m - 2$ symmetric sets of size $m-1$ and smaller that make up its powerset $\mathrm{Pow}(\mathcal{S})$. It is left to the application to eliminate these. This observation also applies to Problem 2, although to a lesser extent due to the prominence requirement. It does not apply to Problem 3, however, if we assume that the chords of each detected set $\mathcal{S}$ are removed from further consideration, the next set being searched from amongst the elements of $\mathrm{Pow}(\mathcal{C} \setminus \mathcal{S})$ only. That procedure has the side effect of also eliminating from consideration all candidates in the non-empty set $\mathrm{Pow}(\mathcal{C}) \setminus \mathrm{Pow}(\mathcal{C} \setminus \mathcal{S})$. Comparing the three problems, the two latter ones are closer in character to an interest operator, this being the kind of usage commonly given to image intensity corner and edge detectors.

A further observation is that due to measurement errors, each set has a finite probability of being incorrectly classified. The effect is that *however thorough the search, it cannot solve the above problems with perfect accuracy*. This opens the question whether deterministic search algorithms should only be considered, or whether consideration should also be given to random search algorithms that, although adding some uncertainty on their own, might possibly prove more efficient or have other advantages over deterministic algorithms. We will now turn our attention to deterministic search algorithms, discussing random algorithms in Section 4.5.



(a) (b) (c)

Figure 4.1: Testing a 3D affine structure for skewed symmetry involves (a) testing each pair of line segments connecting any two points for 3D parallelism, (b) (optionally) testing parallel line segments for 3D collinearity of midpoints, and (c) testing any four (or more) parallel segments for coplanarity of midpoints.

4.3 Deterministic Search

Chapter 2 states three equivalent formulations of 3D skewed symmetry. The first of these defines 3D skewed symmetry in terms of the *chords* connecting symmetric point pairs (Section 2.4.1). That formulation was found to be particularly well suited for the development of a statistical detector as reported in Chapter 3. This formulation has also computational advantages.

4.3.1 Sequential Application of Skewed Symmetry Constraints

By the definition given in Section 2.4.1 a 3D skewed symmetric point set constitutes any set of point pairs whose supporting lines are parallel in 3D and whose midpoints lie in a common plane, as depicted in Figure 4.1. Since these two constraints have to hold between all chords in the set, this might seem to suggest applying both tests simultaneously. How-

test	min. no. of chords
parallelism of chords (2.22)	2
coplanarity of midpoints (2.23)	4

Table 4.1: The minimum number of chords required to perform each of the component tests of skewed symmetry as defined in Section 2.5.2.

ever, as seen from Table 4.1 the minimal set of chords required for the test of parallelism is considerably lower than for the test of coplanarity. This suggests using the test of parallelism to prune the set of candidate chords prior to applying the test of coplanarity, in order to gain computational efficiency. The sequential application introduces a couple of auxiliary constraints. Together, the available constraints can be summarized as follows:

1. A single point pair specifies a spatial direction. A single pair can thus act as a basis against which any further pairs can be tested for *parallelism*.

2. A pair of points a short distance apart compared to the size of the uncertainty in the position of the endpoints has a large uncertainty on its direction. As a consequence it will test parallel to chords of a wide range of directions.

3. The minimum number of points spanning a plane is three. Three mid-points in a general position can act as a basis against which any further mid-points can be tested for *coplanarity*.

4. A *collinear* set of basis points will not give meaningful results, since it will test coplanar with any additional point.

Implemented using the statistical tests from Chapter 3, these constraints can be applied to sets of any size above a minimal size, which varies between one and four point pairs for each constraint. The size of the powerset $\mathrm{Pow}(\mathcal{C})$, however, excludes an exhaustive application to all possible sets of point pairs. Applying the constraints while growing each skewed symmetric set from one seed chord to its maximum size seems to be a more viable option.

The procedure of growing symmetric sets by introducing point pairs one at a time introduces the following subtlety. It is clear that the addition of a further point into a set necessarily increases (or at best leaves unchanged) the sum of squared distances of those points from any given linear subspace, including a subspace that minimizes that sum. The *average* distance of the points from the linear subspace, on the other hand, can either increase or decrease. While based on the *chi-squared* distribution, rather than the average, the effect on the tests given in Chapter 3 is that the fact that a set $\mathcal{S}_m$ consisting of m point pairs passes the tests, does not guarantee that every subset of $\mathcal{S}_m$ passes the same tests,

although this may be a more common failure in borderline cases. Consequently, it is not guaranteed that every growth-path into a skewed symmetric set will be unbroken.

4.3.2 Constraint Satisfaction and Directional Partitioning

The problem of symmetry detection within a scene by growing symmetric sets, falls in the category of *constraint satisfaction problems*, along with many classical problems of artificial intelligence (see [RN95]). This category includes well known problems of computer vision such as the matching of data features to object models under geometric model constraints [GLP84, Gri90].[1] The constraint satisfaction problem is solved by a *constraint satisfaction search* [RN95], which can be implemented as a backtracking tree search over a priority-queue holding, in our case, the candidate point pairs from the set $\mathcal{C}$. The queue imposes an ordering on $\mathcal{C}$, which can be used to arrive at a solution more quickly if external knowledge exists as to what constitutes a good candidate pair.

The path from the root of the tree to a node at depth m represents one particular order in which m elements are chosen from $\mathcal{C}$ to form a candidate set $\mathcal{S}$. As a result, the tree will contain *repeated states* in that it represents the same set $\mathcal{S}$ at multiple nodes in the tree, once for each of the $m!$ paths into $\mathcal{S}$. While maintaining every path makes the search impervious to any sensitivity to the order in which the elements that make up $\mathcal{S}$ are selected, it will lead to unnecessary search unless care is taken to avoid expanding a node if the the set it represents has already been tested. Checking for repeated states on the other hand requires its own resources both in the form of storage and computation [RN95]. The problem of repeated states is side stepped by structuring the search tree such that a set $\mathcal{S}$ is only represented at a single node. This means that only one path into each set is ever explored, which can introduce a critical dependence on the order in which the candidates are introduced. A simple algorithm growing skewed symmetric sets in this manner is outlined in Algorithm 4.1.

A reduction in the size of the detection problem is gained from an observation made by Ponce in the case of 2D skewed symmetry [Pon89], that also applies to the 3D case.

[1] It is interesting to compare the problem of matching object models to data features to the interest operator considered here in the context of object recognition. The major difference between the two is that the interest operator searches for matches only *within* a limited set of data features, whereas the model matching can involve searching for matches with an extremely large number of model sets.

$\mathcal{C} \leftarrow$ unordered candidate pairs $\{\mathbf{X}_p, \mathbf{X}_{p'}\}$ of points from an affine structure.
$n \leftarrow |\mathcal{C}|$.

`For` $i = 1 \ldots n-3$:
  `If` c_i is **very short** `then next` i.

  `For` $j = i+1 \ldots n-2$:
    `If` (c_i, c_j) have **any common endpoints**
    `or` (c_i, c_j) are **non-parallel**
    `then next` j.

    `For` $k = j+1 \ldots n-1$:
      `If` (c_i, c_j, c_k) have **any common endpoints**
      `or` (c_i, c_j, c_k) are **non-parallel**
      `or` (c_i, c_j, c_k) have **collinear midpoints**
      `then next` k.
      $S \leftarrow \{c_k, c_j, c_i\}$.

      `For` $l = k+1 \ldots n$:
        `If` (c_l, S) have **any common endpoints**
        `or` (c_l, S) are **non-parallel**
        `or` (c_l, S) have **collinear midpoints**
        `or` (c_l, S) have **non-coplanar midpoints**
        `then next` l.
        Insert c_l into S.
      `next` l.
      `If` size of $S \geq 4$ `then` **insert** S into *skew symmetric*.
    `next` k.
  `next` j.
`next` i.

Algorithm 4.1: Pseudo-code for a simple algorithm for growing skewed symmetric sets from a set of candidate point pairs.

He noted that since a test of parallelism cannot possibly succeed if the supporting lines of candidate pairs of points have widely different spatial directions, it would be possible to partition the set C of all point pairs into sets of similar directions, and solve the smaller detection problem of each partition in isolation. This comes at the price of eliminating

Given two affine views, compute the largest skewed symmetric point sets.

1. Detect interest points in two affine views I and J,
   (e.g. using the Harris interest operator [HS88]).
2. Match feature points between the two views (e.g. using [BZM95])
   to create a set of correspondences $\mathbf{w}_p = \begin{bmatrix}\mathbf{x}_p^I & \mathbf{x}_p^J\end{bmatrix}$.
3. Generate the set $\mathcal{C}$ of unordered pairs of correspondences $\{\mathbf{w}_p, \mathbf{w}_{p'}\}$.
4. Partion $\mathcal{C}$ into $k \times k$ bins according to the image orientation
   of the supporting lines $\mathbf{x}_p^I\mathbf{x}_{p'}^I$ and $\mathbf{x}_p^J\mathbf{x}_{p'}^J$.
5. For each bin of $\mathcal{C}$, starting with the most populated bin:
   6. Grow the largest skewed symmetric point set
      using Algorithm 4.1 and the tests
      from Table 3.12 (or Table 3.13) at a fixed value of α.

Algorithm 4.2: A summary of the deterministic symmetry detection algorithm, starting from two affine views.

any symmetric sets that extend across more than one partition unless multiple overlapping partitionings are used. The computational advantage is significant however, since the size of the search problem is exponential in the number of point pairs under consideration. The complete symmetry detection algorithm is summarized in Algorithm 4.2, starting from the images.

4.3.3 Choosing the Sharpness of Statistical Tests

The statistical tests of degenerate structure derived in Chapter 3 have only two free parameters, the covariance of the data points, assumed known or estimated as described in Section 3.4, and the probability of false rejection α, whose effect is the same regardless of the type of test or the number of points under consideration. To determine a suitable value for α one might reason that for a growth-path leading to the skewed symmetric set $\mathcal{S}_m$ to succeed with probability p, would require that the probability of correct detection for each of the m tests in the path multiplied together should give p. If all tests were set to equal probability of failure α, then this implies $(1-\alpha)^m \stackrel{!}{=} p$ giving $\alpha = 1 - p^{\frac{1}{m}}$. For a

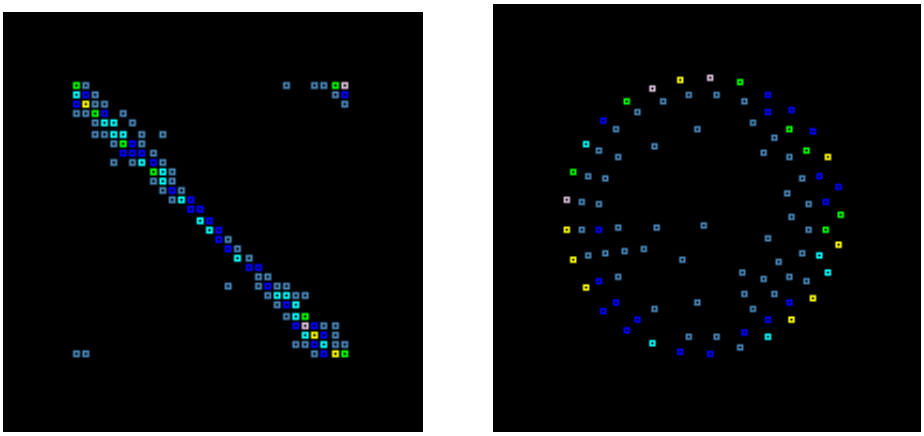
growth-path into a set of size $m = 6$ to have a probability of success $p = 0.95$ would then require each test along the path to be conducted with a rejection probability $\alpha = 0.0085$. Prescribing a rejection probability as low as this implies very generous bounds on the tests, which again means that for the purpose of pruning away incorrect sets these tests prove a blunt instrument.

We would be left with the above situation, if not for the fact that there are multiple paths into each set. At depth d in the search tree there is one node [2] leading to $\mathcal{S}_m$ for each d-chord subset $\mathcal{S}_d \subset \mathcal{S}_m$, of which there are $\binom{m}{d}$. These are in general not all independent, but the multitude of paths at every level except the deepest one, which only contains the set $\mathcal{S}_m$, makes it a reasonable to assume that the last test dominates the previous ones, and that the total probability of detection $P(\mathcal{S}_m)$ be closer to $1 - \alpha_m$, where α_d is the probability of rejection chosen for tests at depth d. Choosing a fixed $\alpha_d = \alpha$ would then result in all skewed symmetric sets be detected with roughly equal probability.

4.4 Implementation and Experiments

The symmetry detector was implemented as described in the previous section, using the statistical tests from Chapter 3. Two variations of the detector were implemented, one taking a 3D affine structure as input, and another applying constraints directly to point correspondences across two affine views. The two variants are identical in most respects other than the dimensionality of the statistical tests and the partitioning of the input. The set $\mathcal{C}$ of all possible point pairs from the P input points was partitioned according to direction of supporting lines as described in the previous section. In the case of affine structure the hemisphere of line directions was partitioned along latitude and longitude. The former was divided up at equal angular intervals, whereas the latter was divided at successively larger intervals at higher latitudes to keep the partitions similar in shape and size. This choice of partitioning over more uniform polygonal tesselations (see eg. [Hor86]) was made on grounds of simplicity of implementation. An example of this partitioning is shown in Figure 4.2(b). For the two-view variant, the set $\mathcal{C}$ of all possible pairs of inter image correspondences was partitioned along the direction of each of the two supporting image

[2] Not counting nodes representing repeated states, since each repeated state consists of exactly the same set of candidates.



(a) Orientation in two views. (b) Orientation in affine structure.

Figure 4.2: Partitioning of the set $\mathcal{C}$ of all pairs of point features according to the orientation of supporting lines. Shown are two partitionings: (a) according to 2D orientations of image lines in two views, and (b) according to 3D orientations in the recovered affine structure (shown mapped onto a hemisphere). Warmer colours indicate a higher concentration of point pairs. The data consists of interest points matched across the two views shown in Figure 4.3.

lines in the two views. This type of partitioning is not the best conceivable, however, since it is oblivious to rotations away from the camera, which manifest themselves as a foreshortening in the image rather than a rotation. Again the choice was made in favor of simplicity. An example of this partitioning is shown in Figure 4.2(a).

Software Implementation

The implementation was written in the C++ programming language, using proven software libraries, amongst which are the *LEDA* library [Näh93] for lists and matrices, the *LAPACK* library [ABB+99] for linear algebra, and *Numerical Recipes in C* routines [PTVP92] for special mathematical functions. The *X-Windows* graphical user interface and file handling was adopted from the *Horatio* vision libraries [McL94]. Where appropriate, interface class definitions were modeled on the *Image Understanding Environment* or *IUE* [KM94], which at the time implementation started was at an early stage of development. The IUE class definitions provided amongst other things high level abstractions for 2D and 3D geometrical objects, sets and other collections of objects, and an object oriented method of graphical output. These features made the coded algorithm clear and concise and free from actual

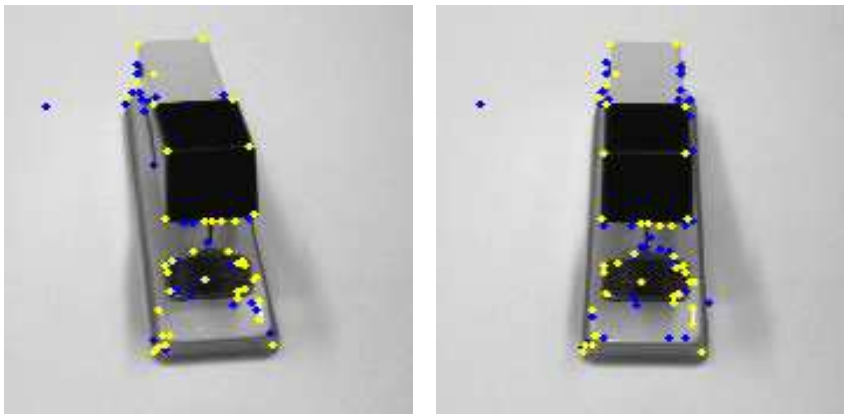


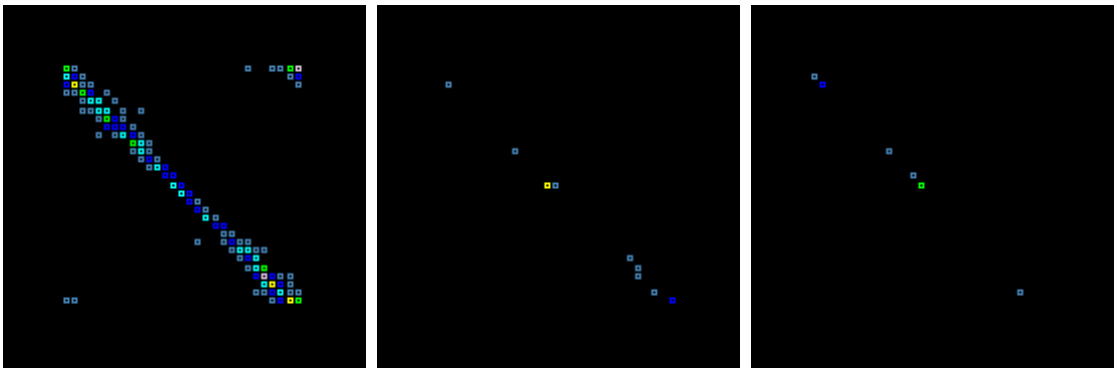
Figure 4.3: Two views of a simple scene obtained under affine viewing conditions. The strongest interest points extracted from each image are superimposed. Points automatically matched between the two views are shown in yellow and unmatched points in dark blue.

implementation detail. The resulting implementation turned out to be both robust and flexible. The benefit of the object oriented design was that major alterations to algorithms and experiments could be executed quickly and without complications, provided they drew on functionality existing within the environment. The drawback of writing the environment as well as the algorithms, was that small changes to algorithms could result in a large implementation effort, if they required that new functionality be implemented. The use of external libraries from different sources imposes some run-time overhead since data must be converted between the formats used by the different constituent libraries. Compared to large software environments, such as the official IUE implementation, however, compile and run times are very reasonable. The 20.000 lines of code[3] were compiled using GNU g++ under SunOS and Linux operating systems on Sparc and i386 architectures, respectively.

4.4.1 Automatic Detection from Interest Points

This section demonstrates fully automatic detection of skewed symmetry from image pairs, by means of interest points automatically detected and matched between the two views. The interest points were obtained using the Harris corner detector [HS88], which locates points in the image where the image intensity changes sharply in two directions. Harris

[3]Gross line count of source files excluding libraries.

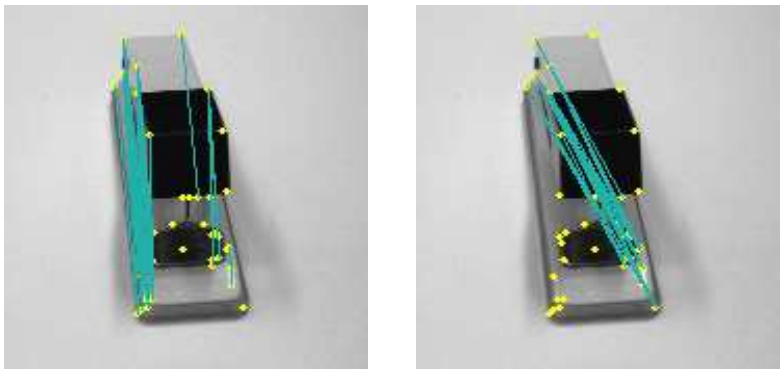


(a) Candidate point pairs. (b) Sets retained at $\alpha = 0.01$. (c) Sets retained at $\alpha = 0.1$.

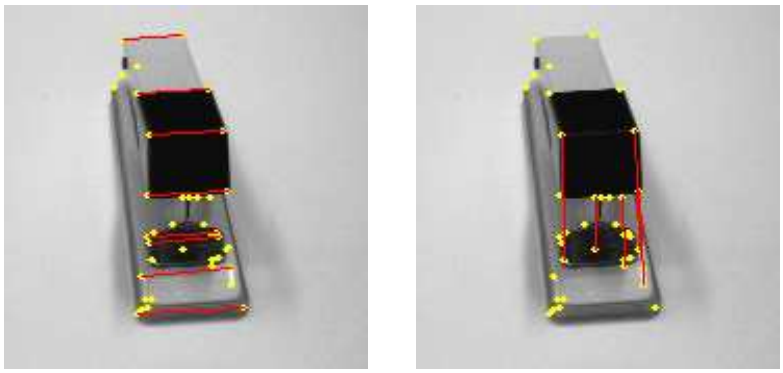
Figure 4.4: Distribution of point pairs across the two-view partitions show the histogram of image orientation of supporting lines in two views of a stapler (Figure 4.3). Shown is the distribution (a) before and (b)-(c) after the detection of skewed symmetry. Warmer colours indicate higher concentration of point pairs in (a) and larger detected skewed symmetric sets in (b)-(c).

corners have been found to stick rather well to scene points under changes in viewpoint (see [SMB00]). For that reason they are widely used for the computation of camera geometry and sparse scene structure from multiple views, and have also been put to use in object recognition [SM97] to select interest points on objects with a degree of repeatability. The name corner detector does not refer to the shape of objects in the scene, but rather to the shape of the image intensity function, which is affected by scene illumination and the reflectance function of surfaces as well as the shape of objects. As a result, one can only expect a partial correspondence between image and scene corners.

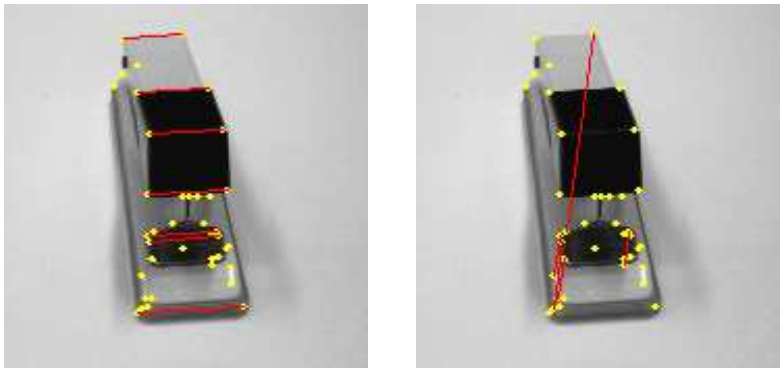
For this experiment we used pairs of images of simple scenes containing untextured bilaterally symmetric 3D objects, like the one shown in Figure 4.3. Interest points were extracted and matched automatically using implementations of the corner matcher by Beardsley et al. [BZM95], kindly provided by Paul Beardsley and Andrew Fitzgibbon. The location of corresponding points in the two views was then used to detect skewed symmetric sets using either (1) the image locations directly or (2) using affine structure computed from the same correspondences.



(a) Two largest partitions of candidate point pairs (47 and 37 pairs).

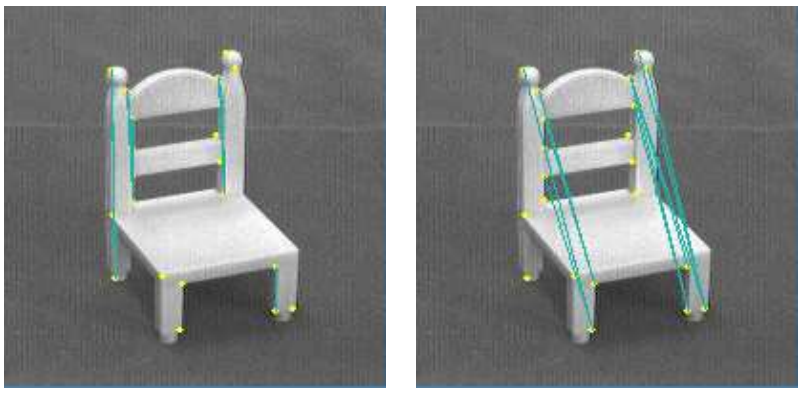


(b) Two largest retained sets at $\alpha = 0.01$ (8 and 5 chords).

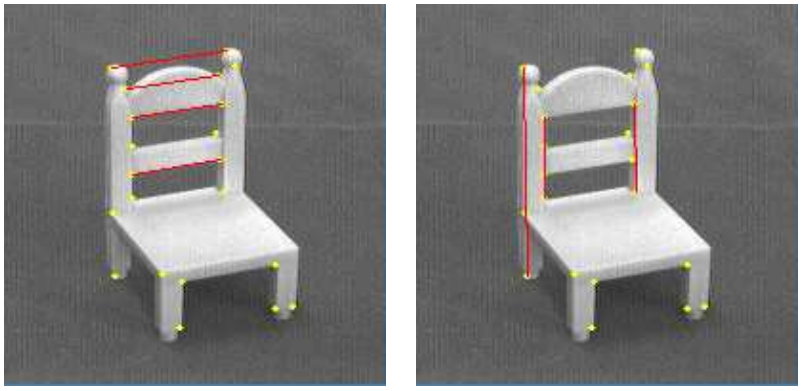


(c) Two largest retained sets at $\alpha = 0.1$ (7 and 5 chords).

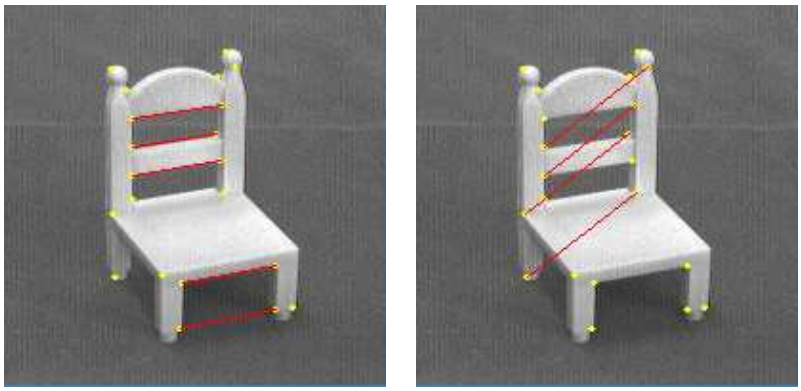
Figure 4.5: Results of a fully automatic detection of 3D skewed symmetry from Harris corner features in two views of a simple scene.



(a) Two largest partitions of candidate point pairs (14 and 11 pairs).

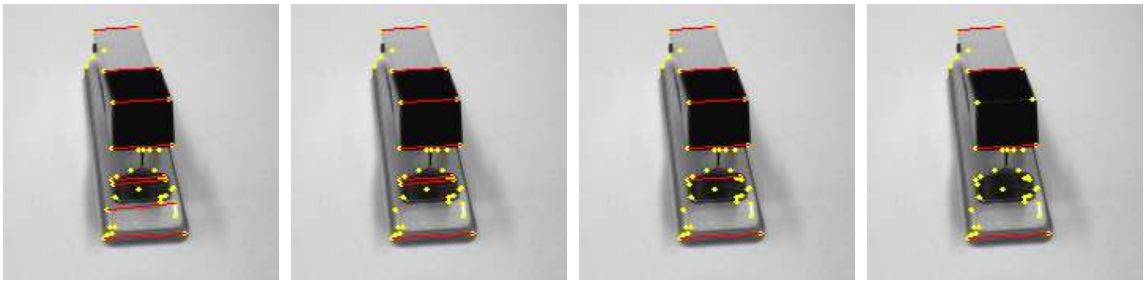


(b) Two largest retained sets at $\sigma = \widehat{\sigma} = 0.6, \alpha = 0.01$ (4 chords).



(c) Two largest retained sets at $\sigma = 0.7, \alpha = 0.01$ (5 and 4 chords).

Figure 4.6: Symmetric sets narrowly missed by the automatic detector. Chords at the front and back of the object (b) fail the test of parallelism at the MLE estimate $\sigma = \widehat{\sigma}$ as given by (3.29) (c) but are successfully detected at a slightly higher choice of σ.



(a) $\alpha = \frac{1}{100}$: 8 chords. (b) $\alpha = \frac{1}{8}$: 7 chords. (c) $\alpha = \frac{1}{4}$: 6 chords. (d) $\alpha = \frac{1}{2}$: 4 chords.

Figure 4.7: Chords missed as the rejection probability α is increased beyond reason. Shown is the largest detected set at specified values of the parameter.

Skewed Symmetry from Affine Views

The results of running the symmetry detector directly on the location of corresponding image features in two simple scenes are shown in Figure 4.5 and Figure 4.6 respectively. The set of all possible pairs of correspondences was partitioned using 12° intervals along the circle of orientations in each view. The results of the partitioning for the first scene is shown in Figure 4.4(a), with the contents of the two largest partitions shown in Figure 4.5(a). The standard deviation σ of the localization error of the interest operator was estimated from the affine reprojection error using the MLE (3.29) giving $\widehat{\sigma} = 0.50$ for the first scene and $\widehat{\sigma} = 0.60$ for the second. These values along with the rejection probability fixed at $\alpha = 0.01$ in accordance with the reasoning given in Section 4.3.3 are the only parameters that require setting.

From the first scene a total of 30 skew symmetric sets containing 4 or more chords were detected. They are distributed among 9 partitions, as shown in Figure 4.4(b). The largest set consists of 8 chords correctly revealing the main bilateral symmetry of the object as shown in Figure 4.5(b). This is the only set containing more than 5 chords. Of the 5 chord sets only the set shown is disjoint from the 8 chord set. Running the detector with a more restrictive $\alpha = 0.1$ resulted in the largest skew symmetric set containing one less chord, or a total of seven, with only one other set containing as many as 5 chords. These are shown in Figure 4.5(c). The total number of skewed symmetric sets dropped to 25 in the 6 partitions shown in Figure 4.4(c). The effect of increasing α beyond this is shown in Figure 4.7.

The second scene generated only half as many point correspondences as the first scene,

21 as compared to 40. Run with the initial set of parameters, $\widehat{\sigma} = 0.60$, and $\alpha = 0.01$ returned only 6 sets none of which contained more than 4 chords. The detected sets of which two are shown in Figure 4.6(b), were confined to the back of the object, none spanning both the front and the back. Upon examination it turned out that the chords at the front and back of the object failed the test of set-wise parallelism. Increasing the estimated value of σ to slightly to 0.70 was sufficient to detect the 5 chord set shown in Figure 4.6(c). Apart from the possibility that this be a chance occurrence, this could also be caused by movement of features due to self-occlusion or a deviation from affine viewing conditions. It may also be that the MLE is on the optimistic side in these conditions.

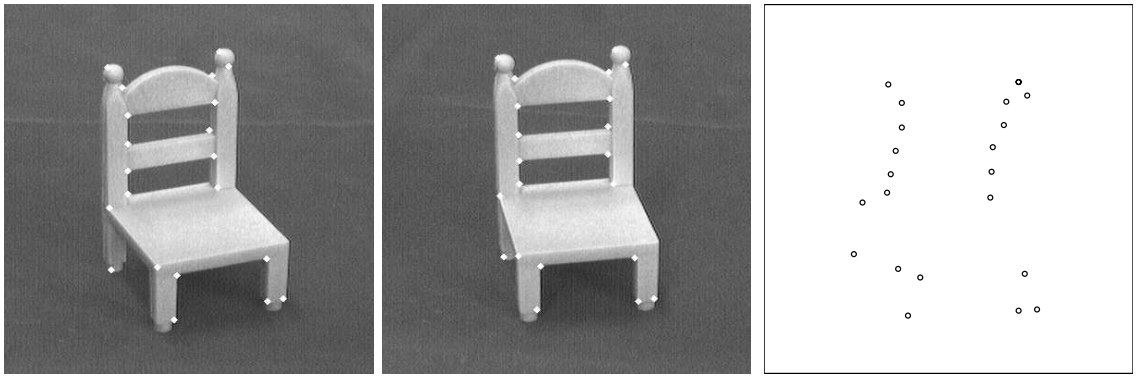
The total number of nodes expanded in the search for symmetry at $\alpha = 0.01$ was in 9534 and 463 for the two scenes. Running time was less than one second on a 733MHz Pentium III. This includes all steps of the symmetry detection algorithm from established point correspondences, but not the initial detection and matching of interested points between the two views. This corresponds to steps 3 through 6 of Algorithm 4.2 on page 81.

Discusson on the Choice of Parameters Values

The optimal tests for degenerate structure derived in the last chapter for use in the symmetry detection algorithm are parametrized in two parameters, whose values need to be given. These are the standard deviation of the feature location error σ, and the acceptable false rejection probability α.

In the experiments in this chapter σ was set at or slightly above the estimated affine reprojection error, given by the MLE $\widehat{\sigma}$ from (3.29). A value of σ above the reprojection error (3.29) may be necessary to accommodate inaccuracies in the affine approximation. Furthermore, the reprojection error $\widehat{\sigma}$ may not fully capture the motion of interest points on the surface of scene objects, especially when this motion is predominantly along epipolar lines. An example of this is the motion of the *apparent contour* of a smooth object surface. A too generous a value for σ will reduce the effectiveness of the pruning, necessary to contain the combinatorics of the search.

Section 4.3.3 reasoned that the value chosen for the significance level α of each *individual* test should be chosen to represent the acceptable failure probability of detecting the



(a) Automatic correspondences across two affine views (b) Recovered structure

Figure 4.8: Affine structure of a simple scene recovered from the location of interest points matched automatically across two views as described in the text.

largest symmetric set. The experimental results appear to support that choice. Notably, the proportion of chords missed from the largest symmetric set appears to roughly follow the specified value of α, as borne out by the results shown in Figure 4.7.

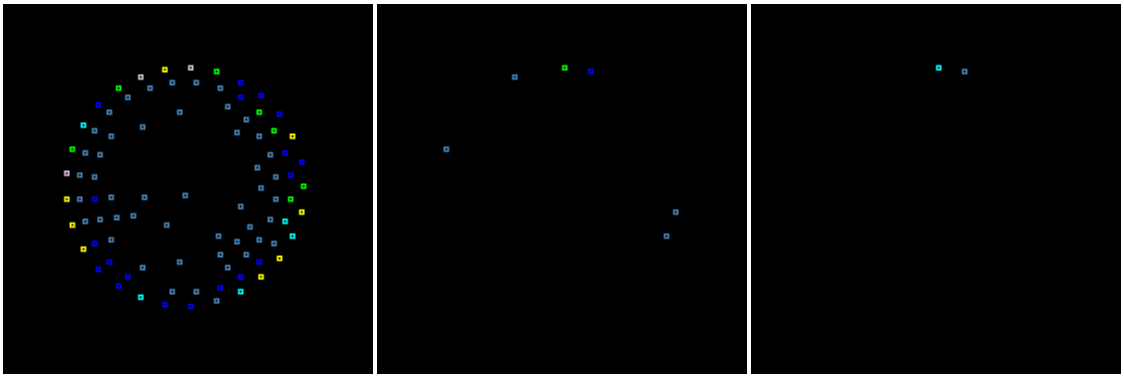
Skewed Symmetry from Affine Structure

The affine structure of the points matched across the two views was obtained using the *isotropic ML factorization* (3.31), as shown in Figure 4.8. This estimator has the property that the probability distribution of each recovered 3D point is isotropic with equal variance, as described by its covariance matrix (3.32). Consequently, the distribution of the orientation of a pair of reconstructed points mapped on a hemisphere is also isotropic, with a variance that depends only on the distance between the points. This means that a uniform partitioning of the hemisphere is viable, while the probability of a set of parallel point pairs being split up in the process is smaller for the longer pairs.

The set of all pairs of points from the affine structure was partitioned using a tessellation of the hemisphere with a 12° mesh. The distribution of the pairs from the first scene is shown in Figure 4.9(a), while the content of the largest partitions from the second scene is shown in Figure 4.10(a). The tests of degenerate structure were conducted with the same parameter values as in the two view case, ie. a covariance matrix (3.32)

$$\Sigma_{\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i} = \widehat{\sigma}^2 \cdot \mathsf{I}_r,$$

with $\widehat{\sigma} = 0.60$, and a fixed rejection probability $\alpha = 0.01$. Again these values resulted



(a) Candidate point pairs. (b) Sets retained at $\alpha = 0.01$. (c) Sets retained at $\alpha = 0.1$.

Figure 4.9: Distribution of point pairs across the partitioned hemisphere of 3D orientations in affine structure recovered from two views of a stapler (Figure 4.3). Shown is the distribution (a) before and (b)-(c) after detection of skewed symmetry. Warmer colours indicate higher concentration of point pairs in (a) and larger detected skewed symmetric sets in (b)-(c).

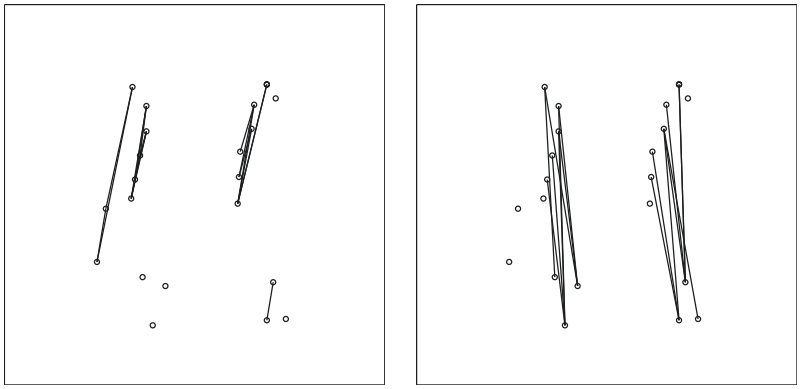
in only 4 chord sets being detected, while increasing the estimated standard deviation to $\widehat{\sigma} = 0.70$, returned the same 5 chord set as in the case of the two view sets. The largest skewed symmetric sets detected from the second scene are shown in Figure 4.10(b).

The total number of nodes expanded in the search for symmetry at $\alpha = 0.01$ was 14082 and 1216 for the two scenes. Despite this the search ran quicker than in the two view case where fewer nodes were expanded. This does not come as a surprise, for the statistical tests performed at each node involve smaller matrices in the case of affine structure.

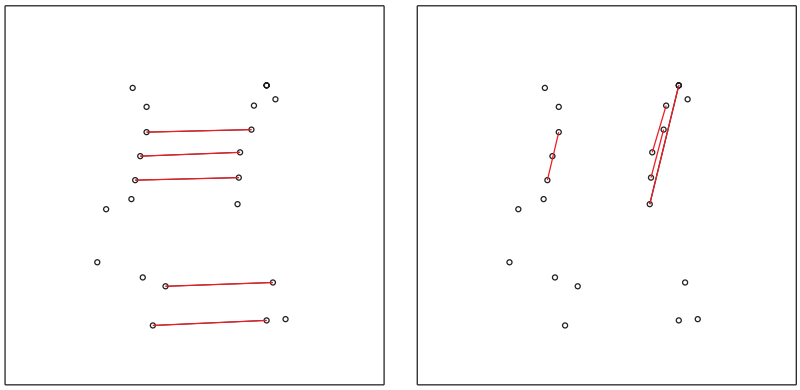
4.4.2 Performance Characterization of Detector in Noise

This section attempts to characterize the performance of the symmetry detector in the presence of measurement noise. The experiment evaluates the effects of (1) different noise levels, (2) varying the rejection probability parameter α, and (3) enforcing the non-collinearity condition on the basis and a minimum segment length. Control over the noise level was maintained by using synthetic data, thus allowing for comparison between different levels of noise. The design of the experiment was largely based on Haralick's recommendations [Har94a, KJPH93].

The experiment was conducted on two views of a synthetic point set. Each pair of views depicted the affine projection of 40 scene points of which 20 constituted bilaterally



(a) Two largest partitions of candidate point pairs (19 and 15 pairs).



(b) Two largest retained sets at $\sigma = 0.7, \alpha = 0.01$ (5 and 4 chords).

Figure 4.10: Results from detecting 3D skewed symmetry in the affine structure extracted from the two views shown in Figure 4.8.

symmetric point pairs. The shape of each point set, the camera viewpoints, and the aspect ratio were all treated as *nuisance variables* in the terminology of [Har94a] and were assigned different randomly chosen values for each pair of views. The size of the object in each image was typically between 100 and 300 pixels in diameter. To the point locations in each view we added isotropic Gaussian noise of standard deviation $\sigma = 0.5$, 1.0, or 2.0 pixels. For each noise level we ran the symmetry detector with several different settings of the user selected rejection probability α. For each value of σ and α we determined the average number of chords missed by the detector, based on 20 image pairs. We repeated this test for similarly generated reference sets of 40 points, none of which had symmetric counterparts. This gave the size of the largest false detected set of symmetric chords.

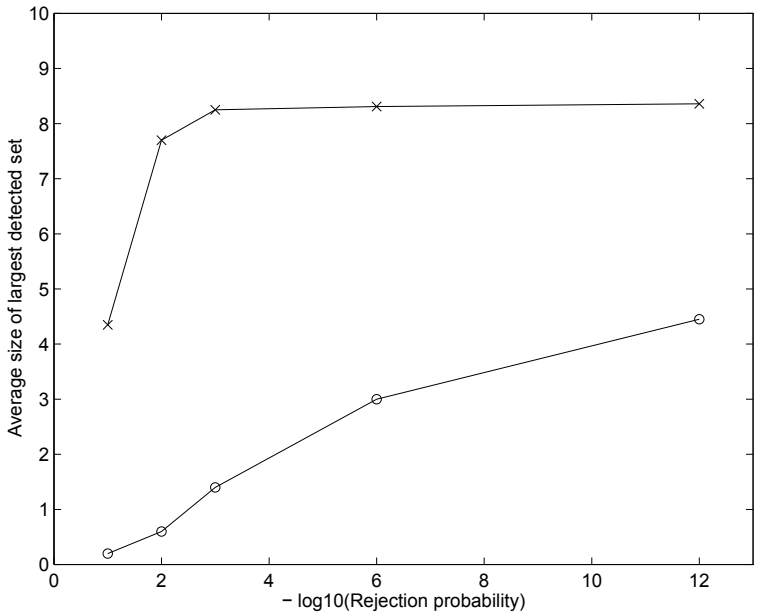


Figure 4.11: The effect of varying the rejection probability α on the size of the largest detected set. The two curves show average results over 40 point data sets, (top) with and (bottom) without an embedded 10 chord symmetric set. The curves depicted here were obtained at an image noise level of $\sigma = 1.0$ pixels, with non-collinearity and minimum segment length not enforced.

Effects of Noise and the Rejection Probability Parameter

The effect of varying the rejection probability α is shown in Figure 4.11. The figure shows the average size of the largest detected set in the presence and absence of symmetry at a specific noise level $\sigma = 1.0$, as the probability α is varied from 0.1 through 10^{-12}. While the smallest detectable set contains four chords, the average size can be less if sets are not detected in every sample image. To summarize the performance at different levels of measurement noise σ, we produced "operating curves" plotting the size of the largest falsely detected set against the number of missed chords, as shown in Figure 4.12. The term *operating curve* is borrowed from statistical detection theory [VT71] where *receiver operating curves* (*ROC*) characterize the performance of binary detectors, by plotting the false detection rate against the rate of missed detection. Figure 4.12 shows that the average number of missed or falsely detected chords goes up with the noise level. Choosing an operating point on each curve as the point where the average number of missed chords equals the average size of false detected sets (as indicated by the intersection with the

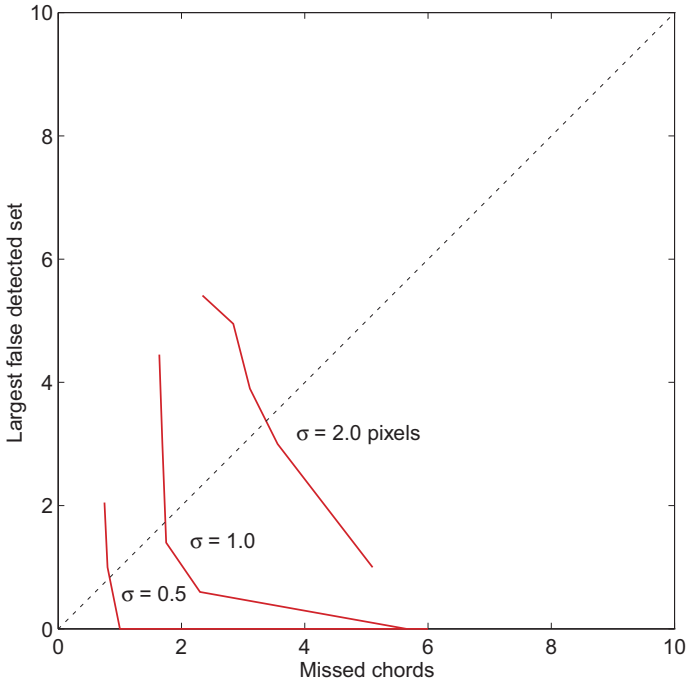


Figure 4.12: Operating curves in terms of the size of the largest falsely detected set and the number of missed chords. The figure summarizes the performance of the symmetry detector at different levels of image noise. Each operating curve summarizes a pair of detection curves like the ones seen in Figure 4.11. Shown here are operating curves corresponding to the case where non-collinearity and minimum segment length are not enforced.

dotted line in Figure 4.12) we see that the error grows from 0.8 chords at $\sigma = 0.5$, to 1.8 chords at $\sigma = 1.0$, and 3.5 chords at $\sigma = 2.0$ pixels. The growth in error is approximately linear with increased noise level.

Effects of Enforcing Non-Collinearity and Minimum Segment Length

In an attempt to quantify the effect of enforcing a minimum segment length and enforcing non-collinearity of midpoints, as implemented in Algorithm 4.1, we compared operating curves for different parameter choices in the manner suggested in [KJPH93]. The parameter settings considered were (1) enforce a minimum segment length of 0, 1σ, 4σ, and 10σ (2) enforce non-collinearity of midpoints. We proceeded by first producing operating curves for each choice of parameters like the example in Figure 4.12. We then chose an operating point on each curve as the point where the average number of missed chords equals the average size of false detected sets. Noting the average error in terms of set size at the

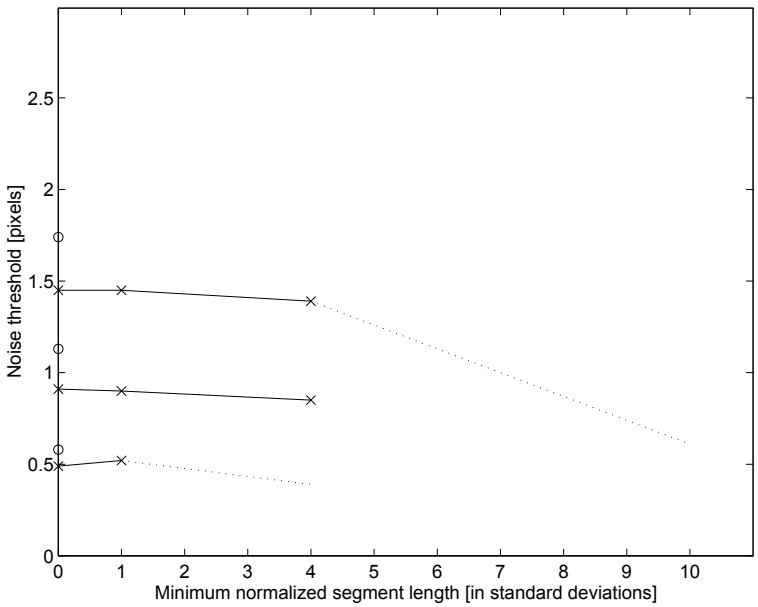


Figure 4.13: The effect of enforcing non-collinearity and discarding line segments of a certain minimum length on the noise thresholds (from bottom) σ_{T1}, σ_{T2}, and σ_{T3}. The noise thresholds are defined as the noise level at which the average size of false detected sets and the average number of missing chords both equal 1, 2, and 3 respectively. The dotted lines indicate less reliable data. For comparison the 'o'-s on the vertical axis show the slightly higher noise thresholds when non-collinearity is not enforced.

operating point for each noise level, we can determine the relation between noise level and number of errors for different choices of parameters. In particular we can determine the *noise thresholds* σ_{T1}, σ_{T2} and σ_{T3}, defined as the noise level at which the average number of errors equals 1, 2 and 3 respectively. In Fig. 4.13 we show how the three noise thresholds vary with the choice of minimum length of line segments at the chosen operating point. From the figure we conclude that varying the minimum segment length from zero to 4σ produces about 5% drop in the noise threshold at the chosen operating point. Increasing the noise threshold beyond that was seen to cause a larger drop, although the results were less reliable.

Introducing the non-collinearity contraint has a similar, but larger reduction in the noise threshold values, as tabulated in Table 4.2. The fact that the noise threshold values decrease when constraints are enforced more tightly implies that the effect of enforcing the constraints is causing a loss of chords. At the selected operating point and for the data sets considered here the benefit of enforcing the constraints is outweighed by the loss of

non-collinearity	noise threshold [pixels]		
	1 error	2 errors	3 errors
not enforced	0.6	1.2	1.7
enforced	0.5	0.9	1.4

Table 4.2: Comparison of the noise thresholds for one, two, and three errors, when non-collinearity is enforced or not enforced.

chords, if only marginally. The main conclusion drawn from Table 4.2, however, is that the performance of Algorithm 4.1 is seen to degrade gracefully with increased noise, as measured by the near-linear increase between the three noise thresholds over a range of parameter settings considered in this experiment. The observed degradation in performance is the result of increased perturbation of the geometric structure. It should be noted here, that the experiment was conducted with a known variance of the image noise. The use of an inappropriate value for the image variance may lead to different results.

Computational Cost

The average number of nodes in the search tree expanded during the search that is the number of sets $\mathcal{S}$ from $\mathrm{Pow}(\mathcal{C})$ examined for skewed symmetry, was evaluated at each of the threshold points in Figure 4.13. This number turned out to be close to constant around 4,600, irrespective of rejection probability α and noise level σ. This fact is likely governed by the common operating point. Outside this operating point, however, the average number of nodes expanded varied from approx. 1,400 through 10,000 as values for α and σ were varied throughout this experiment. The total size of the search space prior to partitioning, is given by the cardinality of $\mathrm{Pow}(\mathcal{C})$ as 10^{234}.

4.5 Searching by Random Sampling

It was seen in Section 4.3 that the classical constraint satisfaction search introduces inefficiencies in the form of repeated states, that are only eliminated at the cost of considerable implementation effort and storage. We will now look at an alternative search strategy based on random sampling.

Constraint satisfaction based on random sampling was introduced by Fischler and Bolles [FB81] in their *Random Sampling Consensus* (RANSAC) algorithm. In the last few years, starting with the work of Torr and Murray [TM97], RANSAC has found renewed use in computer vision as a robust estimator of geometric constraints. Unlike many common robust estimators (see [TM97]), RANSAC does *not* require that the set of inliers be a considerable proportion of the total data set. For this reason RANSAC can also be applied to search problems in which the proportion of inliers is very low, as can be the case in symmetry detection, although this comes at the price of a large number of samples. In this section we will consider the application of RANSAC to the search for symmetry, and bring to light certain advantageous properties of that approach.

Search Effort in Random Sampling

When drawing a sample of size p from a data set consisting of n point pairs, the probability of the whole sample belonging to a given subset of size k, such as a particular k-chord symmetric set, is given by

$$P_n = \frac{\binom{k}{p}}{\binom{n}{p}} = \frac{k}{n} \cdot \frac{k-1}{n-1} \cdot \ \cdots \ \cdot \frac{k-p+1}{n-p+1}, \tag{4.1}$$

assuming that each point pair is drawn without replacement. The probability of selecting at least one such sample in m trials is

$$\gamma = 1 - (1 - P_n)^m. \tag{4.2}$$

The number of samples m needed to to select any k-chord set with a specified probability γ can therefore be computed as

$$m = \frac{\log(1-\gamma)}{\log(1-P_n)}. \tag{4.3}$$

This should be compared to the total size of the search space, which is the total number of distinct subsets of size p,

$$c \stackrel{\text{def}}{=} \left(\begin{array}{c} n \\ p \end{array} \right), \tag{4.4}$$

all of which would be examined in an exhaustive algorithm. In Figure 4.14 the probability of selection γ as given by (4.2) is plotted as a function of the number of samples m, for

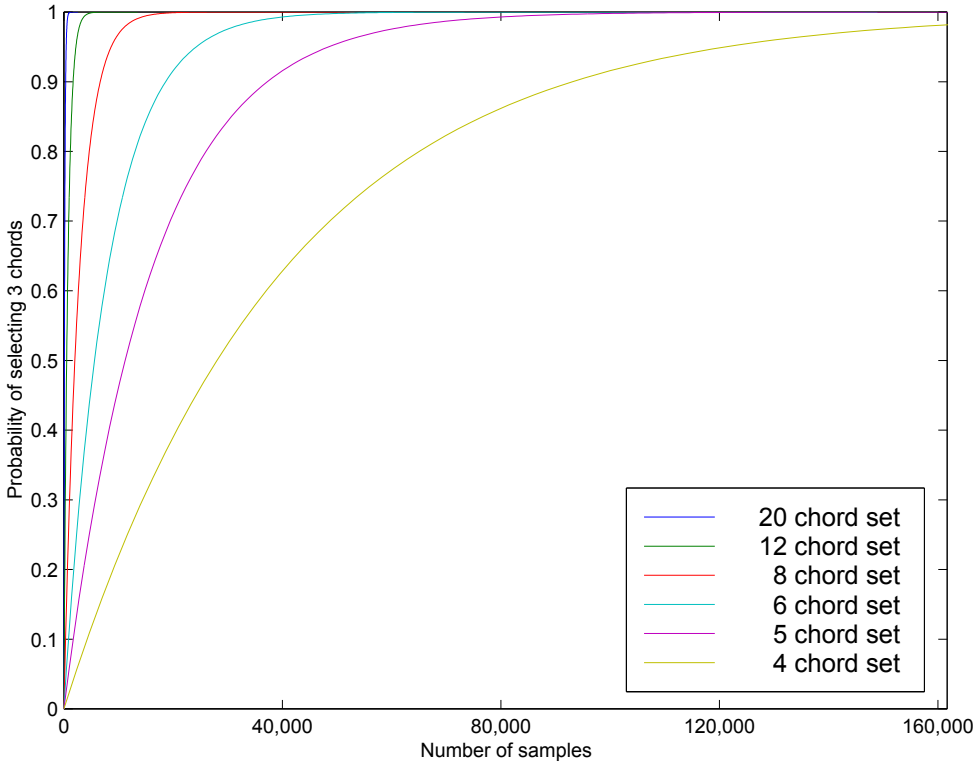


Figure 4.14: Probability of selecting at least one sample consisting of three point pairs from a k-chord symmetric set contained in a set of 100 point pairs, as a function of the number of samples. An exhaustive search would involve examining all $161,700$ distinct triples of point pairs.

the case of repeatedly selecting $p = 3$ point pairs out of a set of $n = 100$ pairs. The extent of the x-axis equals the size of the search space as given by (4.4). It is seen that a high probability of selection is reached long before the whole space is sampled.

When (4.3) is used to determine the sample size of the random search (for a specified γ and k), the search effort increases with the size n of the data set at the same *rate* as an exhaustive search, ie. as $O(\binom{n}{p}) = O(n^p)$. This is readily seen by expanding the denominator of (4.3)

$$m = \frac{\log(1-\gamma)}{\log(1-P_n)} = \frac{\log(1-\gamma)}{-P_n - O\left(P_n^2\right)} \tag{4.5}$$

and substituting (4.1) for P_n. It follows that the *relative* search effort $\frac{m}{c}$ of random sampling compared to exhaustive search in large data sets is independent of n, and has the asymptotic

value

$$\boxed{\lim_{n\to\infty} \frac{m}{c} = \frac{-\log(1-\gamma)}{\binom{k}{p}}.} \tag{4.6}$$

Equation (4.6) encapsulates the flexibility offered by random sampling, showing how the search effort can be varied through the choice of the selection probabilty γ for a given set size k. For the proposed task of repeatedly selecting $p = 3$ point pairs to form a basis for a candidate plane of symmetry this relative effort is

$$\lim_{n\to\infty} \frac{m}{c} = \frac{-\log(1-\gamma)}{\frac{1}{6}k(k-1)(k-2)}. \tag{4.7}$$

The relative search effort $\frac{m}{c}$ required in order to select at least one 3-chord basis from a k-chord symmetric set is tabulated in Table 4.3 for select values of the specified probability of success γ. The required number of samples m is given as a percentage of the total size c of the search space. To allow each candidate pair to be tested as a part of many different bases, each pair must be returned to the data set prior to selecting the next sample. Unless a record is kept of every sample selected, one can never be certain that the whole search space has been searched. An extreme requirement of selecting very small sets with a very high probability may therefore lead to the number of samples required exceeding the size of the search space, as is indeed observed in one case in Table 4.3.

no. of chords	selection probability γ		
	0.90	0.95	0.99
4	57.56 %	74.89 %	115.13 %
5	23.03 %	29.96 %	46.05 %
6	11.51 %	14.98 %	23.03 %
8	4.11 %	5.35 %	8.22 %
12	1.05 %	1.36 %	2.09 %
20	0.20 %	0.26 %	0.40 %

Table 4.3: The relative search effort $\frac{m}{c}$ required in order to select at least one 3-chord basis from a k-chord symmetric set with a given probability γ.

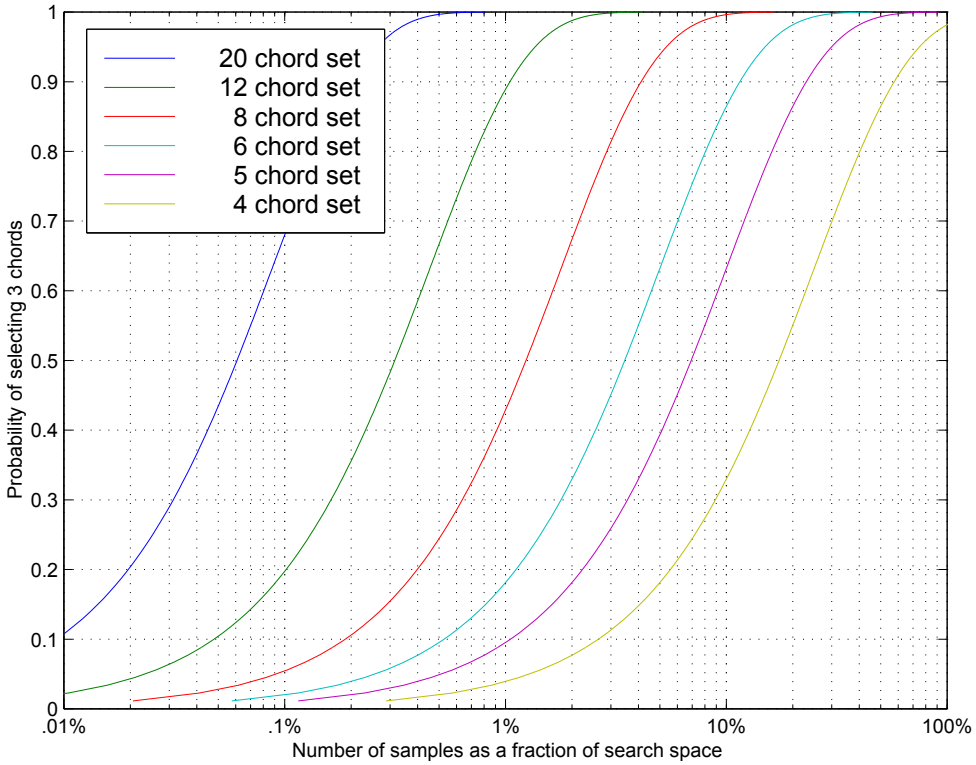


Figure 4.15: Probability of selecting 3 chords from a large set of point pairs, for different sizes of the symmetry set, shown as a function of the number of samples.

Selecting the Sample Size

In practice the size of the largest symmetry set within an unknown data set will not be known beforehand. The search must therefore be terminated at some point, either after a pre-specified number of samples or based on the size of the largest detected set. At any time during the search, as given by the number of samples m already examined, the probability γ that a basis from a symmetric set of a given size k has been selected can be determined using (4.6). That this probability γ increases rapidly with the the number of chords k in the symmetric set is readily seen from Figure 4.15. The effect is that *large symmetric sets have a higher probability of being selected early* on in the search, and even if the search is terminated early, a more sizeable proportion of the larger sets will have been recovered than of the smaller sets. This works to our advantage if we are mostly interested in the largest sets. Equally useful is the property of sampling randomly that even in the case when no symmetric set has been recovered in m samples, we can break off the search

no. of chords	proportion of search space sampled					
	0.3%	1.4%	5.3%	15%	30%	75%
4	0.011	0.053	0.193	0.451	0.698	<u>0.950</u>
5	0.026	0.127	0.414	0.776	<u>0.950</u>	0.999
6	0.051	0.238	0.657	<u>0.950</u>	0.998	1.000
8	0.137	0.534	<u>0.950</u>	1.000	1.000	1.000
12	0.440	<u>0.950</u>	1.000	1.000	1.000	1.000
20	<u>0.950</u>	1.000	1.000	1.000	1.000	1.000

Table 4.4: The probability γ of selecting 3 chords belonging to any one particular symmetric set at least once for different sample sizes and different sized symmetric sets. The sample sizes are chosen such as to correspond to a selection probability $\gamma = 0.95$ for the particular sizes of symmetric sets in Table 4.3.

knowing that the probability of having missed a k-chord set is not greater than $1 - \gamma$.

To determine an appropriate value for the number of samples m, one can specify a selection probability for one particular set size. For example one could specify that all 8-chord sets be selected with probability 0.95. From Table 4.3 we read that this means that the number of samples m should be 5.35% of the size of the search space or $\frac{5.35}{100}c$. Having decided on that number of samples, we can read the the probability of selection for different sized symmetric sets from the third column on the right hand side of Table 4.4. It should be noted that the probability of *selection* γ considered here only refers to the search. It differs from the final probability of *detection*, which also depends on the bounds used in determining the set of inliers to the basis.

4.6 Implementation and Results

4.6.1 Implementation

The complete randomized symmetry detection process is summarized in Algorithm 4.4, starting from the images. This process makes repeated use of the RANSAC algorithm to find the largest degenerate subset of points of a particular rank, both for the symmetry search and for the initial frame-to-frame matching. The search for the largest degenerate subset is summarized in Algorithm 4.3. Together, the two algorithms describe a simple

Given a set of point correspondences in two or more affine views, find the largest rank r subset.

Repeat for m samples:

1. Randomly pick $r+1$ correspondences to act as a basis for an r-dimensional subspace in the measurement space.
2. Compute the SVD of the measurement matrix (3.18) containing only the $r+1$ basis vectors.
3. Test the basis for degeneracy by applying the dimensionality test (3.47) to the singular values with $r_0 = r - 1$. Return to step 1, if not rejected.
4. Compute the ML estimators for the restriction matrix $\widehat{\mathsf{B}}_q$ and $\widehat{\mathbf{b}}_q$ using (7.15) and (7.16).
5. Repeat for each of the given correspondences $\mathbf{w}_i$:
   (a) Compute the Mahalanobis distance $d_i^2 = (\widehat{\mathsf{B}}_q\mathbf{w}_i - \widehat{\mathbf{b}}_q)^T(\widehat{\mathsf{B}}_q\mathbf{w}_i - \widehat{\mathbf{b}}_q)$ to the subspace spanned by the basis.
   (b) Reject as outlier if $d_i^2 > \chi^2_{q,\,\alpha}$
6. Retain the set of inliers if the number of inliers exceeds all previous samples, discarding any previously retained set. Repeat from step 1.

Algorithm 4.3: A method finding the largest rank r subset in a set of correspondences using RANSAC.

process, that is easily implemented.

The randomized search by RANSAC in Algorithm 4.3 differs in many important respects from the deterministic traversal of the search space presented in Algorithm 4.1 on page 80. There are some notable similarities however. Both search algorithms repeatedly select a minimal set of candidates to form a basis for a subspace of the desired dimension. Both algorithms apply the tests of degeneracy, developed in the last chapter, to this basis to ensure that it spans the subspace. In the randomized search, each basis is selected independently of the previous one, whereas in Algorithm 4.1, successive bases are highly correlated. Algorithm 4.1 goes on to grow an inlier set from the basis by successive testing

Given two affine views, compute the largest skewed symmetric point set $\mathcal{M}$.

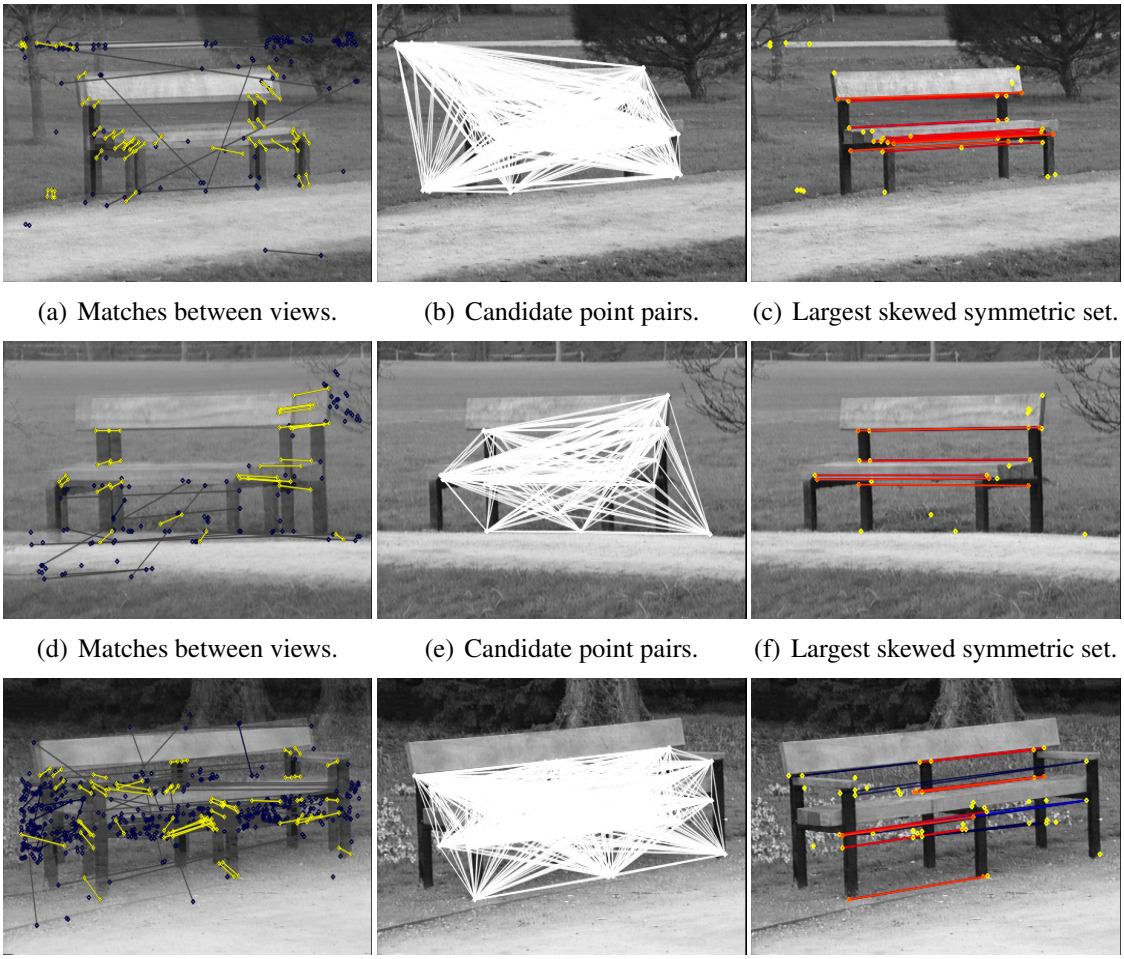
1. Detect interest points in two affine views I and J, (e.g. using the Harris interest operator [HS88]).
2. Create a set of candidate matches between interest points in view I and J.
3. Find the largest set of matches $\mathbf{w}_p = \left[\mathbf{x}_p^I \; \mathbf{x}_p^J\right]^T$ consistent with affine epipolar geometry (7.13), by applying Algorithm 4.3 to the candidate set with rank $r = 3$.
4. Create a set $\mathcal{C} = \{\mathbf{w}_p, \mathbf{w}_{p'}\}$ of unordered pairs of correspondences.
5. Find the largest subset $\mathcal{P} \subset \mathcal{C}$ of parallel chords (2.22), by applying Algorithm 4.3 to the set $\mathcal{C}_- = \{\mathbf{w}_p - \mathbf{w}_{p'}\}$ with rank $r = 1$.
6. Find the largest non-coplanar subset $\mathcal{M} \subset \mathcal{P}$ with coplanar midpoints (2.23), by applying Algorithm 4.3 to the set $\mathcal{P}_+ = \{\mathbf{w}_p + \mathbf{w}_{p'}\}$ with rank $r = 2$, testing (2.24) in step 3.
7. Retain $\mathcal{M}$ if it has the highest cardinality so far.
8. Remove $\mathcal{P}$ from $\mathcal{C}$ and repeat from step 5.

Algorithm 4.4: A randomized process for detecting 3D skewed symmetry, starting from two (or more) affine views.

and inclusions of elements from the candidate set, while RANSAC tests each remaining candidate against the basis without inclusion. Both algorithms use the cardinality of the largest detected inlier set as the final selection criterion.

4.6.2 Results

Examples of running the randomized symmetry detection process in Algorithm 4.3 are shown in Figures 4.16-4.18. The data are digitized camcorder images of an outdoor scene with a resolution of 512 by 272 pixels. The viewing conditions are assumed weak perspective. The entire process runs automatically from a pair of images. The point features are detected using the Harris interest operator. The matching of features between views is done enforcing the affine epipolar constraint as described in Algorithm 4.4, and detailed in



(a) Matches between views. (b) Candidate point pairs. (c) Largest skewed symmetric set.

(d) Matches between views. (e) Candidate point pairs. (f) Largest skewed symmetric set.

(g) Matches between views. (h) Candidate point pairs. (i) Largest skewed symmetric set.

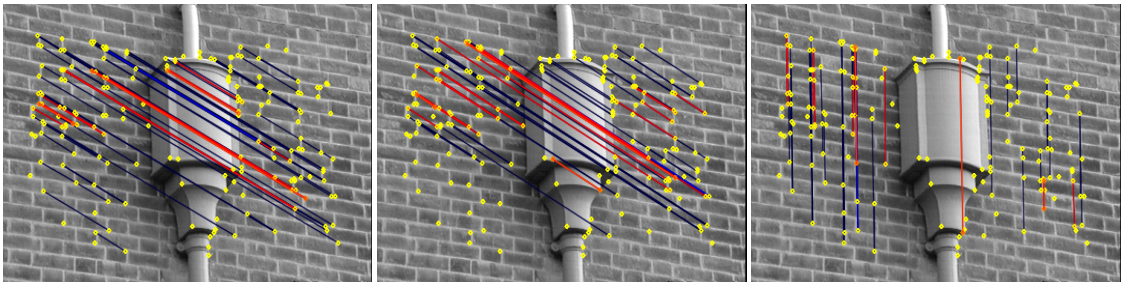
Figure 4.16: The largest 3D skewed symmetric point set detected by the randomized search method in Algorithm 4.4.

Chapter 7. The test for degeneracy in Algorithm 4.3 ensures that the set of candidate points is non-coplanar.

The symmetry detection is run with estimated image noise level of $\sigma = 0.6$ pixels, which is the mean reprojection error of the inlier set to the affine epipolar constraint. The false rejection probability was selected between $\alpha = 0.01$ and $\alpha = 0.05$. As expected, occasional outliers, e.g. non-parallel chords, were observed at the lower value. At the higher value inclusion in the set was clearly more dependent on the selected basis. The degeneracy threshold was set at 3 times the quantile of the chi-square distribution to enforce non-coplanarity of the detected set. The matching was done from an initial exhaustive set



(a) Matches between views. (b) Candidate point pairs. (c) Largest skewed symmetric set.



(d) Further skewed symmetric sets.

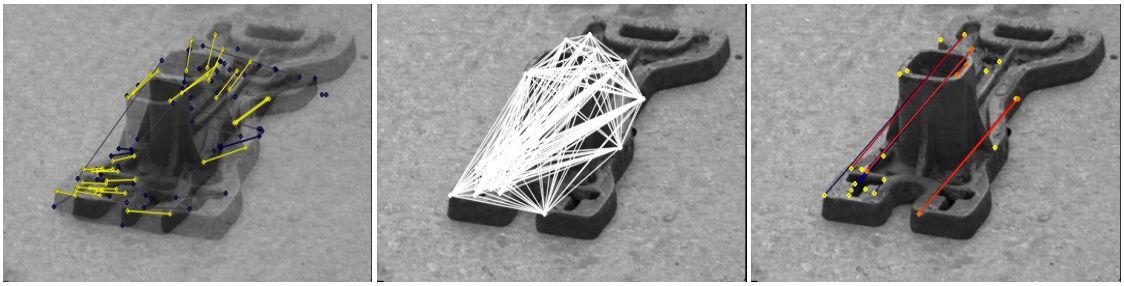
Figure 4.17: An example of the influence of a regular background texture on the symmetry detection. The detected 3D skewed symmetric point sets (red) are dominated by the background texture.

of correspondences between points in the joint image. Line segments shorter than 30 pixels were excluded the set in favor of a smaller initial set.

For the objects containing a single distinctive plane of bilateral symmetry seen in Figure 4.16 (c), (f), the symmetric point set is invariably returned as the largest skewed symmetric set. In the object in Figure 4.16 (i), which contains more than one plane of symmetry, the location of the detected plane of symmetry varies with the detected set of parallel line segments. As seen in the figure, the skewed symmetric set with the highest cardinality does not necessarily coincide with a plane of bilateral symmetry.

Influence of Texture

Figure 4.17 shows an apparently bilaterally symmetric object mounted on a background with a regular texture. The vast majority of the detected features lies on the textured background. The regularity of the texture gives rise to a range of directions of line segments, each populated by a number of segments. A single segment off the plane, i.e. on the foreground object, suffices to define a valid 3D skewed symmetric set. As a result we observe



(a) Matches between views. (b) Candidate point pairs. (c) Largest skewed symmetric set.



(d) Matches between views. (e) Candidate point pairs. (f) No 3D skewed symmetric set.

Figure 4.18: Examples of insufficient feature data for symmetry detection. In the images in the top row (a-c), not enough symmetrically arranged features are matched between the two views (a) to allow the main plane of object symmetry to be detected. In the images on the bottom row (d-f), the detected features are confined to a single plane so that the 3D symmetric shape of the object goes undetected.

large number of detected sets, dominated by segments on the planar background, with only minimal evidence of a 3D structure.

Influence of the Feature Detection and Matching

The symmetry detection algorithm takes as input matched point features across two (or more) views. Figure 4.18 shows to cases, where the symmetry detection fails due to lack of input data. In figure (a) the detected features on the left hand side of the symmetric object are correctly matched between views, while matching fails on the right hand side. As a consequence there is insufficient evidence, for the detection of the main plane of bilateral symmetry. A comparatively insignificant solution is returned in Figure 4.18 (c). In figure (d) the relative abundance of matched features is almost entirely contained within a single plane. Consequently, there is insufficient evidence to detect a 3D bilateral symmetry.

4.7 Constraining the Search Further

As implemented in Algorithm 4.4, only the cardinality of the skewed symmetric set is used to select among the different sets detected. The entire set of matches between all feature points is considered for examination, although the random sampling algorithm can limit the proportion of candidates tested. There is clearly scope for both reducing the search space and providing tighter constraints for the selection among the detected skewed symmetric sets.

4.7.1 Additional Constraints for Pruning or Verification

Stronger Evidence for 3D Symmetry

Algorithm 4.3 guarantees that the three line segments forming the basis are non-coplanar. In a scene with a large number of planar line segments, one segment off the plane suffices, however, to define a 3D symmetric point set. In the example of Figure 4.17 we see that many skew symmetric point sets are generated with only such minimal evidence of the presence of a 3D structure. The number of spuriously detected sets could possibly be reduced by requiring more than one segment to lie off a dominant plane.

Euclidean Constraints

In the selection of solutions, there is no preference for more or less skewed symmetric sets. In the next chapter we will investigate the possibility of adding partial knowledge about the geometry of the image sensor, to aid in the selection of those skewed symmetric sets that correspond to bilateral symmetry in a Euclidean frame.

Local Geometric Invariants

The combinatorial cost of testing features globally across the image, has prompted the computation of invariant local signatures [VMUO95] that can provide stronger indication for matching. A volume defined by four end-points of symmetric chords would be the same on either side of a skewed reflection as discussed in Section 2.4.3.

Sub-Groupings

A symmetric structure may contain symmetrically arranged local structure at a scale smaller than the overall scale of the object. Small-scale symmetries can be detected within a limited neighbourhood, reducing the number of feature pairs that need to be inspected. If each feature were initially matched to a fraction a of the total number of features N, then the complexity of the search would go down by same factor a.

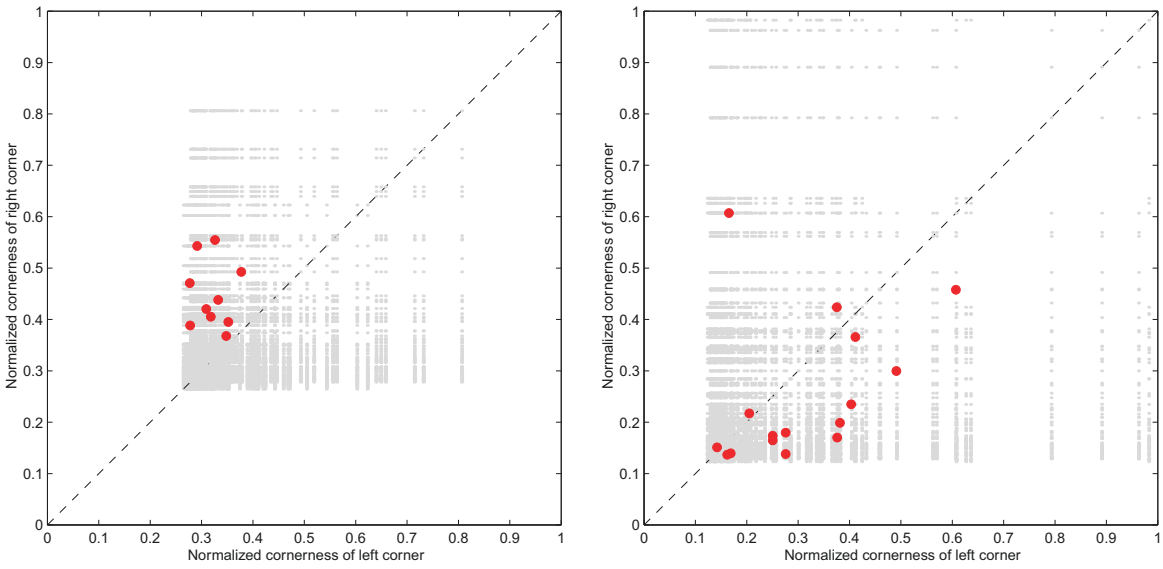
Lower Dimensional features

Bilaterally symmetric three-dimensional objects may contain planar surfaces in a symmetric arrangement. Planar invariants to 3D affine transformation include the ratio of areas, and the ratio of length of parallel lines. Planar surfaces may also contain symmetric figures. Windows on buildings provide a case in point.

4.7.2 Making Use of Weak Constraints

This chapter has focused on the detection of symmetry using geometric constraints, the only information drawn from appearance being the location of sharp intensity changes in the image (Section 4.4.1). It was noted in Section 4.3 that the sequential search for symmetry provided an opportunity to speed up the search at little extra cost by sorting the candidate point pairs prior to searching. To do so, however, requires some external information on which to rank the candidates, which ideally would measure how likely a pair of image points is to correspond to symmetric scene points. To be of use, however, a ranking function needs only to outperform random ordering. Pre-sorting the candidates in this way has the potential of speeding up the search even in the case of weak evidence, without compromising the optimality of the geometric solution.

The first place to look for this kind of information would be in the image intensity function around each point pair. A straight forward correlation of image patches often used in tracking, or perhaps rather correlation with a reflected image patch, is not entirely accurate because of the unknown 3D shape of the surfaces and the different aspect on each of the two points. In this respect this problem shows striking similarity with two well known matching problems: *wide-baseline stereo matching* [PZ98b, PZ98a, Bau00, TV00] and *object recognition by matching to stored views* [SC96, SM97]. The apparent dissimilarity caused



(a) chair (b) stapler

Figure 4.19: Scatterplots of the cornerness [HS88] of symmetric point features (red circles) plotted on the background of the cornerness of all possible pairs of point features. Each plot is drawn from a couple of views of an object. Similar cornerness of symmetric features would be indicated by a concentration of symmetric features around the diagonal line.

by the reflection about the plane of symmetry being non-Euclidean, is easily removed by reflecting one of the two image patches. The two above mentioned problems have proven quite tough, and have been the subject of many studies. Among the proposed methods are grey value invariants to rotation [SM97], warping local image patches by either a homography [PZ98b, PZ98a] or an affine transformation [Bau00, TV00], as well as non-geometric measures such as *grey value histograms* [SC96]. We have not had the opportunity to test any of the above, but while conducting the experiments described in Section 4.4.1, we checked the viability of using the much less sophisticated measure of *cornerness* [HS88], an integral component of the interest operator used to obtain the image features. The initial results, evaluated on the simple scenes used in those experiments, were not particularly encouraging, as made evident by the scatterplots shown in Figure 4.19.

External information about the goodness of candidates can also be used to speed up search by random sampling. If a function measuring the goodness of candidates is available, then rather than allotting all candidates an equal probability of being selected, this probability can be weighted individually by the goodness function. That way the better

candidates are more likely to be selected for testing earlier on in the search.

4.8 Conclusions

This chapter has shown that the detection of 3D skewed symmetry from unconnected point features, using no other information about a feature than its location, is a task of high computational complexity. Despite this, enough headway has been made in managing this complexity through selective testing, to allow us to demonstrate fully automatic detection of 3D skewed symmetry from affine views of uncluttered scenes. The optimal tests derived in the preceding chapter contributed significantly to this result, by enabling us to cope with inexact feature locations and by automatic threshold selection.

It was also seen that the classical constraint satisfaction search introduces inefficiencies in the form of repeated states, that are only eliminated at the cost of considerable implementation effort and storage. A simpler way of searching is offered by random sampling as demonstrated on moderately cluttered scenes. The relative efficiency of the two approaches has yet to be established.

The potentially greatest reduction in computation time, however, could be expected to be had from a richer description of the features, either in the form of connectedness or topology, or by pose invariant local descriptors. This kind of information can be expected to play a similar rôle in a practical application of symmetry detection as in other hard matching problems such as wide-baseline stereo and object recognition.

5

Verification of Euclidean Symmetry

5.1 Overview

This chapter investigates means to verify the existence of bilateral rather than skewed symmetry in the scene. Section 5.3 presents two linear orthogonality constraints on the absolute conic or its dual for each bilaterally symmetric point set. Section 5.4 develops a statistical test of bilateral symmetry from three symmetric sets in a general arrangement, while Section 5.5 catalogues the degenerate arrangements. Section 5.6 discusses the possibilities, and the pitfalls, of incorporating assumptions about the internal parameters of the affine camera to test for bilateral symmetry in fewer than three sets. Section 5.7 then develops the statistical tests, by simultaneously applying constraints on the absolute conic and its dual. The machinery developed in this chapter can be applied more generally to the rectification of an affine structure from any hypothesized orthogonality relations, as reported in [Thó00].



Figure 5.1: A skewed symmetric bench in Euclidean space. A common sight in Olympia, Greece.

5.2 Introduction

So far we have considered the detection of skewed rather than bilateral symmetry, for the simple reason that the information to distinguish between the two is not available in affine views. The missing information is the unknown global mapping between the affine structure and the Euclidean structure of the scene. This mapping could be obtained in a seperate calibration step, either by placing a calibration object in the scene or by making certain assumptions about the camera parameters to effect self-calibration [Qua96]. The fact that knowledge about the affine-to-Euclidean mapping is required in order to determine whether a skewed symmetry is bilateral, also means that knowledge about bilateral symmetry in the scene provides constraints on this mapping. Given enough symmetric sets we should thus be able to *work out* the mapping, and with it the Euclidean structure of the whole scene. The assumption that every skewed symmetry in an affine frame must be a bilateral symmetry in the Euclidean frame is a false one, as is seen by the example in Figure 5.1. If we have more skewed symmetric sets than the bare minimum, however, then we should be able to

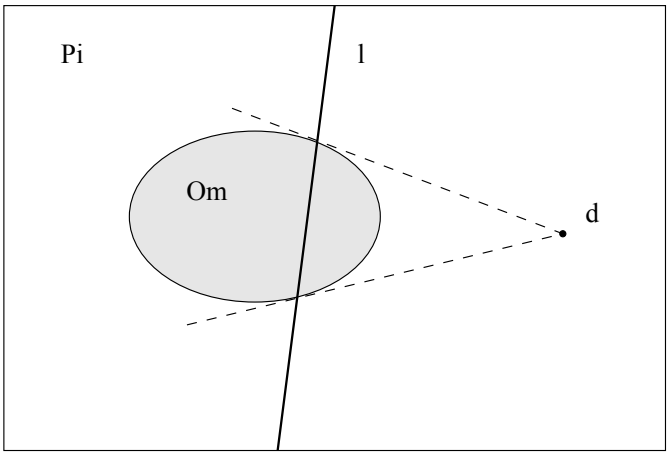


Figure 5.2: Pole-polar relationship with respect to the absolute conic.

verify whether they are mutally consistent with the hypothesis that they are all bilaterally symmetric in a Euclidean frame. Finally, it should also be possible to reduce the number of sets, required for testing, in a hybrid method incorporating reasonable assumptions about the camera parameters. To provide a usable method, we need to formalize the constraints, provide a statistical test to account for measurement errors, and to look at degenerate cases. Those are the tasks of this chapter.

5.3 Euclidean Consistency from Three Symmetric Sets

The knowledge of a bilaterally symmetric point set in an affine structure provides partial constraints on the Euclidean structure of space. We will now formulate the partial constraints as two constraints on the absolute conic or equivalently on its dual.

The constraint provided by a bilaterally symmetric set is that the chords connecting symmetric point pairs must be orthogonal to the plane of symmetry. This translates to a *pole-polar relationship* [SK52, MZ92a, ZLA98], w.r.t. the absolute conic Ω, between the chord direction $\mathbf{d}$ and the line $\mathbf{l}$ where the plane of symmetry meets π_∞, the plane at infinity. The pole-polar relationship, shown geometrically in Figure 5.2, is encapsulated in each of the two equivalent transformations or *polarities* of π_∞

$$\mathbf{l} = \Omega\mathbf{d}, \qquad \text{and} \qquad \mathbf{d} = \Omega^{-1}\mathbf{l} = \Omega^{*}\mathbf{l}, \tag{5.1}$$

where Ω^* is the dual conic. Choosing any two distinct points $\mathbf{d}_1$, $\mathbf{d}_2$ on $\mathbf{l}$, we obtain from

(5.1) the two orthogonality relations in the point conic Ω

$$\boxed{\begin{aligned}\mathbf{d}_1^T\Omega\mathbf{d} &\stackrel{!}{=} 0\\ \mathbf{d}_2^T\Omega\mathbf{d} &\stackrel{!}{=} 0.\end{aligned}} \tag{5.2}$$

Dually, any two distinct lines $\mathbf{l}_1$, $\mathbf{l}_2$ on π_∞ intersecting in $\mathbf{d}$, give us equivalent relations in the line conic Ω^*

$$\boxed{\begin{aligned}\mathbf{l}_1^T\Omega^*\mathbf{l} &\stackrel{!}{=} 0\\ \mathbf{l}_2^T\Omega^*\mathbf{l} &\stackrel{!}{=} 0.\end{aligned}} \tag{5.3}$$

It follows that *three* planes of symmetry, in a general position relative to each other, together provide six constraints on the five parameters of the absolute conic. Since five constraints suffice to determine Ω (up to scale), and with it the Euclidean structure of space (again up to scale), the extra constraint can be used to test the three planes for consistency with the hypothesis that the skewed symmetry detected in the affine frame is due to a bilateral symmetry in Euclidean space. It should be noted that since there is only one affine-to-Euclidean transformation for the whole structure, the three bilaterally symmetric point sets need not have any points in common. The proposed method for recovering scaled Euclidean structure from affine views with the aid of detected bilateral symmetry is summarized in Algorithm 5.1. The necessary statistical tests are developed in the next section.

5.4 Testing for Orthogonality

The previous section showed how three planes of bilateral symmetry must satisfy six orthogonality relations which, for a general arrangement of planes, fully constrain the absolute conic in the affine frame. The current section develops a statistical test of consistency between orthogonality relations and an optimal estimator of the absolute conic. This seemingly simple problem is confounded by the facts that (i) the structure elements assumed to be orthogonal are corrupted by measurement noise as illustrated in Figure 5.3, (ii) that the angle measure is not linear in the errors, and (iii) that the assumption of orthogonality may in some cases be unfounded.

Assuming normal errors on the structure, we make the following contributions. We show that for all but the shortest direction vectors the distribution of their inner product

Given two or more affine views, compute a scaled Euclidean reconstruction using detected skewed symmmetry.

1. Detect $n \geq 3$ skewed symmetric point sets with at least 3 distinct chord directions, using Algorithm 4.2 or Algorithm 4.4.
2. Repeat for m samples:
3. Randomly pick three skewed symmetric sets with distinct chord directions.
4. Test the resulting six orthogonality relations (5.2) for consistency using Algorithm 5.2, estimating the absolute conic $\widehat{\Omega}$ in the process.
5. For each remaining skewed symmetric set:
   (a) Compute the estimated variance of the two orthogonality relations using the unbiased estimator (5.15).
   (b) Reject as outlier if $\frac{\left(\widehat{\mathbf{d}}_{k_1}^T \widehat{\Omega} \widehat{\mathbf{d}}'_{k_1}\right)^2}{\widehat{\sigma}_{k_1}^2} + \frac{\left(\widehat{\mathbf{d}}_{k_2}^T \widehat{\Omega} \widehat{\mathbf{d}}'_{k_2}\right)^2}{\widehat{\sigma}_{k_2}^2} > \chi^2_{2,\alpha}$.
6. Retain $\widehat{\Omega}$ if the number of inliers exceeds all previous samples and repeat from step 2.
7. Compute the Cholesky decomposition of the absolute conic $\widehat{\Omega} = \widehat{\mathsf{K}}^{-T}\widehat{\mathsf{K}}^{-1}$.
8. Multiply the estimated affine structure by $\widehat{\mathsf{K}}^{-1}$ to obtain the scaled Euclidean reconstruction $\widehat{\mathsf{X}}_E = \widehat{\mathsf{K}}^{-1}\widehat{\mathsf{X}}_A$.

Algorithm 5.1: Outline of a method for obtaining a scaled Euclidean reconstruction from two or more affine views by means of detected bilateral symmetries.

is indeed well described by a normal distribution, and we give an unbiased estimator of its mean and its variance. Using this distribution we give a likelihood ratio test of the hypothesis of $K > 5$ mutually consistent orthogonality relations, and a maximum likelihood estimator of the absolute conic based on the same relations, thus providing optimal rectification of the structure. We also provide simple means of computing these in a short succession of linear steps, and assess the validity of that approach. These results are not specific to symmetry detection, but can be used to test orthogonal directions and to upgrade an affine structure to a scaled Euclidean structure, as reported in [Thó00].

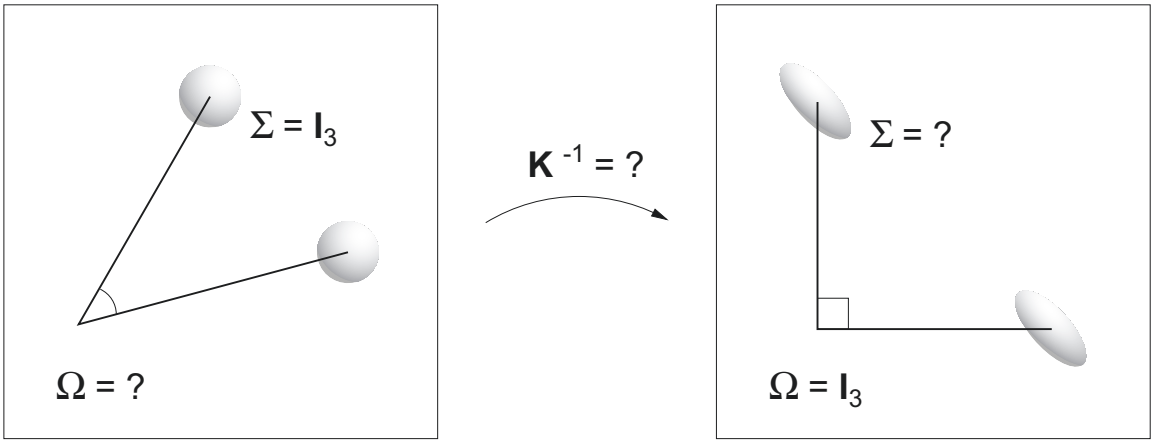


Figure 5.3: Rectification amounts to the estimation of an unknown transformation K^{-1}, which maps the affine structure (left) into Euclidean space (right). The unknown transformation takes the known (isotropic) error distribution of the affine structure into an unknown non-isotropic distribution. Since orthogonality is enforced in the Euclidean space, the unknown error distribution must be taken into account in the estimation of K^{-1}.

Problems with Least Squares Estimation

Given $K \geq 5$ orthogonality constraints of the type (5.2) we can obtain a least squares solution for Ω in a linear algorithm [Qua96, LZ99], along the lines of the well known *eight point algorithm* [Har95a] for the estimation of the fundamental matrix. This is achieved by writing the components of the matrix

$$\Omega = \begin{bmatrix} \omega_1 & \omega_2 & \omega_3 \\ \omega_2 & \omega_4 & \omega_5 \\ \omega_3 & \omega_5 & \omega_6 \end{bmatrix}$$

in the vector

$$\omega \stackrel{\text{def}}{=} \begin{bmatrix} \omega_1 & \omega_2 & \omega_3 & \omega_4 & \omega_5 & \omega_6 \end{bmatrix}^T$$

and collecting the six coefficients of each of the K orthogonality relations in the $K \times 6$ matrix D to pose the minimization problem

$$\left\| \mathsf{D}\omega \right\| \stackrel{!}{=} \min, \quad \text{subject to} \quad \left\| \omega \right\| \stackrel{!}{=} 1. \tag{5.4}$$

The solution $\omega = \widehat{\omega}$ which minimizes (5.4) is the least eigenvector of the matrix D, i.e. the unit eigenvector corresponding to the smallest eigenvalue.

When we are faced with uncertainty in the form of (i) *errors on measured* directions between points in the affine structure and possibly also (ii) *erroneous assumptions about*

orthogonality in the scene, we run into problems with this direct linear method. Firstly, it provides no proper mechanism to combine the K constraints such as to minimize the resulting error in Ω. Secondly, it provides no means of checking an over-constrained set of $K \geq 6$ constraints for consistency, since it is by no means clear, whether and then how the residual error of the least squares fit given by the smallest eigenvalue of D should be interpreted. It should be noted that common assumptions about isotropic measurement errors are of no help here, since the constraint is being enforced in the unknown Euclidean space, as illustrated in Figure 5.3. In the remainder of this section we attempt to resolve both the above issues.

5.4.1 Distribution of the Inner Product

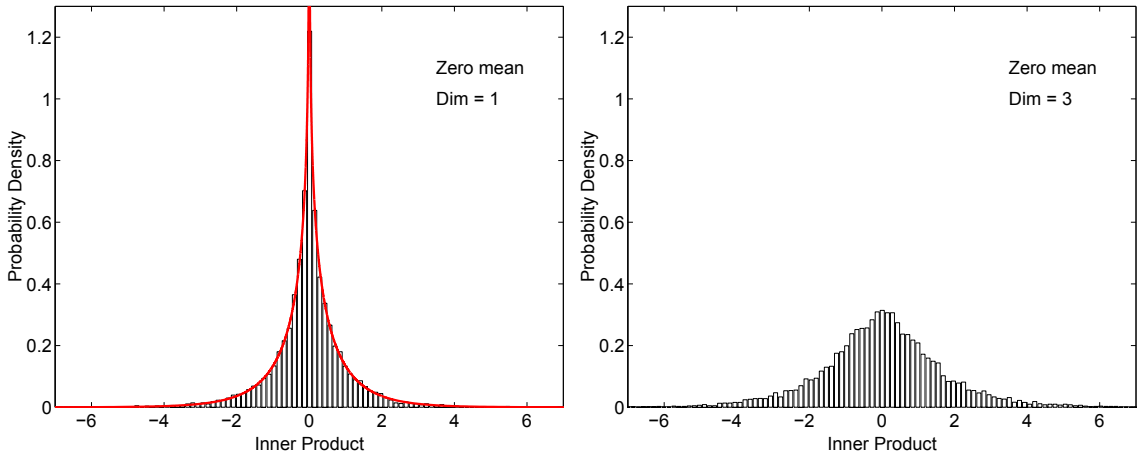
Consider the two deterministic directions $\mathbf{d}_k$ and $\mathbf{d}'_k$ in the affine frame whose estimates $\widehat{\mathbf{d}}_k$ and $\widehat{\mathbf{d}}'_k$ are corrupted by additive random noise according to

$$\widehat{\mathbf{d}}_k = \mathbf{d}_k + \mathbf{n}_k, \quad \text{and} \quad \widehat{\mathbf{d}}'_k = \mathbf{d}'_k + \mathbf{n}'_k.$$

Let us assume that the errors are independent, and have a normal distribution $\mathbf{n} \sim N(\mathbf{0}, \Sigma)$ and $\mathbf{n}' \sim N(\mathbf{0}, \Sigma')$. Using the erroneous estimates, the Euclidean inner product from (5.2) expands to

$$\begin{aligned}\widehat{c}_k \overset{\text{def}}{=} & \widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}'_k \\ = & (\mathbf{d}_k + \mathbf{n}_k)^T \Omega (\mathbf{d}'_k + \mathbf{n}'_k) \\ = & \mathbf{d}_k{}^T \Omega \mathbf{d}'_k + \mathbf{d}_k{}^T \Omega \mathbf{n}'_k + \mathbf{d}'_k{}^T \Omega \mathbf{n}_k + \mathbf{n}_k{}^T \Omega \mathbf{n}'_k. \end{aligned} \tag{5.5}$$

The first term $\mathbf{d}_k{}^T \Omega \mathbf{d}'_k$ on the right hand side of (5.5) is fixed. Since $\mathbf{d}_k{}^T \Omega$ and $\mathbf{d}'_k{}^T \Omega$ are fixed vectors, the next two terms are univariate normal by the definition of a multivariate normal random variable [MKB79, p. 60]. The final term $\mathbf{n}_k{}^T \Omega \mathbf{n}'_k$ is not normal, but is insignificant compared to the first three terms unless both $\widehat{\mathbf{d}}_k$ and $\widehat{\mathbf{d}}'_k$ are on the magnitude of the measurement noise. The last term is also a weighted sum of products of univariate normal variables. While a single product of two normal variables has the non-normal distribution shown in Figure 5.4(a), the sum of three products shown is Figure 5.4(b) is starting to resemble a normal distribution. For the time being we will therefore consider



(a) Univariate case. (b) 3D multivariate case.

Figure 5.4: The empirical p.d.f. of the inner product of two independent normal random variables with zero mean and unit variance. The solid red line in (a) shows the distinctively non-normal theoretical distribution $\frac{1}{\pi}K_0(|x|)$, where K_0 is the modified Bessel function of the second kind.

the distribution of the final term close enough to normal for our purposes, bearing in mind that any deviation from a normal distribution will only show at a low SNR.

To complete the specification of the distribution of $\widehat{c}_k$ we need to provide its mean and variance. Having assumed that the errors on the two direction vectors are zero mean and uncorrelated, i.e. $\mathrm{E}\{\mathbf{n}_k\} = \mathrm{E}\{\mathbf{n}'_k\} = \mathbf{0}_3$ and $\mathrm{E}\left\{\mathbf{n}_k{\mathbf{n}'_k}^T\right\} = 0_{3\times3}$, the mean of $\widehat{c}_k$ is given by

$$c_k \stackrel{\text{def}}{=} \mathrm{E}\{\widehat{c}_k\} = {\mathbf{d}_k}^T\Omega\mathbf{d}'_k, \tag{5.6}$$

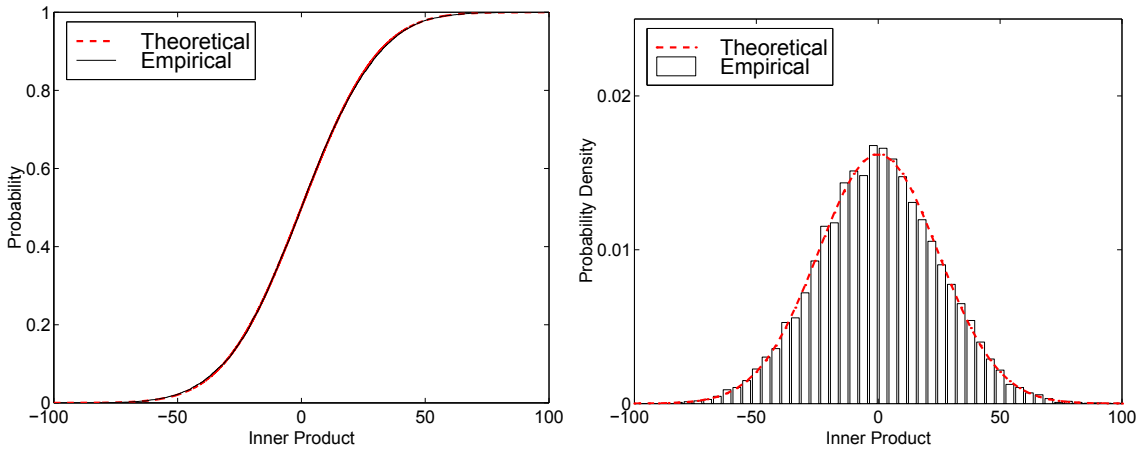
since all subsequent terms of (5.5) vanish under the expectation. The variance of $\widehat{c}_k$ is given by

$$\begin{aligned}\sigma_k^2 \stackrel{\text{def}}{=}&\mathrm{E}\left\{(\widehat{c}_k - c_k)^2\right\}\\ =&\mathrm{E}\left\{\left({\mathbf{d}_k}^T\Omega\mathbf{n}'_k + {\mathbf{d}'_k}^T\Omega\mathbf{n}_k + {\mathbf{n}_k}^T\Omega\mathbf{n}'_k\right)^2\right\}\\ =&{\mathbf{d}_k}^T\Omega\Sigma'_k\Omega\mathbf{d}_k + {\mathbf{d}'_k}^T\Omega\Sigma_k\Omega\mathbf{d}'_k + \mathrm{tr}\,\Omega\Sigma_k\Omega\Sigma'_k.\end{aligned} \tag{5.7}$$

In the case of standardized errors in the affine structure ($\Sigma_k = \Sigma'_k = \mathsf{I}_{3\times3}$), this is simply

$$\sigma_k^2 = {\mathbf{d}_k}^T\Omega^2\mathbf{d}_k + {\mathbf{d}'_k}^T\Omega^2\mathbf{d}'_k + \mathrm{tr}\,\Omega^2. \tag{5.8}$$

It follows that we expect the conditional p.d.f. of the Euclidean inner product (5.5) of two



(a) Cumulative probability distribution.

(b) Probability density.

Figure 5.5: The theoretical small noise distribution (red) of the *inner product of two orthogonal direction vectors* in Euclidean 3-space, as given by (5.9) and (5.8), and the empirical distribution (black) of synthetically generated direction vectors with 10% RMS added noise.

direction vectors in the affine structure to be closely approximated by

$$p(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}'_k | c_k, \Omega) \approx \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{ -\frac{\left(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}'_k - c_k\right)^2}{2\sigma_k^2} \right\}. \tag{5.9}$$

The approximate distribution given by (5.9) is shown plotted against the true distribution in Figure 5.5. The results of a Kolmogorov-Smirnov test [BNB96] of the distribution (5.9) are given in Table 5.1. They indicate that the distribution of c_k is indeed closely approximated by the normal distribution (5.9), when the RMS noise level is no more than than half the RMS size of the direction vectors, provided of course that the errors on the direction vectors

noise	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
1%	0.590	0.877	pass
10%	0.613	0.846	pass
50%	1.29	0.071	pass
100%	3.26	10^{-9}	fail

Table 5.1: Results of the Kolmogorov-Smirnov test on the null distribution of the scalar product given by (5.9) and (5.8), when $\Omega = \mathsf{I}_{3\times 3}$. The data are those of Figure 5.5. The trials were repeated with different levels of RMS noise, relative to the size of the direction vectors.

are normal as assumed.

5.4.2 Testing Relations for Consistency

We now seek a statistical test of the hypothesis that $K > 5$ independent relations of the type (5.2) are indeed all orthogonality relations

$$H_0 : \ \mathbf{d}_k^T \Omega \mathbf{d}_k' = 0, \ \forall k \in \{1, \ldots, K\},$$

against the alternative hypothesis that some of them are not

$$H_1 : \ \mathbf{d}_k^T \Omega \mathbf{d}_k' \neq 0, \ \exists k \in \{1, \ldots, K\},$$

when our knowledge of $\mathbf{d}_k$ and $\mathbf{d}_k'$ is uncertain. Collecting the individual inner products c_k together in the vector $\mathbf{c} \stackrel{\text{def}}{=} \{c_k\}$, we can rewrite the hypotheses in the form

$$H_0 : \mathbf{c} = \mathbf{0}_K \quad \text{against} \quad H_1 : \mathbf{c} \neq \mathbf{0}_K.$$

There is no single correct way of deciding one way or the other, but the answer that minimizes the probability of mis-detection at a set probability of false alarm is given by the *likelihood ratio test* [BNB96]. This requires us to specify the joint conditional p.d.f. or *likelihood function* of each hypothesis. From (5.9) we obtain the unconstrained likelihood function

$$\begin{aligned} L(\mathbf{c}, \Omega) &\stackrel{\text{def}}{=} \prod_{k=1}^{K} p(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k' | \mathbf{c}, \Omega) \\ &\approx \frac{1}{\left[(2\pi)^k \prod_{k=1}^{K} \sigma_k^2\right]^{1/2}} \cdot \exp\left\{ -\tfrac{1}{2} \sum_{k=1}^{K} \frac{\left(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k' - c_k\right)^2}{\sigma_k^2} \right\}, \end{aligned} \tag{5.10}$$

corresponding to the alternative hypothesis, with the null distribution

$$L(\mathbf{0}, \Omega) \approx \frac{1}{\left[(2\pi)^k \prod_{k=1}^{K} \sigma_k^2\right]^{1/2}} \cdot \exp\left\{ -\tfrac{1}{2} \sum_{k=1}^{K} \frac{\left(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k'\right)^2}{\sigma_k^2} \right\}. \tag{5.11}$$

The likelihood ratio of the hypotheses is defined as

$$\Lambda \stackrel{\text{def}}{=} \frac{\max L_{H_0}}{\max L_{H_1}} = \frac{\max_{\Omega} L(\mathbf{0}, \Omega)}{\max_{\{\mathbf{c}, \Omega\}} L(\mathbf{c}, \Omega)}. \tag{5.12}$$

We start by maximizing the unconstrained likelihood $L(\mathbf{c}, \Omega)$, given in (5.10). The exponent in (5.10) is always less or equal to zero. Hence the exponential is maximized by the choice $c_k = \widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k'$, at which the exponent vanishes, leaving only the normalizing constant giving

$$\max_{\{\mathbf{c},\Omega\}} L(\mathbf{c}, \Omega) \approx \max_{\Omega} \frac{1}{\left[(2\pi)^k \prod_{k=1}^K \sigma_k^2\right]^{1/2}}.$$

The variance σ_k^2 depends on the matrix Ω through (5.7). The matrix of the absolute conic Ω is only recoverable from orthogonality constraints up to scale, however, so the overall scale of σ_k^2 is cannot be determined. This is not of major concern, since the relative size of the variance between the constraints is not affected, but without a constraint on the scale of Ω the estimates of σ_k^2 can be made to vanish, leaving the likelihood functions (5.10) and (5.11) unbounded and the likelihood ratio (5.12) undefined. It is particularly convenient to set the scale of Ω by fixing $\prod_{k=1}^K \sigma_k^2$ to an arbitrary positive constant. This causes the normalizing constants of (5.10) and (5.11) to cancel so the logarithm of the likelihood ratio (5.12) simplifies to

$$\boxed{-2 \log \Lambda \approx \min_{\Omega} \sum_{k=1}^{K} \frac{\left(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k'\right)^2}{\sigma_k^2}.} \tag{5.13}$$

We note that the parameters of Ω that minimize (5.13) also maximize the likelihood function (5.11), thus providing the *maximum likelihood estimator* of Ω from K orthogonality constraints.

Since we have shown that $\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k'$ is approximately $N(0, \sigma_k^2)$, it follows immediately that $\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k' / \sigma_k$ is $N(0, \mathsf{I})$ so the right hand side of (5.13) is χ^2 distributed with $K - 5$ degrees of freedom. We note that this approximate distribution, which is valid for all K, is identical to the general asymptotic distribution (as $K \mapsto \infty$) predicted by the theory of likelihood ratio tests [BNB96].

5.4.3 Unbiased Estimator of the Variance

The one remaining thing needed in order to compute the likelihood ratio (5.13), is to estimate the variance σ_k^2, given Ω and the measured directions $\widehat{\mathbf{d}}_k$ and $\widehat{\mathbf{d}}_k'$. Substituting the

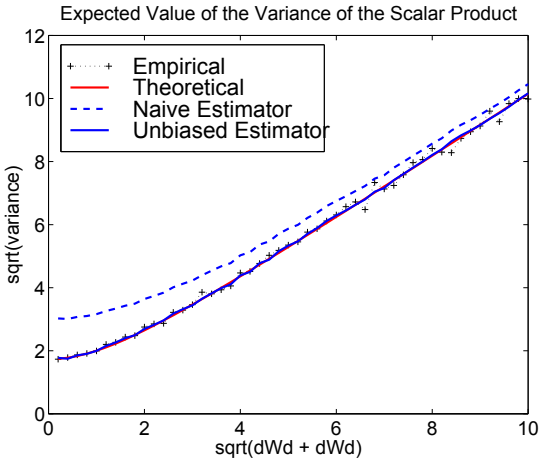


Figure 5.6: The expected value of the variance of $c_k = \widehat{\mathbf{d}}^T \Omega \widehat{\mathbf{d}}'$ shown as a function of ${\mathbf{d}_k}^T \Omega^2 \mathbf{d}_k + {\mathbf{d}'_k}^T \Omega^2 \mathbf{d}'_k$. Shown is an empirical RMS value of c_k from 1000 trials using synthetic data. Also shown is the theoretical value σ_k given by (5.7), as well as the RMS value of each of the two estimators $\widehat{\sigma}_k$ (5.14) and (5.15), obtained from the same data.

measured directions straight into (5.7) we obtain the naïve estimator

$$\widehat{\sigma}_k^2 \stackrel{\text{def}}{=} \widehat{\mathbf{d}}_k^T \Omega \Sigma'_k \Omega \widehat{\mathbf{d}}_k + \widehat{\mathbf{d}}_k'^T \Omega \Sigma_k \Omega \widehat{\mathbf{d}}'_k + \operatorname{tr} \Omega \Sigma_k \Omega \Sigma'_k. \tag{5.14}$$

This estimator is biased, however, since

$$\begin{aligned} \mathrm{E}\left\{\widehat{\mathbf{d}}^T \Omega \Sigma'_k \Omega \widehat{\mathbf{d}}\right\} =& \mathrm{E}\left\{(\mathbf{d}_k + \mathbf{n}_k)^T \Omega \Sigma'_k \Omega (\mathbf{d}_k + \mathbf{n}_k)\right\} \\ =& \mathbf{d}_k^T \Omega \Sigma'_k \Omega \mathbf{d}_k + \mathrm{E}\left\{\mathbf{n}_k^T \Omega \Sigma'_k \Omega \mathbf{n}_k\right\} \\ =& \mathbf{d}_k^T \Omega \Sigma'_k \Omega \mathbf{d}_k + \operatorname{tr} \Omega \Sigma_k \Omega \Sigma'_k, \end{aligned}$$

and similarly for $\mathrm{E}\left\{\widehat{\mathbf{d}}'^T \Omega \Sigma_k \Omega \widehat{\mathbf{d}}'\right\}$, giving

$$\mathrm{E}\left\{\widehat{\sigma}_k^2\right\} = \sigma_k^2 + 2 \operatorname{tr} \Omega \Sigma_k \Omega \Sigma'_k \neq \sigma_k^2.$$

Subtracting the fixed bias $2 \operatorname{tr} \Omega \Sigma_k \Omega \Sigma'_k$ from (5.14) we obtain the unbiased estimator

$$\boxed{\widehat{\sigma}_k^2 \stackrel{\text{def}}{=} \widehat{\mathbf{d}}_k^T \Omega \Sigma'_k \Omega \widehat{\mathbf{d}}_k + \widehat{\mathbf{d}}_k'^T \Omega \Sigma_k \Omega \widehat{\mathbf{d}}'_k - \operatorname{tr} \Omega \Sigma_k \Omega \Sigma'_k,} \tag{5.15}$$

which for standardized errors is

$$\boxed{\widehat{\sigma}_k^2 = \widehat{\mathbf{d}}_k^T \Omega^2 \widehat{\mathbf{d}}_k + \widehat{\mathbf{d}}_k'^T \Omega^2 \widehat{\mathbf{d}}'_k - \operatorname{tr} \Omega^2.} \tag{5.16}$$

The unbiased estimator (5.15) significantly improves the estimate of the variance of the inner product of relatively short direction vectors $\widehat{\mathbf{d}}_k$ and $\widehat{\mathbf{d}}'_k$, as is seen from Figure 5.6.

Given 5 or more pairs of orthogonal direction vectors together with their covariance matrices, compute the ML estimate of the absolute conic and test for consistency.

1. Initialize $\Omega \leftarrow \Omega_0$, (e.g. using $\Omega_0 = \mathsf{I}_3$).
2. Repeat steps 3. through 4. two or more times:
   3. Estimate the variance of each of the $K \geq 5$ orthogonality relations using (5.15):
      $$\widehat{\sigma}_k^2 \leftarrow \widehat{\mathbf{d}}_k^T \Omega \Sigma_k' \Omega \widehat{\mathbf{d}}_k + \widehat{\mathbf{d}}_k'^T \Omega \Sigma_k \Omega \widehat{\mathbf{d}}_k' - \operatorname{tr} \Omega \Sigma_k \Omega \Sigma_k'.$$
   4. Estimate the entries of Ω as the least eigenvector of the linear minimization problem (5.13):
      $$\sum_{k=1}^K \frac{\left(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k'\right)^2}{\widehat{\sigma}_k^2} \stackrel{!}{=} \min .$$
5. Reject the set of orthogonality relations if the smallest eigenvalue in step 4. is larger than $\chi^2_{K-5,\alpha}$

Algorithm 5.2: Computation of the ML estimate of the absolute conic in a simple iterative algorithm.

It should be noted, however, that for very short $\widehat{\mathbf{d}}_k$ and $\widehat{\mathbf{d}}_k'$ (5.15) can evaluate to a zero or negative estimate of variance. This is of a minor concern here, since such an occurrence is easily detected and dealt with, and we are more interested in obtaining an unbiased estimate of variance over the range of reasonable SNR.

5.4.4 Computing the Likelihood Ratio

Substituting the unbiased estimator of the variance (5.15) into the expression for the log-likelihood ratio (5.13) gives, in the case of standardized errors,

$$-2 \log \Lambda \approx \min_{\Omega} \sum_{k=1}^{K} \frac{\left(\widehat{\mathbf{d}}_k^T \Omega \widehat{\mathbf{d}}_k'\right)^2}{\widehat{\mathbf{d}}_k^T \Omega^2 \widehat{\mathbf{d}}_k + \widehat{\mathbf{d}}_k'^T \Omega^2 \widehat{\mathbf{d}}_k' - \operatorname{tr} \Omega^2}. \tag{5.17}$$

The minimization problem described by this equation can be solved by a suitable numerical method. We note, however, that the numerator should vary rapidly close to the correct solution, while the denominator varies slowly. We therefore propose a simplified algorithm,

consisting of alternating between the two linear estimation problems: (i) estimating the denominator σ_k^2 using (5.15) and (ii) minimizing the original expression (5.13) with σ_k^2 fixed. The method is summarized in Algorithm 5.2. The distribution of the test statistic computed using 1, 2, and 5 iterations of the algorithm is shown in Figure 5.7 for varying number K of orthogonality relations and noise level. As is seen in Figure 5.7, the test statistic computed using only two iterations of Algorithm 5.2 follows the expected chi-square distribution closely for noise levels up to 3-10% of the RMS size of the direction vectors being tested.

5.5 Degenerate Arrangements of Symmetry Planes

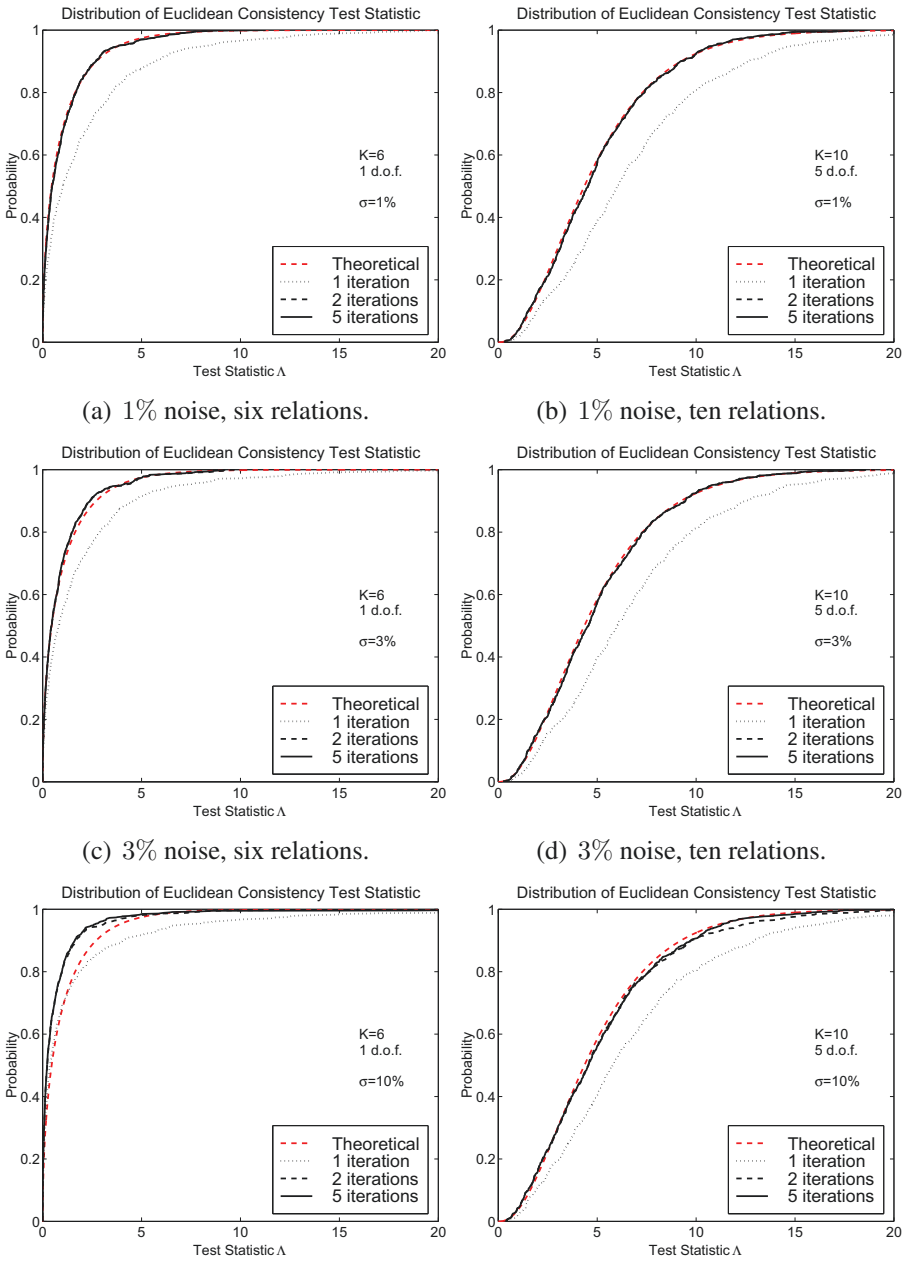
Special arrangements of the three planes of symmetry may reduce the number of independent constraints from the required six. These are catalogued in Table 5.2. The rank of the

relation between normals	relation between planes	rank of constraint
coplanar	intersect in parallel lines	2+2+0 = 4
all repeated	all parallel	2+0+0 = 2
two repeated	two parallel	2+0+2 = 4
one orthogonal pair	one orthogonal pair	2+1+2 = 5
all orthogonal	all orthogonal	2+1+0 = 3

Table 5.2: Special arrangements of three planes of symmetry, which produce fewer than the expected six independent orthogonality relations.

degenerate set is shown in the rightmost column of the table. It was obtained by geometric reasoning, summing up the gain in rank as each symmetry plane is added to the set. The results were verified by examining the rank of the D in (5.4) using simulated data. Since a critical arrangement is tested for using these same constraints, the presence of a critical arrangement has the sole effect of making that set trivially pass our test. We also note, however, that a necessary condition for a triplet of symmetry sets to be in a general arrangement is that the three constituent sets are pairwise in a general arrangement. Consequently, a test constructed from two sets should be of rank four.

We observe two connections to the literature. First, we note that the first two cases in Table 5.2 correspond to the cases of planar motion and pure translation respectively,



(a) 1% noise, six relations.

(b) 1% noise, ten relations.

(c) 3% noise, six relations.

(d) 3% noise, ten relations.

(e) 10% noise, six relations.

(f) 10% noise, ten relations.

Figure 5.7: Empirical (black) and theoretical (red) null distribution of the test statistic (5.17) of K *consistent orthogonality relations*, using the simple iterative Algorithm 5.2. Each figure shows the theoretical χ^2_{K-5} distribution, as well as the empirical distribution of the test statistic after 1, 2, and 5 iterations of the algorithm. The RMS noise level is measured relative to the magnitude of the direction vectors.

two motions critical to self-calibration of projective cameras, and studied in depth by other authors [AZH96, Stu97, ZLA98]. Secondly, the case of all three planes being mutually orthogonal can be described in terms of a *self-polar triangle* [SK52] on the plane at infinity, a notation recently used by Liebowitz and Zisserman [LZ99] in the context of calibrating a projective camera from known orthogonal directions.

relation between normals	condition on poles	condition on polars	dependencies
coplanar	$[\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3] = 0$	$[\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3] = 0$	2
all repeated	$\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{d}_3$	$\mathbf{l}_1 = \mathbf{l}_2 = \mathbf{l}_3$	4
two repeated	$\mathbf{d}_1 = \mathbf{d}_2$	$\mathbf{l}_1 = \mathbf{l}_2$	2
one orthogonality	$\mathbf{l}_1^T \mathbf{d}_2 = 0$	$\mathbf{l}_1^T \mathbf{d}_2 = 0$	1
all orthogonal	$\mathbf{l}_k^T \mathbf{d}_{k'} = 0, \forall k \neq k'$	$\mathbf{l}_k^T \mathbf{d}_{k'} = 0, \forall k \neq k'$	3

Table 5.3: The critical arrangement of three planes of symmetry described in Table 5.2 referred to the plane at infinity.

In Table 5.3 we provide an algebraic formulation of the constraints in Table 5.2. The bottom two constraints in Table 5.2 are obtained as described in [LZ99]. The rightmost column of the table shows the number of dependencies dictated by the algebraic constraints. Added together, the rightmost columns of the two tables Table 5.3 and Table 5.2 gives the total number of six constraints.

5.6 Incorporating Partial Camera Calibration

Section 5.3 gave relations satisfied by three or more bilaterally symmetric point sets, that effectively constrain the absolute conic in the affine frame. For the purpose of bilateral symmetry detection, a test requiring fewer sets would be more effective for discarding non-bilateral sets. If we are to test fewer than three sets for consistency, however, we must provide additional Euclidean information to fix some of the degrees of freedom in the absolute conic. One possibility is to draw such information from the scene, in the form of orthogonality relations (right-angles, aspect ratios) not induced by symmetry. This would require prior interpretation of the scene, and is therefore not so desireable. Another possibility is to draw similar information from the construction of the camera (skew-free sensor, known or constant aspect ratio), as is done in the self-calibration of projective cameras.

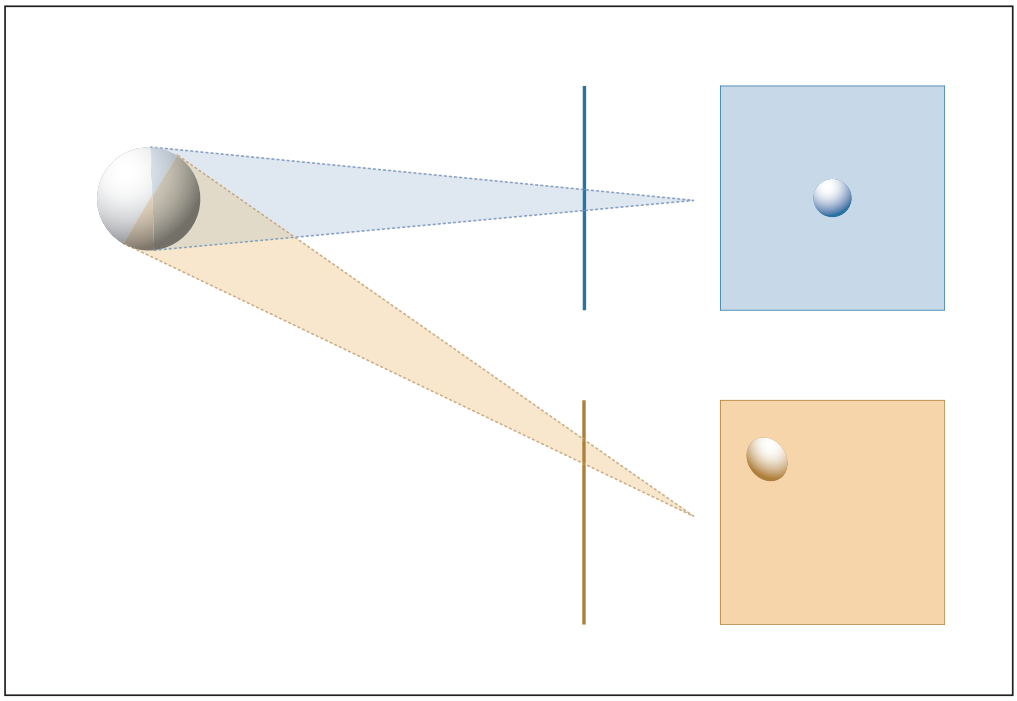


Figure 5.8: Two pin-hole views with constant intrinsics can give rise to two affine views with radically differing affine intrinsics.

In the most common use of the affine camera, as a piecewise approximation of a pin-hole camera, this is not as straight-forward as we shall now see.

5.6.1 Self-Calibration using Assumptions about Camera Intrinsics

The constraint of fixed or partially known intrinsics has been used to self-calibrate an affine camera from multiple views of an unknown scene. A projection matrix M obtained by factorization is related to the true projection matrix M' by a linear transformation D [TK92]

$$\Delta\mathsf{W} = \mathsf{MX} = \mathsf{MDD}^{-1}\mathsf{X} = (\mathsf{MD})(\mathsf{D}^{-1}\mathsf{X}) = \mathsf{M}'\mathsf{X}'$$

Decomposing the true projection matrix for the i-th view M'_i into a 2×2 lower triangular matrix A_i containing the *intrinsic parameters* of the affine camera and a 2×3 rotation matrix R_i [Har94b, Qua96] results in

$$\mathsf{M}_i\mathsf{D} = \mathsf{M}'_i = \mathsf{A}_i\mathsf{R}_i$$

which leads to the relation

$$\mathsf{M}_i\mathsf{D}\mathsf{D}^T\mathsf{M}_i^T = \mathsf{A}_i\mathsf{A}_i^T$$

between the dual of the absolute conic $\Omega^* = \mathsf{D}\mathsf{D}^T$ and its affine dual image (or DIAC), the line conic

$$\omega_i^* = \begin{bmatrix} \mathsf{A}\mathsf{A}^T & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{bmatrix}.$$

This constraint was used by [Qua96] to self-calibrate the camera, under the assumption that the camera intrinsics are fixed up to scale

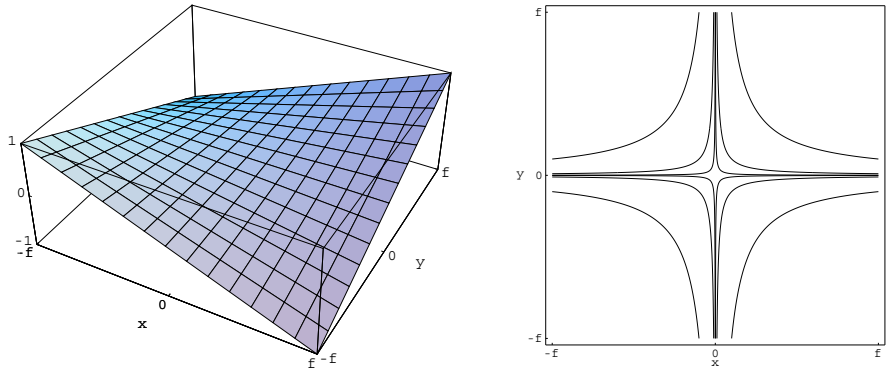
$$\mathsf{A}_i = k_i \mathsf{A}$$

resulting in the self-calibration constraint

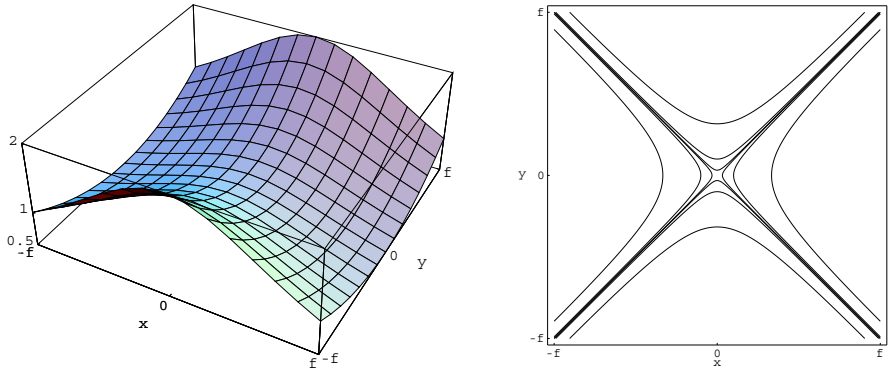
$$\omega_i^* = k_i^2 \omega^*.$$

Limitations of the Assumption of a Fixed Intrinsics

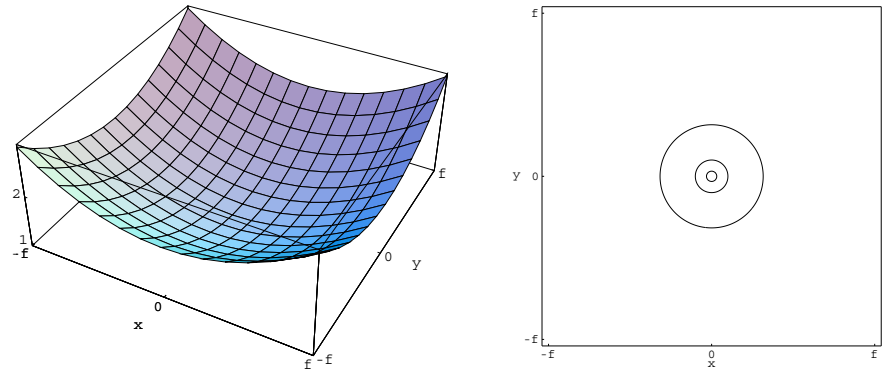
In the perspective model of a pin-hole camera, the intrinsic parameters are coupled from the camera motion. They depend only on internal physical camera parameters: the focal length, aspect ratio, sensor skew, and the principal point. When the affine camera model is used as a piecewise approximation to pin-hole projection, the unknown intrinsic parameters of the affine camera vary with location on the image plane of pin-hole camera. This can be explained by an example. Figure 5.8 shows two views of a sphere, as it moves past a pin-hole camera. In the first view the sphere is close to the principal axis of the camera, so that its image falls close to the principal point. In the second view the sphere is imaged further away from the principal axis, so that its image falls away from the principal point. The first image of the sphere is a circle, but the second image is an elongated ellipse. This is not an artefact of lens distortion, but a perspective effect caused by the image plane cutting the projection cone of the sphere's contour at an angle. The two views can be described by two perspective cameras sharing same intrinsics, only the extrinsic parameters differ between the cameras. The two views can also be well described by affine cameras, but in this case the intrinsics differ between the two cameras. The first camera has zero skew and unit aspect ratio, the second has neither. With respect to calibration, we note that assumptions about the geometry of the camera sensor do not translate directly into constraints on the intrinsic parameters of the affine cameras, except perhaps under special viewing conditions.



(a) Variation in the off-diagonal element ω_{21}.



(b) The ratio of the diagonal elements ω_{22}/ω_{11}.



(c) The ratio of the singular values of ω.

Figure 5.9: Variation of the elements of the affine DIAC ω^* with the position of the image of the reference point $\mathbf{p}_0$ in the para-perspective camera model. The contours on the right show .1%, 1%, and 10% variation in the parameters of the affine DIAC from the fixed values of the projective DIAC corresponding to a skew free sensor with unit aspect ratio.

Quantifying the Variation of the Intrinsics

Figure 5.9 shows how the parameters of the affine DIAC ω^* vary over the image plane of the pin-hole camera. We indentify the position in the image plane with the location of

the *reference point* $\mathbf{p}_0$ in a *para-perspective camera* model. From the figure it is apparent that the assumption of fixed affine intrinsics is only valid if $\mathbf{p}_0$ can be taken as *fixed*. It is also important to note that the assumption of a skew-free sensor with unit aspect ratio is in general not meaningful. Figure 5.9(a) and Figure 5.9(b) show how the off-diagonal element ω^*_{21} and the ratio of the diagonal elements $\omega^*_{22}/\omega^*_{11}$ vary with the position of the reference point in the image. It follows that the knowledge of a skew-free sensor with unit aspect ratio can only be taken to imply $\omega^*_{21} = 0$, and $\omega^*_{22}/\omega^*_{11} = 1$ when the reference point is known to be in the centre of the image. This is the special case of a *weak perspective camera*.

Self-calibration under Varying Intrinsics

The problems associated with assuming constant intrinsics of the affine camera, might suggest methods of self-calibration that allow the camera intrinsics to vary, similar to those recently demonstrated successfully for projective cameras [PKV98, dAHR98]. However these methods require at least one off-diagonal element of the calibration matrix A be known to be zero. In the affine case this would have to be the one and only off-diagonal element $a_{21} = 0$, implying the assumption of the weak perspective camera model. In this case A and therefore ω^* is diagonal, and calibration is achieved from 5 views in a linear algorithm, as in the projective case. This method of calibrating the affine camera was first suggested by Quan [Qua96].

We have thus seen that self-calibration of the affine model approximating a pin-hole camera implies special viewing conditions. We have identified the two following cases:

1. Para-perspective approximation around a fixed reference point in the image with constant intrinsics, and

2. Weak perspective approximation with constant or varying intrinsics.

5.6.2 Constraints from the Weak Perspective Camera

Under the weak perspective model assumption we can equate the affine camera calibration parameters to the physical parameters of the perspective camera, and draw on partial

knowledge about the camera calibration, such as a skew-free image sensor and known aspect ratio.

Orthogonal Image Sensor

An affine camera P_a projects the *absolute disk quadric* Q^* onto the *dual image of the absolute conic* (or *DIAC*) ω^* by the projection equation (A.4)

$$\omega^* = \mathsf{P}_a \mathsf{Q}^* \mathsf{P}_a^T = \mathsf{P}_a \begin{bmatrix} \Omega^* & \mathbf{0}_3 \\ \mathbf{0}_3^T & 0 \end{bmatrix} \mathsf{P}_a^T = \begin{bmatrix} \mathsf{M}\Omega^*\mathsf{M}^T & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{bmatrix} \tag{5.18}$$

as shown in Appendix A. On the left hand side of (5.18) we see that only four of the nine elements of the affine DIAC are not trivially zero, namely

$$\omega_{ij}^* = \mathbf{m}_i^T \Omega^* \mathbf{m}_j, \qquad i,j \in \{1,2\}. \tag{5.19}$$

For a weak perspective camera with an orthogonal (skew free) image sensor we can assume $\omega_{12}^* = 0$, and we have the well known [TK92, Qua96] constraint

$$\boxed{\mathbf{m}_1^T \Omega^* \mathbf{m}_2 \stackrel{!}{=} 0} \tag{5.20}$$

on Ω^*, the dual of the absolute conic. This gives us one constraint per view on the five parameters of Ω^*. By adding the knowledge of a skew-free image sensor under weak perspective imaging conditions, we obtain a Euclidean consistency constraint between two planes of symmetry from two views. Using (5.20) together with the dual form of the symmetry constraints (5.3), we can apply the combined constraints in a *linear* algorithm. Adding two more views gives us a Euclidean consistency constraint on a single plane of symmetry, provided that the four views are obtained under a general motion, a rather strong requirement in many practical situations.

Known Aspect Ratio

If we know the aspect ratio $\omega_{11}^*/\omega_{22}^*$ in any given view, we can always choose the units of the coordinate axes for that view such that the aspect ratio is unity. In this case we have $\omega_{11}^* = \omega_{22}^*$, and we obtain the constraint

$$\mathbf{m}_1^T \Omega^* \mathbf{m}_1 \stackrel{!}{=} \mathbf{m}_2^T \Omega^* \mathbf{m}_2.$$

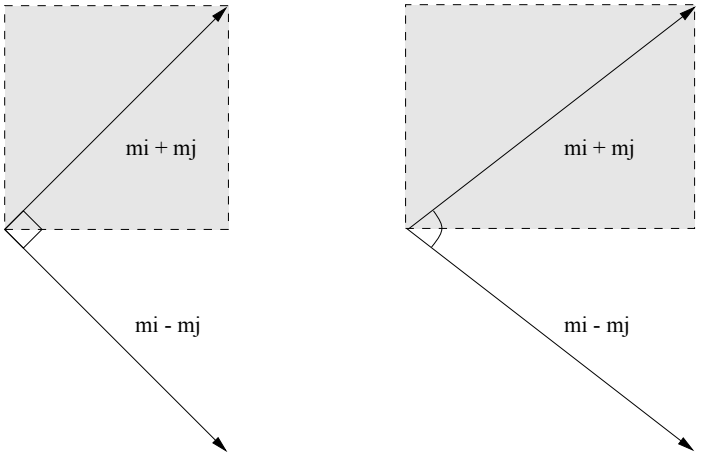


Figure 5.10: Unit aspect ratio expressed as an orthogonality constraint.

By observing that Ω^* is symmetric, we can rewrite this constraint as

$$\boxed{(\mathbf{m}_1 - \mathbf{m}_2)^T \Omega^* (\mathbf{m}_1 + \mathbf{m}_2) \stackrel{!}{=} 0,} \tag{5.21}$$

which is an orthogonality constraint between the two lines $\mathbf{m}_i - \mathbf{m}_j$ and $\mathbf{m}_i + \mathbf{m}_j$, as illustrated in Figure 5.10.

Having expressed the knowledge of an orthogonal sensor with a known aspect ratio as two linear constraints on Ω^* for each view we can solve for the five unknowns of Ω^* to obtain a Euclidean reconstruction from three views in a *linear* algorithm. Compare this to the non-linear algorithm presented in [TK92] under the stronger assumption of orthography.

5.7 Testing for Orthogonality using Partial Calibration

We now look at the problem of testing skewed symmetric sets for bilateral symmetry, while at the same time enforcing constraints imposed by partial knowledge about the camera calibration. The problem here is that the most natural constraints involve different conics. The orthogonality relation between points in the structure (5.2) involves the absolute conic Ω, while the camera constraints (5.20), (5.21) involve the dual conic Ω^*.

5.7.1 Mixing Constraints on Ω and Ω^*

We have seen that five orthogonality constraints fully constrain the absolute conic or its dual. Given five or more constraints of the type (5.2) we obtain the absolute conic Ω in a

linear algorithm. In the dual space the same applies, and we obtain Ω^* in a linear algorithm from five or more constraints of the types (5.3), (5.20), or (5.21). What now, if we have a total of five or more constraints on Ω and Ω^*, but less than five constraints on either conic? Since Ω is symmetric, we have

$$\Omega^* \stackrel{\text{def}}{=} \Omega^{-T} = \Omega^{-1}$$

and we end up with a set of five or more homogeneous linear constraints in the parameters of Ω and Ω^{-1}. Looking at the constraints on the inverse, we can discard the scalar factor $\det \Omega$, which leaves us with a quadratic equation in Ω. The resulting problem is therefore non-linear, as opposed to linear in the case of systems of equations in only one of the two conics.

5.7.2 Expressing Constraints in the Dual Conic Ω^*

We can avoid mixing constraints on the two conics by using only the dual form of each constraint. The camera constraints (5.20) and (5.21) are naturally represented in this form, as is the dual form of the scene constraints (5.3)

$$\mathbf{l}_1^T \Omega^* \mathbf{l} \stackrel{!}{=} 0$$
$$\mathbf{l}_2^T \Omega^* \mathbf{l} \stackrel{!}{=} 0.$$

Each of these enforces the orthogonality of two planes via their respective line of intersection with the plane at infinity. We may construct the three lines from available point measurements by choosing any two points (directions) on each line, to obtain

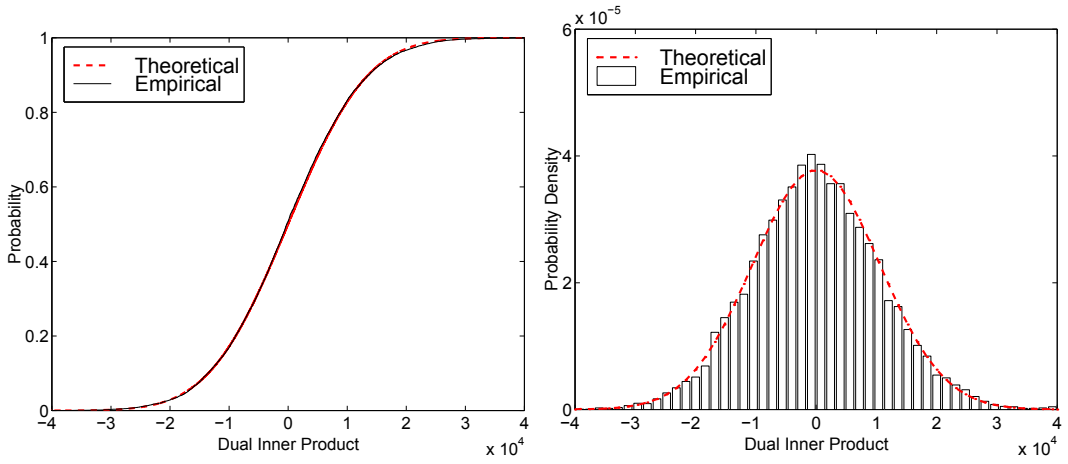
$$(\mathbf{d}_1 \times \mathbf{d}_1')^T \Omega^* (\mathbf{d} \times \mathbf{d}') = 0, \tag{5.22}$$
$$(\mathbf{d}_2 \times \mathbf{d}_2')^T \Omega^* (\mathbf{d} \times \mathbf{d}') = 0. \tag{5.23}$$

The last constraints are quadrilinear in the directions, as opposed to the bilinear constraints (5.2), but linear in the dual conic Ω^*.

5.7.3 Dual Likelihood Ratio Test

Taking into account additive measurement errors in the directions, the dual form on the left hand side of (5.22) becomes

$$(\widehat{\mathbf{d}}_k \times \widehat{\mathbf{d}}_k')^T \Omega^* (\widehat{\mathbf{d}} \times \widehat{\mathbf{d}}') \stackrel{\text{def}}{=} \left((\mathbf{d}_k + \mathbf{n}_k) \times (\mathbf{d}_k' + \mathbf{n}_k')\right)^T \Omega^* \left((\mathbf{d} + \mathbf{n}) \times (\mathbf{d}' + \mathbf{n})'\right).$$



(a) Cumulative probability distribution. (b) Probability density.

Figure 5.11: The theoretical small noise distribution (red) of the *dual inner product of two orthogonal line directions* in Euclidean space, as given by (5.25), and the empirical distribution (black) of synthetically generated direction vectors with 5% RMS added noise.

Under the assumption of small noise we are content with a first order approximation of this expression

$$\begin{aligned}(\widehat{\mathbf{d}}_k \times \widehat{\mathbf{d}}'_k)^T \Omega^* (\widehat{\mathbf{d}} \times \widehat{\mathbf{d}}') \approx\ & (\mathbf{d}_k \times \mathbf{d}'_k)^T \Omega^* (\mathbf{d} \times \mathbf{d}') \\ & + (\mathbf{n}_k \times \mathbf{d}'_k)^T \Omega^* (\mathbf{d} \times \mathbf{d}') \\ & + (\mathbf{d}_k \times \mathbf{n}'_k)^T \Omega^* (\mathbf{d} \times \mathbf{d}') \\ & + (\mathbf{d}_k \times \mathbf{d}'_k)^T \Omega^* (\mathbf{n} \times \mathbf{d}') \\ & + (\mathbf{d}_k \times \mathbf{d}'_k)^T \Omega^* (\mathbf{d} \times \mathbf{n}'),\end{aligned}$$

which has mean $\mu = (\mathbf{d}_k \times \mathbf{d}'_k)^T \Omega^* (\mathbf{d} \times \mathbf{d}')$ and variance

$$\begin{aligned}\sigma^2 = & (\mathbf{d} \times \mathbf{d}')^T \ \Omega^* [\mathbf{d}'_k]_\times \Sigma_{\mathbf{n}_k \mathbf{n}_k} [\mathbf{d}'_k]_\times{}^T \Omega^* (\mathbf{d} \times \mathbf{d}') \\ & + (\mathbf{d} \times \mathbf{d}')^T \ \Omega^* [\mathbf{d}_k]_\times \Sigma_{\mathbf{n}'_k \mathbf{n}'_k} [\mathbf{d}_k]_\times{}^T \Omega^* (\mathbf{d} \times \mathbf{d}') \\ & + (\mathbf{d}_k \times \mathbf{d}'_k)^T \Omega^* [\mathbf{d}']_\times \Sigma_{\mathbf{n}\mathbf{n}} \ [\mathbf{d}']_\times{}^T \Omega^* (\mathbf{d} \times \mathbf{d}') \\ & + (\mathbf{d}_k \times \mathbf{d}'_k)^T \Omega^* [\mathbf{d}]_\times \Sigma_{\mathbf{n}'\mathbf{n}'} \ [\mathbf{d}]_\times{}^T \Omega^* (\mathbf{d} \times \mathbf{d}')\end{aligned} \tag{5.24}$$

or for standardized errors

$$\begin{aligned}\sigma^2 = &(\mathbf{d}\times\mathbf{d}')^T\ \Omega^*[\mathbf{d}'_k]_\times[\mathbf{d}'_k]_\times{}^T\Omega^*(\mathbf{d}\times\mathbf{d}')\\ &+(\mathbf{d}\times\mathbf{d}')^T\ \Omega^*[\mathbf{d}_k]_\times[\mathbf{d}_k]_\times{}^T\Omega^*(\mathbf{d}\times\mathbf{d}')\\ &+(\mathbf{d}_k\times\mathbf{d}'_k)^T\Omega^*[\mathbf{d}']_\times\ [\mathbf{d}']_\times{}^T\Omega^*(\mathbf{d}\times\mathbf{d}')\\ &+(\mathbf{d}_k\times\mathbf{d}'_k)^T\Omega^*[\mathbf{d}]_\times\ \ [\mathbf{d}]_\times{}^T\Omega^*(\mathbf{d}\times\mathbf{d}').\end{aligned} \tag{5.25}$$

The null distribution of a *single* constraint (5.22) was compared to a normal distribution with variance given by the small noise approximation (5.25). The approximate theoretical distribution is shown plotted against the empirical distribution in Figure 5.11. The results of the Kolmogorov-Smirnov tests are given in Table 5.4 for different noise levels. From the table we infer that the approximate distribution is close when the noise level is around 10% or less. This is somewhat higher sensitivity to noise than in the case of constraints on the absolute conic, shown in Table 5.1.

noise	$\sqrt{n}D_n$	p-value	Test at $\alpha = 0.01$
1%	1.14	0.147	pass
3%	1.17	0.128	pass
10%	1.47	0.025	pass
30%	1.82	0.002	fail
100%	9.16	10^{-73}	fail

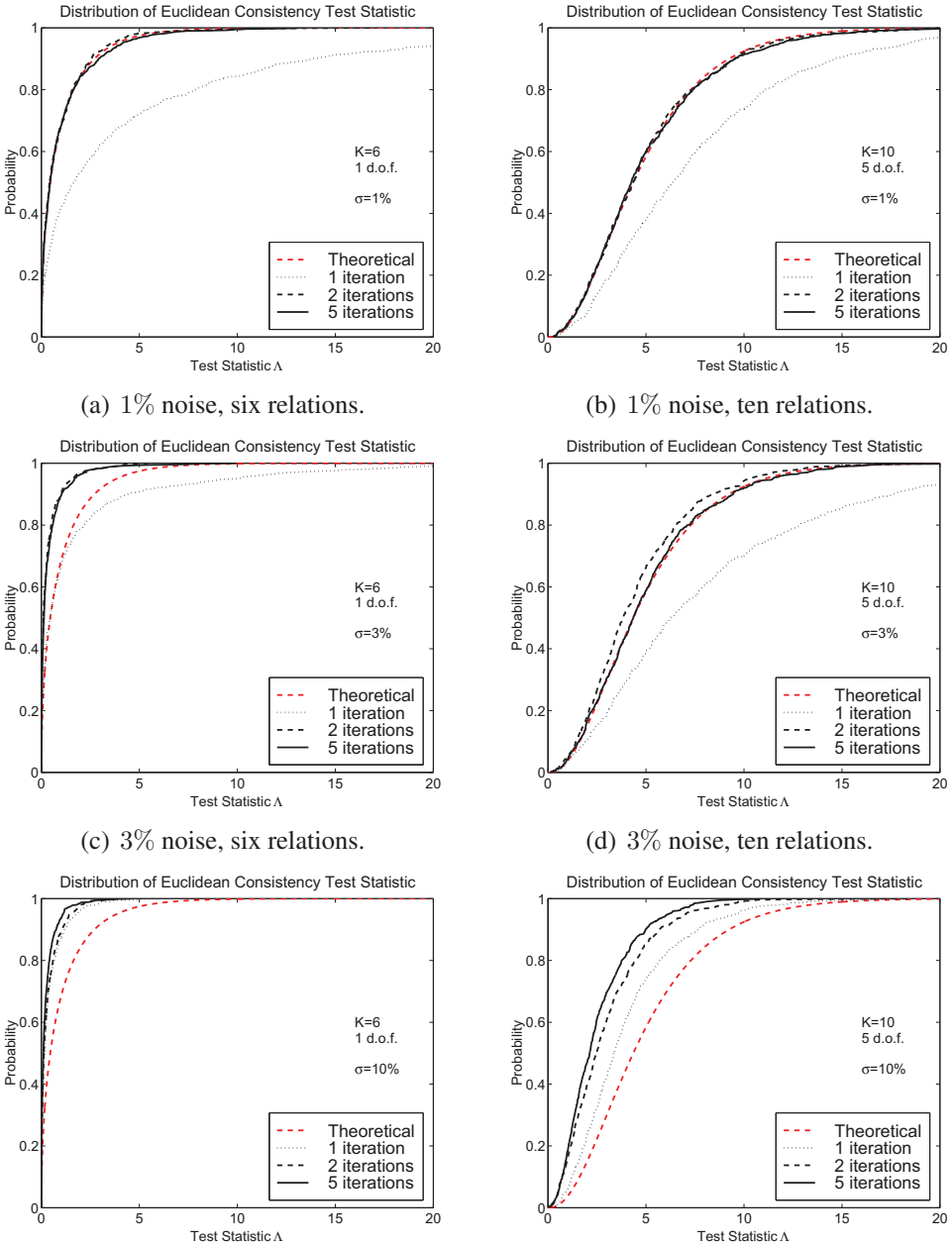
Table 5.4: Results of the Kolmogorov-Smirnov test on the null distribution of the dual scalar product given by (5.25), when $\Omega = \mathsf{I}_{3\times3}$. The data are those of Figure 5.11. The trials were repeated with different levels of RMS noise, measured relative to the magnitude of the direction vectors.

Testing Dual Relations for Consistency

Proceeding as in Section 5.4.2, we derive a test of the hypothesis that $K > 5$ independent relations between planes of the type (5.22) are indeed all orthogonality relations

$$H_0 : \mathbf{c}^* = \mathbf{0}_K \quad \text{against} \quad H_1 : \mathbf{c}^* \neq \mathbf{0}_K,$$

where we have collected the individual dual forms $c_k^* \stackrel{\text{def}}{=} (\mathbf{d}_k \times \mathbf{d}'_k)^T\Omega^*(\mathbf{d}\times\mathbf{d}')$ together in the vector $\mathbf{c}^* \stackrel{\text{def}}{=} \{c_k^*\}$. Following the derivation of (5.13), we obtain the approximate



(a) 1% noise, six relations.

(b) 1% noise, ten relations.

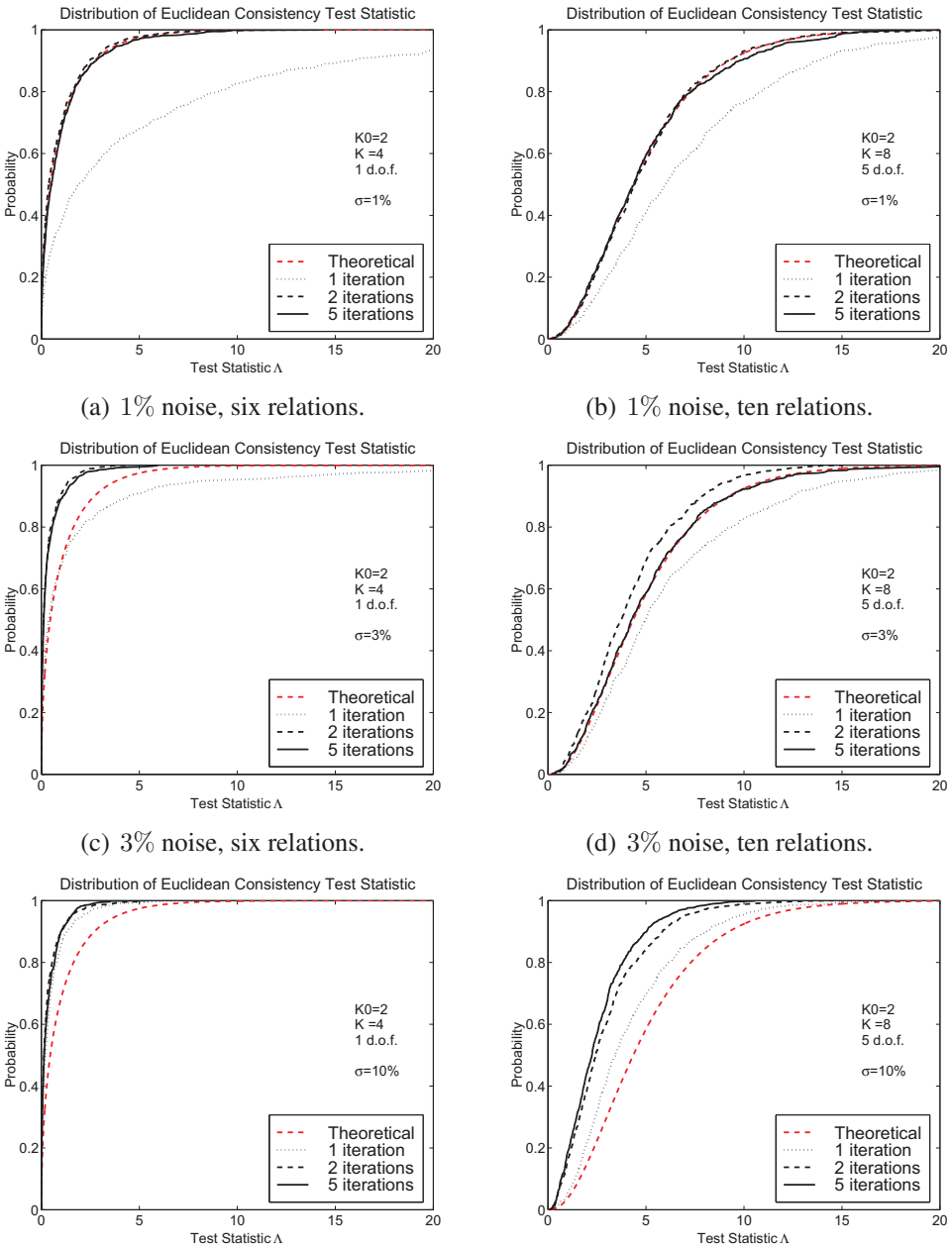
(c) 3% noise, six relations.

(d) 3% noise, ten relations.

(e) 10% noise, six relations.

(f) 10% noise, ten relations.

Figure 5.12: The empirical (black) and theoretical (red) null distribution of the test statistic (5.26) of *K orthogonal planes in an affine structure with known variance*. Each figure shows the theoretical χ^2_{K-5} distribution, as well as the empirical distribution of the test statistic after 1, 2, and 5 iterations of (5.26). The RMS noise level is measured relative to the magnitude of the direction vectors.



(a) 1% noise, six relations. (b) 1% noise, ten relations.

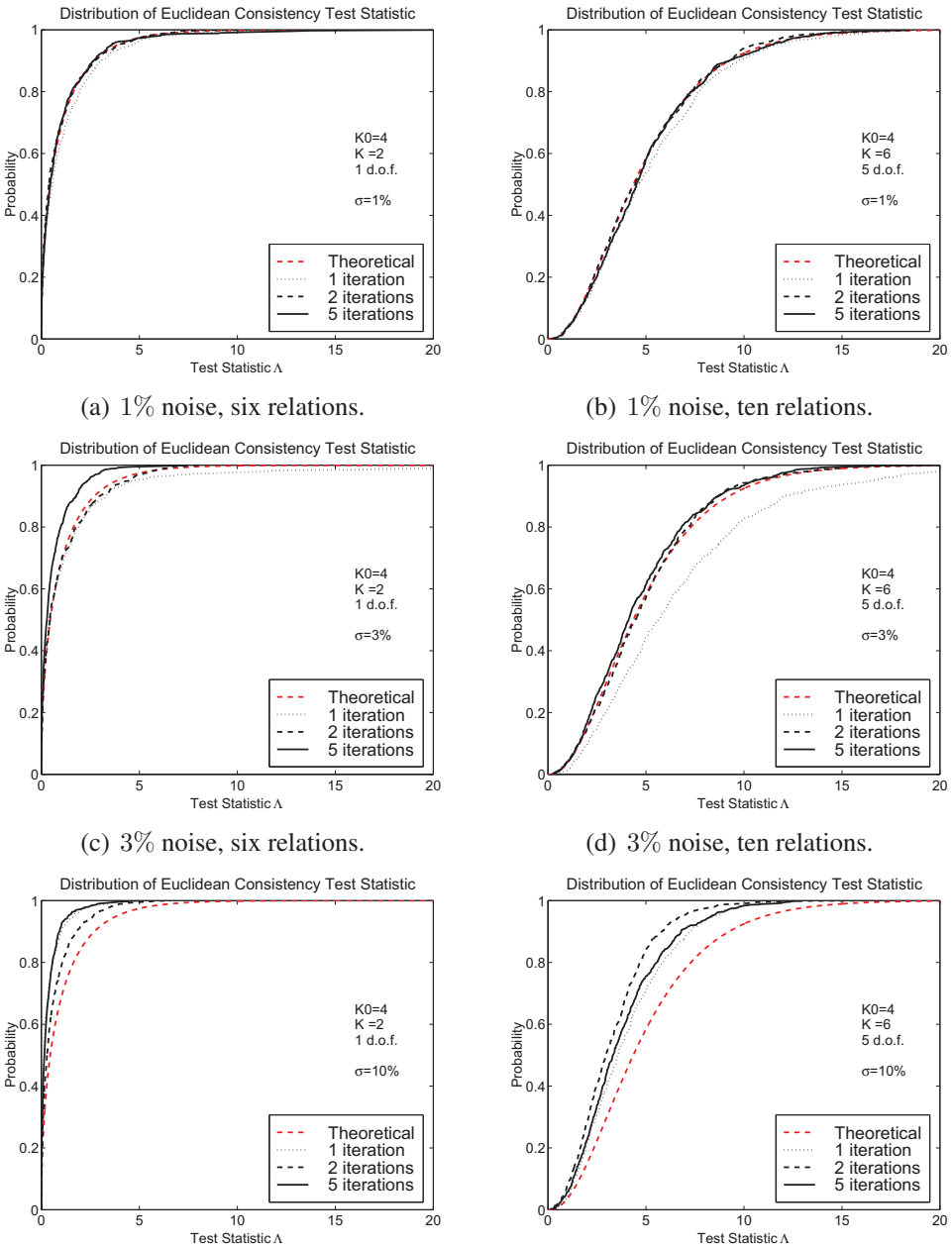
(c) 3% noise, six relations. (d) 3% noise, ten relations.

(e) 10% noise, six relations. (f) 10% noise, ten relations.

Figure 5.13: The empirical (black) and theoretical (red) null distribution of the test statistic (5.26) of K orthogonal planes in an affine structure with known variance and *two known camera constraints*.

log-likelihood ratio

$$-2\log\Lambda \approx \min_{\Omega^*} \sum_{k=1}^{K} \frac{\left((\widehat{\mathbf{d}}_k \times \widehat{\mathbf{d}}'_k)^T \Omega^* (\widehat{\mathbf{d}} \times \widehat{\mathbf{d}}')\right)^2}{\sigma_k^2}. \tag{5.26}$$



(a) 1% noise, six relations.

(b) 1% noise, ten relations.

(c) 3% noise, six relations.

(d) 3% noise, ten relations.

(e) 10% noise, six relations.

(f) 10% noise, ten relations.

Figure 5.14: The empirical (black) and theoretical (red) null distribution of the test statistic (5.26) of K orthogonal planes in an affine structure with known variance and *four known camera constraints*.

The likelihood ratio of a full set of constraints (5.26) was computed in a simple iterative scheme as in Section 5.4.4 alternating between evaluating the numerator and the denominator of (5.26) using the small noise approximation of the variance (5.25), (compare

Algorithm 5.2). The resulting statistic and the predicted chi-square distribution are shown together in Figures 5.12-5.14 for different numbers of camera constraints. The figures show that the distribution of the test statistic is well described by the chi-square distribution when the measurement errors are quite small. They indicate that a test based on (5.26) and (5.25) will be fairly accurate at 1% noise level for the minimum 6 constraints, increasing to 3% for 10 constraints.

5.8 Conclusions

This chapter has considered the problem of verifying bilateral symmetry in detected skewed symmetric sets.

1. It was shown how the orthogonality constraint between true bilateral chords and the corresponding plane of symmetry can be expressed in two linear equations parameters of the absolute conic, or equivalently its dual.

2. Three planes of symmetry in a general position provide six constraints on the absolute conic, fully determining the conic with one constraint to spare.

3. If the planes of symmetry are not in a general position the set of constraints will drop rank. The situations under which this happens were given.

4. A statistical test of the hypothesis that a set of six or more orthogonality relations are indeed consistent with a single interpretation of Euclidean space are given.

    - For all but the shortest direction vectors (with norm less than 2σ) is the distribution of their Euclidean inner product (5.5) well described by a normal distribution with mean (5.6) and variance (5.7). An unbiased estimator of the variance is (5.15).
    - The likelihood ratio of the hypothesis of K orthogonality relations in an affine structure is (5.13), which has a chi-square distribution with $K - 5$ degrees of freedom. The maximum likelihood estimate of the absolute conic in the affine frame is also provided by (5.13)

- The likelihood ratio and the MLE of the absolute conic can be computed in a succession of linear steps by alternatively estimating the numerator and denominator of (5.13).

5. It is shown how assumptions about the intrinsic parameters of the affine camera are formulated as linear orthogonality relations in the dual of the absolute conic.

6. At the same time it is shown that the usual assumptions of constant intrinsics based on physical parameters of the pin-hole camera only hold for the affine camera in the special cases of weak perspective and para-perspective.

7. Mixing linear constraints on the absolute conic and its dual results in a system of linear and quadratic equations in one of the conics.

8. Using the dual form of the pole-polar relations results in a linear system in the dual conic. The resulting estimation problem is slightly further away from the measurements than the first problem, because the expressions involve lines, or in the case of point features cross products of point, with the results that the approximate distributions are more sensitive to noise.

Part II

Affine Matching

6

Affine Multilinear Matching Constraints

6.1 Overview

In this chapter we specialize the projective unifocal, bifocal, trifocal, and quadrifocal tensors to the affine case by means of the affine camera in its original homogeneous form. We show how the tensors obtained relate to the registered tensors encountered in previous work, and obtain an affine specialization of known projective multifocal relations between points *and* lines in terms of those tensors. In Section 6.3 we discuss the registered affine camera model, and define the registered tensors. In Section 6.4 we specialize the projective tensors to the case of the affine camera, and write each tensor in terms of the registered tensors. In Section 6.5 we specialize known projective incidence relations and transfer equations to the affine case, using the results of Section 6.4. The work in this chapter was presented in [TM99] and [HTM99].

6.2 Introduction

The geometry of multiple views has been the subject of intense research over the last decade [HZ00]. Among the most useful results this effort has produced are the matching constraints between corresponding points and lines in two, three, and four *uncalibrated* views. The main advance over earlier work [LH81, SA91] is that scene structure and camera motion can be recovered without any prior knowledge of camera calibration. The resulting reconstruction is known up to a projective ambiguity which can be reduced or removed completely using relatively sparse information about the cameras or the scene. This is possible because the matching constraints are *incidence relations* between rays of sight connecting scene points, image points, and the camera centres, and as such are *projective* relations. The algebraic relations between corresponding points and lines are multilinear, and were originally characterized by the the *fundamental matrix* [Fau92, FLM92, Har92, HGC92] of two views, the *trilinearities* [Sha94] or *trifocal tensor* [VL93, VFL96, Har95b, Har97] of three views, and the *quadrifocal tensor* [Tri95a, Tri95b] of four views.

In the case of projective cameras, the multilinear relations connecting corresponding points and lines across two or more views are now well known (see [HZ00]). In a notation originated by Triggs [Tri95a, Tri95b], with recent additions by Heyden [Hey98], these relations are formulated using four tensors: the unifocal[1], bifocal, trifocal, and quadrifocal tensors. Correspondences are established across multiple views in a process involving the estimation of relevant tensors, and once the tensors are known, points and lines can be transferred between views without explicit 3D reconstruction. The tensors are overparametrized, with non-linear constraints existing between the tensor elements, a fact that somewhat hampers their estimation. While the one constraint on the elements of the fundamental matrix was known from the start [Fau92, FLM92], the corresponding constraints on the elements of the tri- and quadrifocal tensors proved elusive. Partial constraints on the trifocal tensor were given by Papadopoulo and Faugeras [PF98] and on the quadrifocal tensor by Heyden [Hey98]. Only after the publication of the current work on the affine tensors [TM99] was the minimal set of constraints on the projective tensors worked out by Canterakis [Can00] in the case of the trifocal tensor and by Shashua and Wolf [SW00] in

[1]The term *uni*focal conforms better to the Latin sequence *bi-*, *tri-*, and *quadrifocal* than the Greek *mono*focal, as Hartley has pointed out.

the case of the quadrifocal tensor.

The Affine Case

In the case of affine cameras, a corresponding set of multifocal relations has only been partially uncovered [SZB95, Tor95, QK97, QOM98]. The compact representation by registered affine tensors obtained from inhomogeneous 2×3 projection matrices, and used successfully to develop multiple-view relations between finite points and directions [QK97, QOM98] is not fully general, but has serious shortcomings in that:

- it does not immediately suggest corresponding relations between lines, and
- it fails to reveal the existence of quadrifocal relations and an affine quadrifocal tensor.

We shall see that projective tensors, specialized to the affine case [FP98, BL98, MC98], are free of these limitations.

The work presented in this chapter combines two ingredients. The first of these is the unification in a single framework of the projective multi-view tensors achieved by Triggs [Tri95a, Tri95b] with additions by Heyden [Hey98] of three entities previously studied in isolation: (1) the well known and well understood fundamental matrix of two uncalibrated views, (2) the even more useful but not fully understood trifocal tensor, and (3) a much less well understood and largely unused quadrifocal tensor. The second ingredient is the simpler form of the tensors in the case of affine cameras already pointed out by Faugeras and Papadopoulo [FP98], and a simpler form still in the case of registered affine cameras, not studied previously. Bringing together these two ingredients allows us to produce a set of simple relations connecting points and lines in multiple affine views in terms of the affine tensors. This brings about two things. Firstly, the problematic over-parameterization in the tensors is reduced to such a degree, that the few remaining constraints on the elements of the important affine trifocal tensor can be identified and given a geometric interpretation. Secondly, the estimation of the trifocal tensor is simplified. For point correspondences optimal ML estimation of all the affine tensors is achieved through factorization, with the implication that more complicated methods are only required when dealing with line correspondences.

This treatment is divided into two chapters. The current chapter thrashes out the simplified tensors and the multi-view relations, whereas the subsequent chapter investigates the constraints between tensor elements, affine triangulation, and ML estimation of the tensors. The application of these tensors to both monocular and stereo tracking is reported in [HTM99, Hay00].

Notation

For the most part we will follow Heyden's notation, as introduced in [Hey98]. An important difference is that we will denote all the unregistered tensors by capital letters, reserving lower case letters for the registered tensors defined in the next section.

6.3 The Registered Multifocal Tensors

In this section we discuss the registered affine camera model, briefly review known affine multifocal constraints, and define registered multifocal tensors in analogy to the projective tensors given in [Hey98], and discuss their shortcomings compared to the projective multifocal tensors.

Affine Camera Models

The *affine camera* P_a was defined by Mundy and Zisserman [MZ92a] as the special case $p_1^3 = p_2^3 = p_3^3 = 0$ of the projective camera, i.e.

$$\mathsf{P}_\mathrm{a} = \begin{bmatrix} p_1^1 & p_2^1 & p_3^1 & p_4^1 \\ p_1^2 & p_2^2 & p_3^2 & p_4^2 \\ 0 & 0 & 0 & p_4^3 \end{bmatrix}, \tag{6.1}$$

mapping scene points to image points in an *affine view* by the equation $\mathbf{x} = \mathsf{P}_\mathrm{a}\mathbf{X}$. The affine camera is taken to be uncalibrated, i.e. the values of its parameters are unknown.

When working with more than one affine view of a scene, it is common practice to register the coordinates in each image to the image of a common scene point, often chosen as the centroid of points visible in all views. The registered image coordinates $\Delta\mathbf{x} = \mathbf{x} - \overline{\mathbf{x}}$ are also known in the literature as *relative coordinates* or *difference vectors* [SZB95]. We will now consider the effects of registration on the camera model. Without loss of generality

we can choose the common scene point as the origin of the world frame. With the image of the origin of the world frame given by

$$\mathsf{P}_{\mathrm{a}} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_4^1 \\ p_4^2 \\ p_4^3 \end{bmatrix}$$

it follows that $p_4^1 = p_4^2 = 0$ after registration. Hence the projection matrix of the *registered camera*[2] is

$$\mathsf{P}_{\mathrm{ar}} = \begin{bmatrix} p_1^1 & p_2^1 & p_3^1 & 0 \\ p_1^2 & p_2^2 & p_3^2 & 0 \\ 0 & 0 & 0 & p_4^3 \end{bmatrix}. \tag{6.2}$$

The registered camera can be represented more compactly by the *inhomogeneous* 2×3 projection matrix

$$\mathsf{M} \stackrel{\text{def}}{=} \frac{1}{p_4^3} \begin{bmatrix} p_1^1 & p_2^1 & p_3^1 \\ p_1^2 & p_2^2 & p_3^2 \end{bmatrix}. \tag{6.3}$$

Using M instead of P_{ar}, the affine projection equation splits into two equations. For finite points, we have the familiar inhomogeneous projection equation [SZB95]

$$\Delta\mathbf{x} = \mathsf{M}\Delta\mathbf{X},$$

and for points at infinity or *directions* the homogeneous projection equation [QK97]

$$\lambda\mathbf{d} = \mathsf{M}\mathbf{D},$$

where the homogeneous 3-vector $\mathbf{D}$ represents a point on the plane at infinity and the homogeneous 2-vector $\mathbf{d}$ is a point on the line at infinity in the image.

Registered Multilinear Relations

Starting from the inhomogeneous registered camera model (6.3), existing literature on the tensors of affine views [QK97, KH98, QOM98] introduces multilinear relations between corresponding image points based on the existence of a nontrivial solution $\left[\begin{smallmatrix} \Delta\mathbf{X} \\ -\lambda \end{smallmatrix}\right]$ to the system

$$\begin{bmatrix} \mathsf{M}_I & \Delta\mathbf{x}_I \\ \mathsf{M}_J & \Delta\mathbf{x}_J \\ \mathsf{M}_K & \Delta\mathbf{x}_K \end{bmatrix} \begin{bmatrix} \Delta\mathbf{X} \\ -\lambda \end{bmatrix} = \mathbf{0}_6,$$

[2] In order to simplify the terminology it is understood that *registered* and *unregistered* entities are affine.

where the capital letters I, J, and K label the different views, and similarly for the system of directions

$$\begin{bmatrix} \mathsf{M}_I & \mathbf{d}_I & 0 & 0 \\ \mathsf{M}_J & 0 & \mathbf{d}_J & 0 \\ \mathsf{M}_K & 0 & 0 & \mathbf{d}_K \end{bmatrix} \begin{bmatrix} \mathbf{D} \\ -\lambda_I \\ -\lambda_J \\ -\lambda_K \end{bmatrix} = \mathbf{0}_6.$$

The matrix in either system must have a nontrivial null-space, and the multilinear relations are obtained by observing that every maximal minor of the matrix must therefore be zero. It follows by Laplace expansion that the coefficients of the multilinear relations, are the 3×3 minors of the *inhomogeneous joint projection matrix* [QK97]

$$\begin{bmatrix} \mathsf{M}_I \\ \mathsf{M}_J \\ \mathsf{M}_K \end{bmatrix}. \tag{6.4}$$

It has been pointed out [QOM98] that the minors come in two types, depending on whether they involve rows from two or three views. We will now proceed to define registered tensors, which account for all $\binom{6}{3} = 20$ minors.

The Registered Unifocal Tensor

In analogy to Heyden's definition for projective cameras [Hey98] we introduce the *registered unifocal tensor*

$${}_{IJ}e^j \stackrel{\text{def}}{=} \begin{vmatrix} \mathbf{m}_I^1 \\ \mathbf{m}_I^2 \\ \mathbf{m}_J^j \end{vmatrix}, \quad j \in \{1, 2\}. \tag{6.5}$$

This tensor consists of the two components of the direction to the epipole of camera I in view J, which in an affine view lies on the line at infinity. Between three views there are six registered unifocal tensors, corresponding to the six epipoles, containing a total of 12 minors.

The Registered Trifocal Tensor

We define the *registered trifocal tensor* as

$${}_{IJK}t^{ijk} \stackrel{\text{def}}{=} \begin{vmatrix} \mathbf{m}_I^i \\ \mathbf{m}_J^j \\ \mathbf{m}_K^k \end{vmatrix}, \quad i, j, k \in \{1, 2\}. \tag{6.6}$$

There is only one such tensor between three views, since reordering the views only affects the overall sign of the tensor. This tensor consists of the remaining 8 minors of (6.4). The registered trifocal tensor ${}^{IJK}t$ is the *dual* [3] of the *trilinear tensor of three views* ${}_{IJK}t$ introduced in [QK97] for use in the trilinear constraint of affine line directions.

Shortcomings of the Registered Tensor Approach

We have just seen that the 20 minors of the inhomogenous joint projection matrix (6.4) define the multilinear relations embodied by the registered unifocal and trifocal tensors. Comparing this to the projective tensors given in [Hey98] we note the absence of both an affine bifocal tensor, which would correspond to the well known *affine fundamental matrix* [SZB95], and an affine counterpart to the quadrifocal tensor known to exist between four projective views. The reason for both these omissions being, as we shall see in the next section, that the corresponding relationships involve four rows of the joint projection matrix and are therefore not captured by 3×3 minors. Knowing that the former exists, we look for an alternative approach to arriving at the affine multifocal tensors.

6.4 Affine Specialization of Projective Multifocal Tensors

The projective multilinear matching relations are obtained in the same way as the registered relations outlined in the previous section, by observing that every maximal minor of the matrix in the system

$$\begin{bmatrix} \mathsf{P}_I & \mathbf{x}_I & 0 & 0 & 0 \\ \mathsf{P}_J & 0 & \mathbf{x}_J & 0 & 0 \\ \mathsf{P}_K & 0 & 0 & \mathbf{x}_K & 0 \\ \mathsf{P}_L & 0 & 0 & 0 & \mathbf{x}_K \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ -\lambda_I \\ -\lambda_J \\ -\lambda_K \\ -\lambda_L \end{bmatrix} = \mathbf{0}_{12}$$

must be zero, since the system has a nontrivial solution. The resulting multilinear constraints have as coefficients the 4×4 minors of the joint projection matrix

$$\begin{bmatrix} \mathsf{P}_I \\ \mathsf{P}_J \\ \vdots \end{bmatrix}, \tag{6.7}$$

[3] See A Note on Dualization on page 156.

constituting the multifocal tensors of projective views. When the projection matrices are those of affine cameras, the minors of (6.7) have special properties, since the bottom row of each affine projection matrix (6.1) has the form

$$\begin{bmatrix} 0 & 0 & 0 & \times \end{bmatrix}.$$

We will distinguish between three types of minors, depending on the number of "bottom rows" a minor contains.

- If there are *two or more* bottom rows in a minor, the minor must vanish, since the bottom rows are linearly dependent, as pointed out in [FP98]. We will call this type of minor a *vanishing minor*.

- If, on the other hand, there is *exactly one* bottom row in the minor, e.g.

$$\begin{vmatrix} 0 & 0 & 0 & \times \\ + & + & + & \times \\ + & + & + & \times \\ + & + & + & \times \end{vmatrix},$$

  then we can apply Laplace expansion along that row, to reduce the minor to a scalar multiple of the left hand 3×3 minor (elements indicated by '$+$'), which by (6.1) and (6.3) is exactly one of the 3×3 minors of the inhomogeneous joint projection matrix (6.4), and hence a registered tensor, by definition (6.5) and (6.6). We will call this type of minor *reducible*.

- Finally, if there are no bottom rows in the minor, we cannot apply the above reduction, and we will call such a minor *irreducible*.

These are the three types of minors for unregistered affine cameras. For registered cameras, the irreducible minors are also zero, since the form of (6.2) implies that that the last column of those minors is identically zero.

6.4.1 The Unifocal Tensor

The *unifocal* (or monofocal) tensor for projective cameras consisting of those 4×4 minors of the joint projection matrix (6.7) containing all three rows from camera I and a single

row from camera J was introduced in [Hey98]. For clarity of the following exposition, we use a normalized form defined as

$$ {}_{IJ}E^j \stackrel{\text{def}}{=} (p_I \cdot p_I \cdot p_I \cdot p_J)^{-1} \cdot \begin{vmatrix} \mathbf{p}_I^1 \\ \mathbf{p}_I^2 \\ \mathbf{p}_I^3 \\ \mathbf{p}_J^j \end{vmatrix}, \quad j \in \{1,2,3\}, \tag{6.8} $$

where p_I refers to the element p_4^3 of P_I. We find that the minor ${}_{IJ}E^3$ vanishes, while the remaining two minors are reducible.

Comparison with the definition (6.5) reveals that the components

$$ {}_{IJ}E^j = \frac{(-1)^{3+4} \cdot p_I}{p_I \cdot p_I \cdot p_I \cdot p_J} \begin{vmatrix} p_I^{11} & p_I^{12} & p_I^{13} \\ p_I^{21} & p_I^{22} & p_I^{23} \\ p_J^{j1} & p_J^{j2} & p_J^{j3} \end{vmatrix} = - \begin{vmatrix} \mathbf{m}_I^1 \\ \mathbf{m}_I^2 \\ \mathbf{m}_J^j \end{vmatrix}, \quad j \in \{1,2\}, $$

are equal, but with opposite sign, to the corresponding components ${}_{IJ}e^j$ of the registered unifocal tensor ${}_{IJ}e$, giving

$$ \boxed{ {}_{IJ}E^j = -{}_{IJ}e^j, \quad j \in \{1,2\}.} \tag{6.9} $$

There is no irreducible minor. As in the projective case [Hey98], the three numbers ${}_{IJ}E^j$ are the homogeneous coordinates of the epipole of camera I in image J. For an affine camera the epipole is on the line at infinity, hence the vanishing third coordinate. It follows that the two components of the registered tensor ${}_{IJ}e$ are the coordinates of the *direction* to the epipole of camera I in image J.

6.4.2 The Bifocal Tensors

We use a normalized form of the two *bifocal tensors* of two views [Hey98] consisting of those 4×4 minors of the joint projection matrix (6.7) containing two rows from camera I and camera J respectively, defined as

$$ {}_{IJ}F_{ij} \stackrel{\text{def}}{=} (p_I \cdot p_I \cdot p_J \cdot p_J)^{-1} \cdot \frac{\epsilon_{ii'i''}}{2} \frac{\epsilon_{jj'j''}}{2} \begin{vmatrix} \mathbf{p}_I^{i'} \\ \mathbf{p}_I^{i''} \\ \mathbf{p}_J^{j'} \\ \mathbf{p}_J^{j''} \end{vmatrix} \tag{6.10} $$

using the 3-valent permutation tensor $\epsilon_{ii'i''}$ and the Einstein summation convention [Zis96, Har97]. It follows that for affine cameras four entries vanish

$$ {}_{IJ}F_{ij} = 0, \quad i,j \in \{1,2\}\,. $$

Of the remaining entries all except the entry ${}_{IJ}F_{33}$ are reducible. By dualizing the unifocal tensors, we obtain a simple tensor form for the reducible components

$$\boxed{\begin{array}{lll} {}_{IJ}F_{i3} = & -{}^{JI}e_i, & i \in \{1,2\}, \\ {}_{IJ}F_{3j} = & -{}^{IJ}e_j, & j \in \{1,2\}. \end{array}} \tag{6.11}$$

We write the bifocal tensor in matrix form to show the relation to the registered unifocal tensors ${}_{IJ}e$ and ${}_{JI}e$

$$\{{}_{IJ}F_{ij}\} = \begin{bmatrix} 0 & 0 & -{}_{JI}e^2 \\ 0 & 0 & {}_{JI}e^1 \\ -{}_{IJ}e^2 & {}_{IJ}e^1 & {}_{IJ}F_{33} \end{bmatrix}. \tag{6.12}$$

This is the affine fundamental matrix between views J and I, given in [SZB95] in a different formulation. In the case of registered cameras the irreducible minor

$${}_{IJ}F_{33} = (p_I \cdot p_I \cdot p_J \cdot p_J)^{-1} \cdot \begin{vmatrix} \mathbf{p}_I^1 \\ \mathbf{p}_I^2 \\ \mathbf{p}_J^1 \\ \mathbf{p}_J^2 \end{vmatrix}$$

is zero.

6.4.3 The Trifocal Tensors

We use a normalized form of each of the three *trifocal tensors* of three views [Tri95b, Hey98], consisting of those 4×4 minors of the joint projection matrix (6.7) containing rows from all three cameras I, J, and K, defined as

$${}^{JK}_{I}T^{jk}_{i} \stackrel{\text{def}}{=} (p_I \cdot p_I \cdot p_J \cdot p_K)^{-1} \cdot \frac{\epsilon_{ii'i''}}{2} \begin{vmatrix} \mathbf{p}_I^{i'} \\ \mathbf{p}_I^{i''} \\ \mathbf{p}_J^{j} \\ \mathbf{p}_K^{k} \end{vmatrix} \tag{6.13}$$

In the case of affine cameras, 11 out of the 27 components in the trifocal tensor vanish. This is also known from [FP98]. We can divide the 12 reducible entries into two groups. The eight entries $i, j, k \in \{1,2\}$ are up to sign the components of the registered trifocal tensor ${}^{IJK}t$, according the following mapping of indices

$$\begin{array}{lll} {}^{JK}_{I}T^{jk}_{1} = & {}^{IJK}t^{2jk}, & j,k \in \{1,2\} \\ {}^{JK}_{I}T^{jk}_{2} = & -{}^{IJK}t^{1jk}, & j,k \in \{1,2\}. \end{array}$$

The two equations can be written more compactly as

$$\boxed{{}_{I}^{JK}T_{i}^{jk} = {}_{I}^{JK}t_{i}^{jk}, \qquad i,j,k \in \{1,2\},} \tag{6.14}$$

by dualizing ${}^{IJK}t$ in the first index. The following four entries are up to sign those of the registered unifocal tensors ${}_{IJ}e$ and ${}_{IK}e$

$$\boxed{\begin{array}{ll} {}_{I}^{JK}T_{3}^{j3} = & {}_{IJ}e^{j}, \qquad j \in \{1,2\}, \\ {}_{I}^{JK}T_{3}^{3k} = & -{}_{IK}e^{k}, \qquad k \in \{1,2\}, \end{array}} \tag{6.15}$$

representing the epipole of camera I in views J and K. This leaves four irreducible entries

$${}_{I}^{JK}T_{3}^{jk} = (p_I \cdot p_I \cdot p_J \cdot p_J)^{-1} \cdot \begin{vmatrix} \mathbf{p}_I^1 \\ \mathbf{p}_I^2 \\ \mathbf{p}_J^j \\ \mathbf{p}_K^k \end{vmatrix}, \quad j,k \in \{1,2\}.$$

The Trifocal Tensor and the Reduced Affine Quasi-Tensor

Equations (6.14)-(6.15) show that the trifocal tensor ${}_{I}^{JK}T$ for registered cameras consists exactly of the components of the three registered tensors ${}^{IJK}t$, ${}_{IJ}e$, and ${}_{IK}e$, the remaining components being zero. The same three tensors also constitute the 12 component *reduced affine quasi-tensor* of three views, defined by Kahl and Heyden [KH98] without reference to the trifocal tensor. It follows that the reduced affine quasi-tensor and the trifocal tensor ${}_{I}^{JK}T$ for registered cameras are equivalent descriptions of the geometry of three affine views. The trifocal tensor has the advantages of being a valid tensor, having a clear relation to a corresponding projective entity, and consequently being immediately applicable to well known projective incidence and transfer equations, as is borne out by Section 6.5. Since there is only one *registered* trifocal tensor ${}^{IJK}t$ between three views (Section 6.3), we can obtain the trifocal tensors ${}_{J}^{KI}T$, and ${}_{K}^{IJ}T$ in addition to ${}_{I}^{JK}T$, by supplying only the registered unifocal tensors ${}_{JK}e$, ${}_{JI}e$, and ${}_{KI}e$, ${}_{KJ}e$ respectively. The additional 8 numbers in the four unifocal tensors complete the 20 component *affine quasi-tensor* of three views, also defined in [KH98].

6.4.4 The Quadrifocal Tensor

We use a normalized form of the *quadrifocal tensor* [Tri95b, Hey98], consisting of those 4×4 minors of the joint projection matrix (6.7) containing one row from each of the four

cameras I, J, K, and L, defined as

$$ {}^{IJKL}Q^{ijkl} \stackrel{\text{def}}{=} (p_I \cdot p_J \cdot p_K \cdot p_L)^{-1} \cdot \begin{vmatrix} \mathbf{p}_I^i \\ \mathbf{p}_J^j \\ \mathbf{p}_K^k \\ \mathbf{p}_L^l \end{vmatrix}, \quad i,j,k,l \in \{1,2,3\}. \tag{6.16} $$

Of the 81 components of the affine quadrifocal tensor, 33 components vanish. From the definition (6.6) we obtain

$$ {}^{IJKL}Q^{ijk3} = (-1)^{4+4} \cdot \begin{vmatrix} \mathbf{m}_I^i \\ \mathbf{m}_J^j \\ \mathbf{m}_J^k \end{vmatrix} = {}^{IJK}t^{ijk}, \quad i,j,k \in \{1,2\}, $$

and similarly for the other reducible minors, giving

$$ \boxed{\begin{array}{lll} {}^{IJKL}Q^{ijk3} = & {}^{IJK}t^{ijk}, & i,j,k \in \{1,2\}, \\ {}^{IJKL}Q^{ij3l} = & -{}^{IJL}t^{ijl}, & l,i,j \in \{1,2\}, \\ {}^{IJKL}Q^{i3kl} = & {}^{IKL}t^{ikl}, & k,l,i \in \{1,2\}, \\ {}^{IJKL}Q^{3jkl} = & -{}^{JKL}t^{jkl}, & j,k,l \in \{1,2\}, \end{array}} \tag{6.17} $$

a total of $\binom{4}{1}\,2^3 = 32$ components. The remaining 16 minors ${}^{IJKL}Q^{ijkl}, i,j,k,l \in \{1,2\}$ are irreducible. For registered cameras, the irreducible components are zero, leaving only the 32 components of the four registered trifocal tensors. We know from [FM95] that the projective quadrifocal tensor is completely specified by the projective trifocal and unifocal tensors. The simple form of (6.17) however is special to the affine case.

Reducible Minors of the Affine Multifocal Tensors

Unifocal Tensor (6.9)

$$ {}_{IJ}E^{j} = -{}_{IJ}e^{j}, \qquad j \in \{1,2\}. $$

Bifocal Tensor (6.11)

$$ \begin{aligned} {}_{IJ}F_{i3} &= -{}^{JI}e_{i}, & i \in \{1,2\}, \\ {}_{IJ}F_{3j} &= -{}^{IJ}e_{j}, & j \in \{1,2\}. \end{aligned} $$

Trifocal Tensor (6.14), (6.15)

$$ \begin{aligned} {}^{JK}_{I}T^{jk}_{i} &= {}^{JK}_{I}t^{jk}_{i}, & i,j,k \in \{1,2\}, \\ {}^{JK}_{I}T^{j3}_{3} &= {}_{IJ}e^{j}, & j \in \{1,2\}, \\ {}^{JK}_{I}T^{3k}_{3} &= -{}_{IK}e^{k}, & k \in \{1,2\}. \end{aligned} $$

Quadrifocal Tensor (6.17)

$$ \begin{aligned} {}^{IJKL}Q^{ijk3} &= {}^{IJK}t^{ijk}, & i,j,k \in \{1,2\}, \\ {}^{IJKL}Q^{ij3l} &= -{}^{IJL}t^{ijl}, & l,i,j \in \{1,2\}, \\ {}^{IJKL}Q^{i3kl} &= {}^{IKL}t^{ikl}, & k,l,i \in \{1,2\}, \\ {}^{IJKL}Q^{3jkl} &= -{}^{JKL}t^{jkl}, & j,k,l \in \{1,2\}. \end{aligned} $$

Table 6.1: The reducible minors of the affine multifocal tensors can be written in terms of the registered tensors of section 6.3. These are the only non-zero elements of the affine multifocal tensors when the cameras are registered.

Affine Specialization of the Multifocal Tensors

Unifocal Tensor (6.9)

2 reducible minors	${}_{IJ}E^{j}$	$=$	$-{}_{IJ}e^{j},$	$j \in \{1,2\}.$

Bifocal Tensor (6.11)

4 reducible minors	${}_{IJ}F_{i3}$	$=$	$-{}^{JI}e_{i},$	$i \in \{1,2\},$
	${}_{IJ}F_{3j}$	$=$	$-{}^{IJ}e_{j},$	$j \in \{1,2\}.$
1 irreducible minor	${}_{IJ}F_{33}.$			

Trifocal Tensor (6.14), (6.15)

12 reducible minors	${}^{JK}_{I}T^{jk}_{i}$	$=$	${}^{JK}_{I}t^{jk}_{i},$	$i,j,k \in \{1,2\},$
	${}^{JK}_{I}T^{j3}_{3}$	$=$	${}_{IJ}e^{j},$	$j \in \{1,2\},$
	${}^{JK}_{I}T^{3k}_{3}$	$=$	$-{}_{IK}e^{k},$	$k \in \{1,2\}.$
4 irreducible minors	${}^{JK}_{I}T^{jk}_{3},$			$j,k \in \{1,2\}.$

Quadrifocal Tensor (6.17)

32 reducible minors	${}^{IJKL}Q^{ijk3}$	$=$	${}^{IJK}t^{ijk},$	$i,j,k \in \{1,2\},$
	${}^{IJKL}Q^{ij3l}$	$=$	$-{}^{IJL}t^{ijl},$	$l,i,j \in \{1,2\},$
	${}^{IJKL}Q^{i3kl}$	$=$	${}^{IKL}t^{ikl},$	$k,l,i \in \{1,2\},$
	${}^{IJKL}Q^{3jkl}$	$=$	$-{}^{JKL}t^{jkl},$	$j,k,l \in \{1,2\}.$
16 irreducible minors	${}^{IJKL}Q^{ijkl},$			$i,j,k,l \in \{1,2\}.$

Table 6.2: The projective multifocal tensors specialized to the case of affine cameras. Of the minors constituting the tensors, the reducible minors can be written in terms of the registered tensors of Section 6.3, whereas the irreducible minors can not. All other minors vanish.

6.5 Transfer Equations and Incidence Relations

In this section we specialize the various projective transfer equations and incidence relations between two, three, and four views [Har97, Zis96, Hey98] to affine cameras, using the results of the previous section. For clarity of the following exposition we will give the slightly simpler expressions attainable for registered cameras. The corresponding expressions for registered cameras are obtained in the same manner by retaining the irreducible entries of the tensors, as is shown below on the example of the affine bifocal tensor.

A Note on Dualization

For the purpose of carrying results between dual entitites such as image points and line directions, we introduce the bivalent permutation tensor. We dualize a bivalent tensor in any one index by contracting it with the *bivalent permutation tensor* $\epsilon_{ii'}$, defined by

$$\epsilon_{12} = 1,\ \epsilon_{21} = -1,\ \text{and}\ \epsilon_{11} = \epsilon_{22} = 0, \tag{6.18}$$

by analogy to the well known *trivalent permutation tensor* or *Levi-Civita symbol* $\epsilon_{ii'i''}$ encountered in the last section. Dualization in an index has the effect of moving the index from covariant to contravariant position or vice versa. Dualizing the tensor ${}^{IJK}t$ in the first index thus produces the tensor ${}_{I}^{JK}t$, according to

$${}_{I}^{JK}t_{i'}^{jk} \stackrel{\text{def}}{=} {}^{IJK}t^{ijk} \cdot \epsilon_{ii'}.$$

The tensor does not incur any loss of information under the operation.

The following properties of bivalent dualization are easily verified:

$$a^i \cdot a_i = 0, \tag{6.19}$$

$$a^i \cdot b_i = a_i \cdot b^i. \tag{6.20}$$

The following connection exists between the bi- and trivalent permutation tensors

$$\begin{aligned} \epsilon_{ij3} &= \epsilon_{ij}, \\ \epsilon_{i3k} &= \epsilon_{ki} = -\epsilon_{ik}, \\ \epsilon_{3jk} &= \epsilon_{jk}. \end{aligned} \tag{6.21}$$

Dualization of a bivalent tensor in any one index i, referred to by other authors (e.g. [QK97]), is thus conveniently described by a contraction with the *bivalent permutation tensor* $\epsilon_{ii'}$.

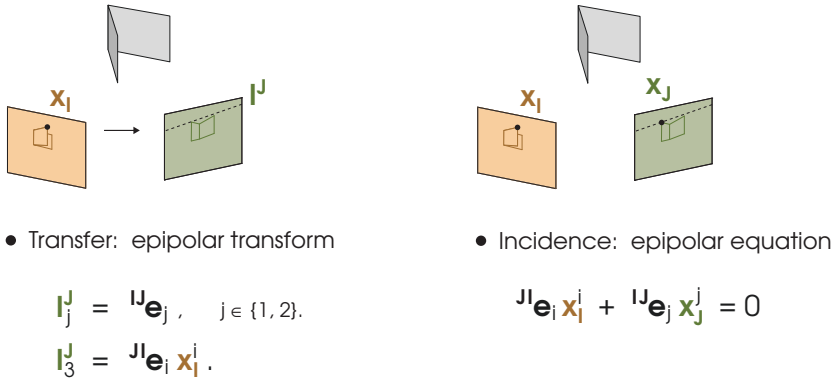


Figure 6.1: Bifocal transfer and incidence, as described by equations (6.22) and (6.24) respectively, representing the epipolar transform and epipolar constraint in tensor form.

6.5.1 Bifocal Transfer Equations

The projective transfer equation [Hey98]

$$l_j^J = {}_{IJ}F_{ij}x_I^i, \quad i \in \{1,2,3\},$$

gives the epipolar line in image J corresponding to a point in image I. Specialized to registered cameras, using (6.11), it becomes

$$\begin{aligned} l_j^J &= {}^{IJ}e_j x_I^3, \quad j \in \{1,2\}, \\ l_3^J &= {}^{JI}e_i x_I^i. \end{aligned}$$

If the point x_I is in inhomogeneous coordinates (implying $x_I^3 = 1$) then these equations simplify to

$$\boxed{\begin{aligned} l_j^J &= {}^{IJ}e_j, \quad j \in \{1,2\}, \\ l_3^J &= {}^{JI}e_i x_I^i. \end{aligned}} \tag{6.22}$$

The first of the two equations gives the fixed direction of the pencil of epipolar lines in view J, as dictated by the direction ${}_{IJ}e$ to the epipole (through ${}^{IJ}e_j \cdot {}_{IJ}e^j = 0$), whereas the second equation selects a particular line from the pencil (see Figure 6.1). The above equation is considerably simpler than the original equation given in [SZB95].

6.5.2 Bifocal Incidence Relations

Using (6.11), the projective incidence relation [Hey98]

$$ {}_{IJ}F_{ij}x_I^i x_J^j = 0 \tag{6.23}$$

becomes

$$ {}^{JI}e_i x_I^i x_J^3 + {}^{IJ}e_j x_J^j x_I^3 = 0, $$

for registered cameras. When used with inhomogeneous coordinates (implying $x_I^3 = x_J^3 = 1$) this equation simplifies to

$$ \boxed{{}^{JI}e_i x_I^i + {}^{IJ}e_j x_J^j = 0.} \tag{6.24}$$

This is the *registered epipolar constraint* [SZB95] in tensor form (see Figure 6.1). Compared to its form in [SZB95], the above equation has the advantage of making explicit the simple relation to the pencils of epipolar lines.

In comparison the unregistered epipolar constraint, obtained from (6.23) using (6.12), becomes

$$ {}^{JI}e_i x_I^i + {}^{IJ}e_j x_J^j + {}_{IJ}F_{33} = 0, $$

which we see, by expanding ${}_{IJ}F_{33}$ in terms of the registered tensors (equation (B.1) of Appendix B.1), is exactly

$$ {}^{JI}e_i \cdot (x_I^i - m_I^{i4}) + {}^{IJ}e_j \cdot (x_J^j - m_J^{j4}) = 0. \tag{6.25}$$

6.5.3 Trifocal Transfer Equations

Transfer of Complete Lines

The projective line transfer equation [Har97]

$$ l_i^I = {}_I^{JK}T_i^{jk} l_j^J l_k^K, \qquad i \in \{1,2,3\}, $$

is specialized to affine cameras, using (6.14)-(6.15), to obtain

$$\begin{aligned} l_i^I &= {}_I^{JK}t_i^{jk} l_j^J l_k^K, & i \in \{1,2\}, \\ l_3^I &= {}_I^{JK}t_3^{jk} l_j^J l_k^K + {}_{IJ}e^j l_j^J l_3^K - {}_{IK}e^k l_3^J l_k^K. \end{aligned} \tag{6.26}$$

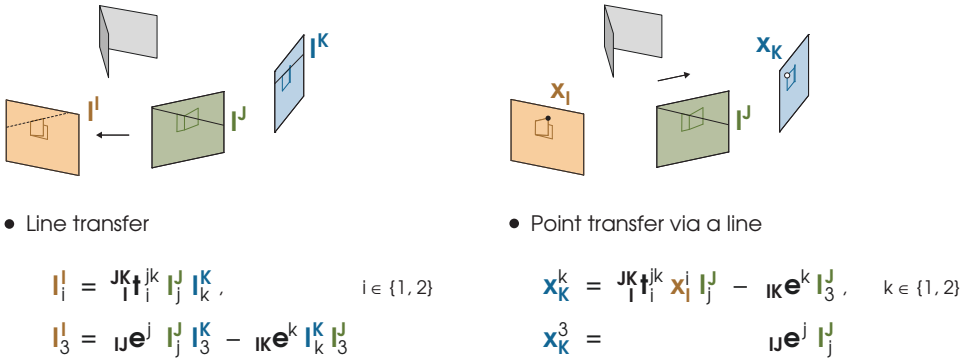


Figure 6.2: Trifocal line and point transfer, as described by equations (6.27) and (6.30) respectively.

which has the registered form

$$\boxed{\begin{aligned} l_i^I &= {}_I^{JK}t_i^{jk} l_j^J l_k^K, & i \in \{1,2\}, \\ l_3^I &= {}_{IJ}e^j l_j^J l_3^K - {}_{IK}e^k l_3^J l_k^K. \end{aligned}} \tag{6.27}$$

These equations transfer the whole line, as depicted in Figure 6.2, rather than just the line direction. They project the linear intersection in the scene of the two planes defined by the image lines $\mathbf{l}_J$ and $\mathbf{l}_K$ and their respective centre of projection into the first view as the line $\mathbf{l}_I$. They are the tensor form of the *affine line transfer equations* first given by Torr [Tor95]. The above form makes explicit the relation to the registered trifocal tensor and the epipoles.

Transfer of Line Directions

We note in (6.27) that the orientation of the transferred line encoded by its first two components is transferred by the registered trifocal tensor, while the offset or finite part of the line is transferred by the epipoles. For line normals (however meaningful in an affine space) we have $l_3^I = l_3^J = l_3^K = 0$, so that (6.27) is reduced to

$$l_i^I = {}_I^{JK}t_i^{jk} l_j^J l_k^K, \quad i \in \{1,2\},$$

and by dualizing we obtain

$$\boxed{d_I^i = {}_{JK}^I t_{jk}^i d_J^j d_K^k, \quad i \in \{1,2\},} \tag{6.28}$$

which is the transfer equation of affine line directions obtained by Quan and Kanade [QK97].

Point Transfer via a Line

The projective equations for point transfer via a line [Zis96, Hey98]

$$x_J^j = {}_I^{JK}T_i^{jk} x_I^i l_k^K,$$
$$x_K^k = {}_I^{JK}T_i^{jk} x_I^i l_j^J$$

similarly become

$$\begin{aligned} x_J^j &= {}_I^{JK}t_i^{jk} x_I^i l_k^K \;+\; {}_{IJ}e^j x_I^3 l_3^K, \quad j \in \{1,2\}, \\ x_J^3 &= \qquad\qquad -\; {}_{IK}e^k x_I^3 l_k^K, \end{aligned}$$

and

$$\begin{aligned} x_K^k &= {}_I^{JK}t_i^{jk} x_I^i l_j^J \;-\; {}_{IK}e^k x_I^3 l_3^J, \quad k \in \{1,2\}, \\ x_K^3 &= \qquad\qquad {}_{IJ}e^j x_I^3 l_j^J, \end{aligned}$$

or if x_I is in inhomogeneous coordinates (implying $x_I^3 = 1$) take on the slightly simpler form

$$\boxed{\begin{aligned} x_J^j &= {}_I^{JK}t_i^{jk} x_I^i l_k^K \;+\; {}_{IJ}e^j l_3^K, \quad j \in \{1,2\}, \\ x_J^3 &= \qquad\qquad -\; {}_{IK}e^k l_k^K, \end{aligned}} \tag{6.29}$$

and as depicted in Figure 6.3

$$\boxed{\begin{aligned} x_K^k &= {}_I^{JK}t_i^{jk} x_I^i l_j^J \;-\; {}_{IK}e^k l_3^J, \quad k \in \{1,2\}, \\ x_K^3 &= \qquad\qquad {}_{IJ}e^j l_j^J. \end{aligned}} \tag{6.30}$$

6.5.4 Trifocal Incidence Relations

A Point and Two Lines

The projective incidence relation of an image point and two image lines given by Hartley [Har97]

$${}_I^{JK}T_i^{jk} x_I^i l_j^J l_k^K = 0 \tag{6.31}$$

is the primary incidence relation between lines and points in three views, and forms the basis for algorithms to estimate the projective tensor [Har97, Har98]. It is specialized to

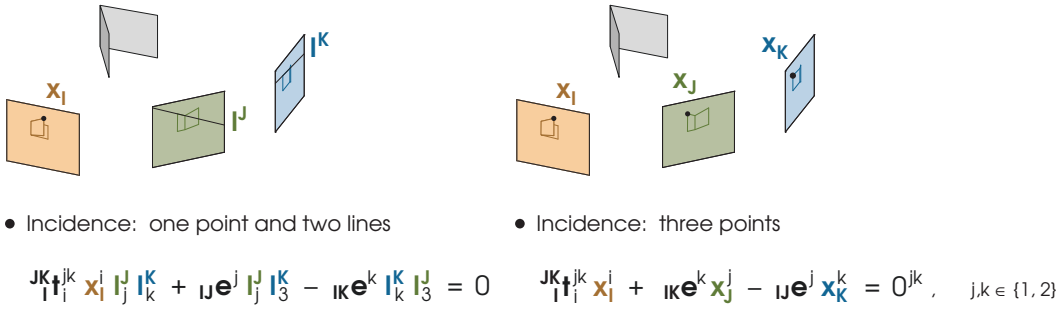


Figure 6.3: Trifocal incidence, as described by equations (6.33) and (6.41) respectively.

the registered case using (6.14)-(6.15) to give

$$ {}^{JK}_{I}t^{jk}_{i} x^i_I l^J_j l^K_k + {}_{IJ}e^j x^3_I l^J_j l^K_3 - {}_{IK}e^k x^3_I l^J_3 l^K_k = 0, \tag{6.32}$$

If x_I is in inhomogeneous coordinates (implying $x^3_I = 1$), we obtain the affine incidence relation between a point and two lines

$$\boxed{ {}^{JK}_{I}t^{jk}_{i} x^i_I l^J_j l^K_k + {}_{IJ}e^j l^J_j l^K_3 - {}_{IK}e^k l^J_3 l^K_k = 0,} \tag{6.33}$$

which is a previously unpublished relation. See Figure 6.3.

Three Image Lines

A projective incidence relation between three views of a line is obtained from (6.31) by replacing x_I by any point on the line l^I, writing $x^i_I = l^I_{i'}\, \epsilon^{ii'i''}$, to produce [Har97]

$$ {}^{JK}_{I}T^{jk}_{i} \epsilon^{ii'i''} l^I_{i'} l^J_j l^K_k = 0^{i''}, \quad i'' \in \{1,2,3\}.$$

Regrouping this expression

$$ \left({}^{JK}_{I}T^{jk}_{i} l^J_j l^K_k \right) l^I_{i'}\, \epsilon^{ii'i''} = 0^{i''}, \quad i'' \in \{1,2,3\}, \tag{6.34}$$

reveals the cross product of the line l^I and the line transferred from views J and K, which should vanish when the two coincide. These relations are specialized to the affine case as follows. In (6.34), consider first the relation corresponding to $i'' = 3$

$$ \left({}^{JK}_{I}T^{jk}_{i} l^J_j l^K_k \right) l^I_{i'}\, \epsilon^{ii'3} = 0.$$

Affine Transfer Equations

Bifocal (6.22)

$$\begin{aligned} l_j^J &= {}^{IJ}e_j, \quad j \in \{1,2\}, \\ l_3^J &= {}^{JI}e_i x_I^i. \end{aligned}$$

Trifocal (6.27), (6.28), (6.30)

$$\begin{aligned} l_i^I &= {}_I^{JK}t_i^{jk} l_j^J l_k^K, \qquad i \in \{1,2\}, \\ l_3^I &= {}_{IJ}e^j l_j^J l_3^K - {}_{IK}e^k l_3^J l_k^K. \end{aligned}$$

$$d_I^i = {}_{JK}^{I}t_{jk}^i d_J^j d_K^k, \quad i \in \{1,2\}.$$

$$\begin{aligned} x_K^k &= {}_I^{JK}t_i^{jk} x_I^i l_j^J - {}_{IK}e^k l_3^J, \quad k \in \{1,2\}, \\ x_K^3 &= {}_{IJ}e^j l_j^J. \end{aligned}$$

Quadrifocal (6.44)

$$\begin{aligned} x_I^i &= {}^{IJK}t^{ijk} l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl} l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl} l_3^J l_k^K l_l^L, \quad i \in \{1,2\}, \\ x_I^3 &= -{}^{JKL}t^{jkl} l_j^J l_k^K l_l^L. \end{aligned}$$

Table 6.3: Affine specialization of projective transfer equations across two, three, and four views.

Since the permutation tensor vanishes when any of its indices are repeated, we have $i \neq 3$ and $i' \neq 3$, so substitution by the registered tensors produces

$$\boxed{\left({}_I^{JK}t_i^{jk} l_j^J l_k^K\right) l_{i'}^I \epsilon^{ii'} = 0.}$$

This is a constraint on the *direction* of the line l^I, as can be seen by writing $d_I^i \stackrel{\text{def}}{=} l_{i'}^I \epsilon^{ii'}$, producing

$${}_I^{JK}t_i^{jk} d_I^i l_j^J l_k^K = 0. \tag{6.35}$$

Affine Incidence Relations

Bifocal (6.24)

$$ {}^{JI}e_i x_I^i + {}^{IJ}e_j x_J^j = 0. $$

Trifocal (6.33), (6.41)

$$ {}_I^{JK}t_i^{jk} x_I^i l_j^J l_k^K + {}_{IJ}e^j l_j^J l_3^K - {}_{IK}e^k l_3^J l_k^K = 0, $$

$$ {}_I^{JK}t_i^{jk} x_I^i + {}_{IK}e^k x_J^j - {}_{IJ}e^j x_K^k = 0^{jk}, \qquad j,k \in \{1,2\}. $$

Quadrifocal (6.45)

$$ {}^{IJK}t^{ijk} l_i^I l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl} l_i^I l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl} l_i^I l_3^J l_k^K l_l^L - {}^{JKL}t^{jkl} l_3^I l_j^J l_k^K l_l^L = 0. $$

Table 6.4: Affine specialization of projective incidence relations across two, three, and four views.

Incidentally, since the direction d_I is a *point* with a vanishing third coordinate, this relation is obtained immediately by substitution into (6.32). Further dualization of (6.35) produces the equivalent equation

$$ \boxed{ {}_{IJK}t_{ijk} d_I^i d_J^j d_K^k = 0. } \tag{6.36} $$

The remaining two relations ($i'' \neq 3$) in (6.34) consist of terms with either $i = 3$ or $i' = 3$. Splitting the sum accordingly, we obtain

$$ \left({}_I^{JK}T_i^{jk} l_j^J l_k^K \right) l_3^I \epsilon^{i3i''} + \left({}_I^{JK}T_3^{jk} l_j^J l_k^K \right) l_{i'}^I \epsilon^{3i'i''} = 0^{i''}, \quad i'' \in \{1,2\}, $$

Substituting the registered tensors and the bivalent permutation tensors, we obtain

$$ -\left({}_I^{JK}t_i^{jk} l_j^J l_k^K \right) l_3^I \epsilon^{ii''} + \left({}_I^{JK}t_3^{jk} l_j^J l_k^K + {}_{IJ}e^j l_j^J l_3^K - {}_{IK}e^k l_3^J l_k^K \right) l_{i'}^I \epsilon^{i'i''} = 0^{i''}, \quad i'' \in \{1,2\}, $$

which is equivalent to the constraints

$$ \boxed{ \left({}_I^{JK}t_i^{jk} l_j^J l_k^K \right) l_3^I - \left({}_I^{JK}t_3^{jk} l_j^J l_k^K + {}_{IJ}e^j l_j^J l_3^K - {}_{IK}e^k l_3^J l_k^K \right) l_i^I = 0_i, \quad i \in \{1,2\}. } $$

We recognize the bracketed expressions from the trifocal line transfer equations (6.26). Denoting the transferred line by l'^I, these relations say

$$l'^I_i l^I_3 - l'^I_3 l^I_i = 0_i \quad i \in \{1,2\}.$$

Three Image Points

The projective incidence relations between three image points [Har97]

$$ {}^{JK}_{I}T^{jk}_i x^i_I \epsilon_{jj'j''} x^{j'}_J \epsilon_{kk'k''} x^{k'}_K = 0_{j''k''}, \quad j'', k'' \in \{1,2,3\}, \tag{6.37}$$

express in tensor form the *trilinearity relationships* [Sha94], as shown in [Har97]. Of the nine relations (6.37) only four are linearly independent. A set of four independent relations is [Har97]

$$x^i_I({}^{JK}_{I}T^{33}_i x^j_J x^k_K - {}^{JK}_{I}T^{3k}_i x^j_J - {}^{JK}_{I}T^{j3}_i x^k_K + {}^{JK}_{I}T^{jk}_i) = 0^{jk}, \qquad j,k \in \{1,2\}. \tag{6.38}$$

For affine cameras the four constraints specialize to

$${}^{JK}_{I}T^{jk}_i x^i_I - {}^{JK}_{I}T^{3k}_3 x^j_J - {}^{JK}_{I}T^{j3}_3 x^k_K = 0^{jk}, \qquad j,k \in \{1,2\}. \tag{6.39}$$

In terms of the registered tensors, we have

$${}^{JK}_{I}t^{jk}_i x^i_I + {}_{IK}e^k x^j_J - {}_{IJ}e^j x^k_K + {}^{JK}_{I}T^{jk}_3 = 0^{jk}, \qquad j,k \in \{1,2\}, \tag{6.40}$$

and for registered cameras

$$\boxed{{}^{JK}_{I}t^{jk}_i x^i_I + {}_{IK}e^k x^j_J - {}_{IJ}e^j x^k_K = 0^{jk}, \qquad j,k \in \{1,2\}} \tag{6.41}$$

as depicted in Figure 6.3. It is shown in appendix B.2.1 that (6.39) are the four *affine trifocal constraints* derived by Torr [Tor95] and re-derived by Quan [QOM98], generalizing the orthographic result of Ullman and Basri [UB91]. As shown in Section 7.5.3, only three of the four relations are linearly independent.

6.5.5 Quadrifocal Transfer Equations

The projective transfer equation [Hey98]

$$x^i_I = {}^{IJKL}Q^{ijkl} l^J_j l^K_k l^L_l, \tag{6.42}$$

describes the transfer of the image of the intersection of the three planes, defined by each of the three image lines and the corresponding optical centre. Specialized to registered cameras, using (6.17), it becomes

$$\begin{aligned} x_I^i &= {}^{IJK}t^{ijk}l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl}l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl}l_3^J l_k^K l_l^L + {}^{IJKL}q^{ijkl}l_j^J l_k^K l_l^L, \quad i \in \{1,2\}, \\ x_I^3 &= -{}^{JKL}t^{jkl}l_j^J l_k^K l_l^L, \end{aligned} \tag{6.43}$$

with the registered form

$$\boxed{\begin{aligned} x_I^i &= {}^{IJK}t^{ijk}l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl}l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl}l_3^J l_k^K l_l^L, \quad i \in \{1,2\}, \\ x_I^3 &= -{}^{JKL}t^{jkl}l_j^J l_k^K l_l^L. \end{aligned}} \tag{6.44}$$

6.5.6 Quadrifocal Incidence Relations

The projective incidence relation [Tri95b, Tri95a]

$${}^{IJKL}Q^{ijkl}l_i^I l_j^J l_k^K l_l^L = 0, \tag{6.45}$$

is satisfied if the four space planes, defined by the four image lines (and the corresponding centre of projection), intersect in a common point. The relation is specialized, using (6.17), to

$$\boxed{{}^{IJK}t^{ijk}l_i^I l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl}l_i^I l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl}l_i^I l_3^J l_k^K l_l^L - {}^{JKL}t^{jkl}l_3^I l_j^J l_k^K l_l^L = 0.} \tag{6.46}$$

A line matched across f views produces $2(f-2)$ constraints. The first two views define a line in space, while each further view produces two matching constraints. Hence we expect a line correspondence across four views to produce four constraints on the quadrifocal tensor. In comparison, the quadrifocal incidence relation (6.45) expresses only one constraint. This is because it is a *concurrence relation*, only expressing the fact that the four back-projected planes intersect in a common point, rather than in a common line.

Replacing only one of the four lines in (6.45) with the pencil of lines through the image of the common point produces

$${}^{IJKL}Q^{ijkl}\epsilon_{ii'i''}x_I^{i'}l_j^J l_k^K l_l^L = 0_{i''}, \quad i'' \in \{1,2,3\}, \tag{6.47}$$

a constraint between a point and three lines. Regrouping

$$\epsilon_{ii'i''}\left({}^{IJKL}Q^{ijkl}l_j^J l_k^K l_l^L\right)x_I^{i'} = 0_{i''}, \quad i'' \in \{1,2,3\},$$

reveals the cross product of the point x_I with the projection into I of the intersection of the three back-projected planes from views J, K, and L, as given by (6.42). Consider first the relation corresponding to $i'' = 3$. This implies $i \neq 3$ and $i' \neq 3$. Substituting the corresponding affine quadrifocal transfer equation (6.43), we obtain

$$\epsilon_{ii'}\left({}^{IJK}t^{ijk}l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl}l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl}l_3^J l_k^K l_l^L + {}^{IJKL}q^{ijkl}l_j^J l_k^K l_l^L\right)x_I^{i'} = 0,$$

which by dualizing the trifocal tensors in the index I and regrouping can be rewritten as

$$\boxed{\begin{aligned}\left({}_I^{JK}t_{i'}^{jk}x_I^{i'}l_j^J l_k^K\right)l_3^L - \left({}_I^{JL}t_{i'}^{jl}x_I^{i'}l_j^J l_l^L\right)l_3^K + \left({}_I^{KL}t_{i'}^{kl}x_I^{i'}l_k^K l_l^L\right)l_3^J \\ +\epsilon_{ii'}\left({}^{IJKL}q^{ijkl}l_j^J l_k^K l_l^L\right)x_I^{i'} = 0.\end{aligned}} \tag{6.48}$$

We recognize the first three bracketed expressions from the trifocal incidence relation (6.32). The tensor q is zero for registered views. When $i'' \neq 3$ we have either $i = 3$ or $i' = 3$. Applying (6.43) to each case we obtain

$$\begin{aligned}-\epsilon_{ii''}\left({}^{IJK}t^{ijk}l_j^J l_k^K l_3^L - {}^{IJL}t^{ijl}l_j^J l_3^K l_l^L + {}^{IKL}t^{ikl}l_3^J l_k^K l_l^L + {}^{IJKL}q^{ijkl}l_j^J l_k^K l_l^L\right)x_I^3 \\ +\epsilon_{i'i''}\left({}^{JKL}t^{jkl}l_j^J l_k^K l_l^L\right)x_I^{i'} = 0_{i''}, \quad i'' \in \{1,2\},\end{aligned} \tag{6.49}$$

which by dualizing the trifocal tensors in the index I and regrouping can be rewritten as

$$\boxed{\begin{aligned}\left(\left({}_I^{JK}t_{i''}^{jk}l_j^J l_k^K\right)l_3^L - \left({}_I^{JL}t_{i''}^{jl}l_j^J l_l^L\right)l_3^K + \left({}_I^{KL}t_{i''}^{kl}l_k^K l_l^L\right)l_3^J + \epsilon_{ii''}\left({}^{IJKL}q^{ijkl}l_j^J l_k^K l_l^L\right)\right)x_I^3 \\ -\epsilon_{i'i''}\left({}^{JKL}t^{jkl}l_j^J l_k^K l_l^L\right)x_I^{i'} = 0_{i''}.\end{aligned}} \tag{6.50}$$

6.6 Conclusions

This chapter specialized all four multifocal tensors to the case of affine cameras, both unregistered and registered. The resulting affine tensors were then used as a vehicle to carry the bulk of all known projective multi-view relations to the case of affine views, albeit mostly shown here are the simpler registered relations. Among the benefits of this work are:

1. Gives a coherent view of the range of relations existing between points and lines in multiple affine views (Tables 6.3 and 6.4 on pages 162 and 163).

2. Exposes commonalities between tensors and between relations previously studied in isolation.

3. Brings to light two new affine relations: The primary trifocal incidence relation between a point and two lines (6.34) , as well as trifocal point transfer via a line (6.30).

4. Reveals the existence of an affine quadrifocal tensor and provides affine quadrifocal transfer (6.44) and incidence relations (6.45).

5. Shows how the entries of the registered affine quadrifocal tensor consist exactly of the entries of the four corresponding trifocal tensors (6.17).

This work paves the way for further study of the constraints between the tensor elements, and the estimation of the tensors. This is the path followed in the next chapter.

7

On the Estimation of the Affine Multifocal Tensors

7.1 Overview

This chapter considers the problem of matching image features between affine views by estimating the multifocal tensors. Section 7.3 discusses the over-parameterization of the affine tensors and specifies the number of constraints that exist between the entries of each tensor. Section 7.4 gives necessary and sufficient constraints on the elements of the affine trifocal tensor. Section 7.5 shows that the optimal ML estimate of the tensors from point matches is achieved by factorization and gives the optimal estimators for all four tensors. Section 7.6 discusses the estimation from line correspondences. Section 7.7 gives the solution to the problem of affine triangulation, which in the affine case is linear and can be given simply in terms of the singular value decomposition. The constraints on the trifocal tensor were presented in [TM99].

7.2 Introduction

In the last chapter we specialized the projective unifocal, bifocal, trifocal, and quadrifocal tensors to the affine case, to obtain an affine specialization of known projective multifocal relations between points and lines. We continue the topic of matching between multiple affine views in this chapter by giving for the first time both necessary and sufficient constraints on the components of the affine trifocal tensor, together with a simple geometric interpretation. Finally, we show how the optimal estimate of all four tensors from point correspondences is achieved through factorization, give a formula for optimal affine triangulation, and discuss the estimation from line correspondences.

7.3 Over-parameterization of the Affine Tensors

The components of the projective multifocal tensors are linearly independent but quadratically highly redundant [Tri95a, Tri95b]. The same goes for the affine tensors, but due to the simpler structure of the tensors made evident in the last chapter, we expect the degree of over-parameterization to be less in the affine case. We will now look to obtain the exact number of constraints existing between the components of each of the affine tensors.

The number of degrees of freedom in the *joint projection matrix* of n views [Tri95b], is given in Table 7.1 for the three cases of projective cameras, unregistered and registered affine cameras. In Table 7.2 we give the number of non-zero components in each of the multifocal tensors, less an overall scale factor. Both are well known in the case of projective views [Tri95b], the affine ones we have derived. The difference between the two gives the redundancy, or the number of constraints between the components of each tensor. This is given in Table 7.3.

As seen from Table 7.3, there are no constraints between the components of the affine bifocal tensor. Hence the bifocal tensor is a minimal parameterization of the constraints between two affine views, for both unregistered and registered cameras. This is a well known property of the *affine epipolar constraint* [MZ92a, SZB95].

number of views	projective	unregistered affine	registered
n	$11n - 15$	$8n - 12$	$6n - 9$
2	7	4	3
3	18	12	9
4	29	20	15

Table 7.1: The number of degrees of freedom in the joint projection matrices of multiple views.

tensor	projective	unregistered affine	registered
F	8	4	3
T	26	15	11
Q	80	47	31

Table 7.2: The number of non-zero components in each of the multifocal tensors, less an overall scale factor.

7.4 Constraints on the Affine Trifocal Tensor

From Table 7.3 we see that the trifocal tensor for affine views is a non-minimal parameterization, expecting three constraints to exist between the tensor components for unregistered views, and two for registered views. We will now continue to derive this minimal set of constraints on the affine trifocal tensor, as well as giving a geometric interpretation of the resulting constraint. We first published this novel result in [TM99]. These same constraints were arrived at independently by Åström et al. [ÅHKO99].

Necessary Constraints on the Trifocal Tensor

A minimal set of eight non-linear constraints on the components of the projective trifocal tensor was only very recently given by Canterakis [Can00]. A key factor in his derivation, as he points out, is the representation of the trifocal tensor by three matrices, the so called *homographic slices*, defined as

$$\mathsf{T}^{1\bullet}_{\bullet} \stackrel{\text{def}}{=} \{T_i^{1k}\},\ \mathsf{T}^{2\bullet}_{\bullet} \stackrel{\text{def}}{=} \{T_i^{2k}\},\ \text{and}\ \mathsf{T}^{3\bullet}_{\bullet} \stackrel{\text{def}}{=} \{T_i^{3k}\}, \tag{7.1}$$

tensor	projective	unregistered affine	registered
F	1	0	0
T	8	3	2
Q	51	27	16

Table 7.3: The number of constraints between the components of each of the multifocal tensors, i.e. the bifocal, trifocal, and the quadrifocal tensor, for projective, unregistered affine, and registered views. This is obtained as the difference between the corresponding entries in Table 7.2 and Table 7.1.

each representing a plane homography. The resulting constraints following by enforcing algebraic properties of homographies. This approch was pioneered by Shashua and coworkers [AS98], who went on to determine the minimal set of 51 algebraic constraints on the projective quadrifocal tensor [SW00] using homographic slices.

The following derivation of the minimal set of three respectively two constraints on the unregistered and registered trifocal tensor, predates the above work by Canterakis and Shashua, however, and builds on a partial set of constraints obtained earlier by Papadopoulo and Faugeras [FP98] and Heyden [Hey98] using an alternative matrix representation of the trifocal tensor [Har97, Hey98]

$$\mathsf{T}_1^{\bullet\bullet} \stackrel{\text{def}}{=} \{T_1^{jk}\},\ \mathsf{T}_2^{\bullet\bullet} \stackrel{\text{def}}{=} \{T_2^{jk}\},\ \text{and}\ \mathsf{T}_3^{\bullet\bullet} \stackrel{\text{def}}{=} \{T_3^{jk}\}, \tag{7.2}$$

termed *correlation slices* [Can00], as each of the three matrices represents a plane mapping from a line in view K onto a point in view J, which in the literature [SK52] is known as a *correlation*.

While the correlation slices proved insufficient in arriving at a minimal set of constraints on the projective trifocal tensor, the following derivations shows that a minimal set of constraints could be obtained in the affine case. What Papadopoulo and Faugeras [FP98, PF98] and Heyden [Hey98] achieved using correlation slices, was to produce two necessary conditions for 27 numbers to constitute the components of a *projective* trifocal tensor

$$\det \lambda_i \mathsf{T}_i^{\bullet\bullet} = 0, \quad \forall \lambda_i, \quad \text{and} \quad \det \left[\mathcal{N}(\mathsf{T}_1^{\bullet\bullet}) \quad \mathcal{N}(\mathsf{T}_2^{\bullet\bullet}) \quad \mathcal{N}(\mathsf{T}_3^{\bullet\bullet}) \right] = 0, \tag{7.3}$$

where the operator $\mathcal{N}(\bullet)$ denotes the null-space of its matrix argument. From [Hey98] we have that the matrix argument of the determinant in the second constraint of (7.3) equals the

fundamental matrix between views J and I, up to scaling of the columns. Since the affine fundamental matrix (6.12) is trivially of rank two [SZB95], the second constraint of (7.3) is always satisfied, and the necessary conditions on the *affine* trifocal tensor are reduced to the single condition

$$\boxed{\det \lambda_i \mathsf{T}_i^{\bullet\bullet} = 0, \quad \forall \lambda_i,} \tag{7.4}$$

known in [FP98, PF98] as the *extended rank constraints*.

A Minimal Set of Constraints on the Affine Trifocal Tensor

It follows from the affine specialization of the trifocal tensor given in Section 6.4.3 that for affine cameras the correlation slices (7.2) take the form[1]

$$\mathsf{T}_1^{\bullet\bullet} \stackrel{\text{def}}{=} \left\{T_1^{jk}\right\} = \begin{bmatrix} \mathsf{t}_1^{\bullet\bullet} & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{bmatrix}, \tag{7.5}$$

$$\mathsf{T}_2^{\bullet\bullet} \stackrel{\text{def}}{=} \left\{T_2^{jk}\right\} = \begin{bmatrix} \mathsf{t}_2^{\bullet\bullet} & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{bmatrix}, \tag{7.6}$$

and

$$\mathsf{T}_3^{\bullet\bullet} \stackrel{\text{def}}{=} \left\{T_3^{jk}\right\} = \begin{bmatrix} \mathsf{t}_3^{\bullet\bullet} & {}_{IJ}\mathbf{e}^{\bullet} \\ -{}_{IK}\mathbf{e}^{\bullet T} & 0 \end{bmatrix}. \tag{7.7}$$

Using (7.5)-(7.7) we can write the necessary condition (7.4) as

$$\det \lambda_i \mathsf{T}_i^{\bullet\bullet} = \begin{vmatrix} \lambda_i \cdot \mathsf{t}_i^{\bullet\bullet} & \lambda_3 \cdot {}_{IJ}\mathbf{e}^{\bullet} \\ -\lambda_3 \cdot {}_{IK}\mathbf{e}^{\bullet T} & 0 \end{vmatrix} = 0, \quad \forall \lambda_i.$$

We know that a determinant is linear in any row or column. Since the element in the bottom right hand corner is zero, we can take λ_3 and $-\lambda_3$ to be multipliers of the third column and the third row respectively. Moreover it is easily shown by Laplace expansion that the presence of the zero element causes the determinant to be linear in the 2×2 sub-matrix $\lambda_i \cdot \mathsf{t}_i^{\bullet\bullet}$. It follows directly, that we can factor the λ_i's out of the determinant to obtain the sum

$$-\lambda_3^2 \cdot \lambda_i \cdot \begin{vmatrix} \mathsf{t}_i^{\bullet\bullet} & {}_{IJ}\mathbf{e}^{\bullet} \\ {}_{IK}\mathbf{e}^{\bullet T} & 0 \end{vmatrix} = 0, \quad \forall \lambda_i.$$

This condition holds for all λ_i if and only if all three determinants vanish, giving the three constraints

$$\begin{vmatrix} \mathsf{t}_i^{\bullet\bullet} & {}_{IJ}\mathbf{e}^{\bullet} \\ {}_{IK}\mathbf{e}^{\bullet T} & 0 \end{vmatrix} = 0, \quad i \in \{1, 2, 3\},$$

[1] This result is also found in [FP98], but without the interpretation in terms of the registered tensors.

which when expanded become

$$ {}_{IJ}e^2\,{}_{IK}e^2 t_1^{11} - {}_{IJ}e^1\,{}_{IK}e^2 t_1^{21} - {}_{IK}e^1\,{}_{IJ}e^2 t_1^{12} + {}_{IK}e^1\,{}_{IJ}e^1 t_1^{22} = 0, \tag{7.8} $$

$$ {}_{IJ}e^2\,{}_{IK}e^2 t_2^{11} - {}_{IJ}e^1\,{}_{IK}e^2 t_2^{21} - {}_{IK}e^1\,{}_{IJ}e^2 t_2^{12} + {}_{IK}e^1\,{}_{IJ}e^1 t_2^{22} = 0, \tag{7.9} $$

and

$$ {}_{IJ}e^2\,{}_{IK}e^2 T_3^{11} - {}_{IJ}e^1\,{}_{IK}e^2 T_3^{21} - {}_{IK}e^1\,{}_{IJ}e^2 T_3^{12} + {}_{IK}e^1\,{}_{IJ}e^1 T_3^{22} = 0. \tag{7.10} $$

We see that the third constraint vanishes trivially for registered cameras, leaving two constraints as expected. We can write the two constraints (7.8)-(7.9) together as

$$ \boxed{ {}^{I}_{JK}t^{i}_{jk} \cdot {}_{IJ}e^{j} \cdot {}_{IK}e^{k} = 0^{i}, \quad i \in \{1,2\},} \tag{7.11} $$

by using the dual trifocal tensor ${}^{I}_{JK}t$, or in the dual form

$$ \boxed{ {}^{JK}_{I}t^{jk}_{i} \cdot {}^{IJ}e_{j} \cdot {}^{IK}e_{k} = 0_{i}, \quad i \in \{1,2\}.} \tag{7.12} $$

Alternatively the constraints (7.11) can be derived from the three view quadratic p-relations of type II, as obtained in Section C.2 of Appendix C

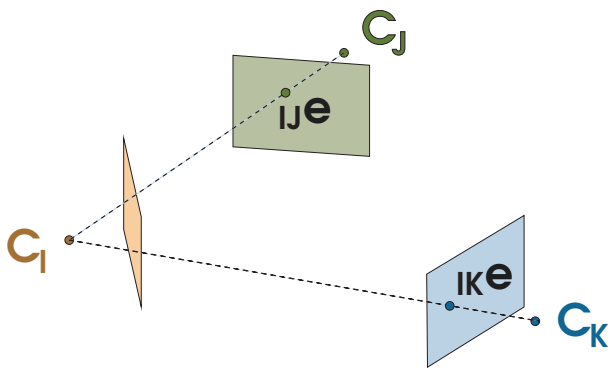
$$ -t^{i21} \cdot_{IJ} e^1 + t^{i11} \cdot_{IJ} e^2 -_{IK} e^1 \cdot_{JI} e^i = 0^i, $$
$$ -t^{i22} \cdot_{IJ} e^1 + t^{i12} \cdot_{IJ} e^2 -_{IK} e^2 \cdot_{JI} e^i = 0^i, $$

by eliminating the unknown tensor ${}_{JI}e$. This provides an alternative proof of necessity. Sufficiency follows by independence and counting argument.

Geometric Interpretation of the Constraints on the Trifocal Tensor

The constraints (7.11) have the form of the transfer equation of affine line directions (6.28). Hence (7.11) can be taken as describing the transfer of corresponding image directions from views J and K into view I. This observation leads us to a simple geometric interpretation of the constraints. From Section 6.3 we know that each of the two unifocal tensors ${}_{IJ}e$ and ${}_{IK}e$ is the direction to the epipole (on the line at infinity) corresponding to camera I in views J and K respectively. It is also well known (see Appendix A), that the centre of projection is the unique null space of any homogeneous projection matrix. Now the constraint (7.11) simply states that the directions

$$ 0 \leftrightarrow {}_{IJ}e \leftrightarrow {}_{IK}e $$



$$ {}_{JK}^{I}t_{jk}^{i}\ {}_{IJ}e^{j}\ {}_{IK}e^{k} = 0^{i}, \qquad i \in \{1, 2\}. $$

Figure 7.1: The constraint on the trifocal tensor states the fact that the epipoles ${}_{IJ}e$ and ${}_{IK}e$ are the images of the centre of camera I.

are in correspondence between views I, J, and K respectively, in accordance with the fact that the three directions are the respective images of the centre of projection of camera I in views I, J, and K.

Relation to the Constraints on the Reduced Affine Quasi-Tensor

Kahl and Heyden [KH98] gave two necessary and sufficient constraints on the components of the *reduced affine quasi-tensor*, an entity we have shown to be equivalent to the trifocal tensor for registered cameras (Section 6.4.3). Kahl and Heyden neither provide a derivation nor an interpretation of their constraints. We will now show that in fact only one of the two constraints they give is correct.

Using the definitions in Section 6.3 we can write the two identities given by [KH98] in terms of the registered tensors ${}^{IJK}t$, ${}_{IJ}e$, and ${}_{IK}e$. Normalizing ${}^{IJK}t$ to $t^{111} = 1$, they are

$$ (t^{122} - t^{112}t^{121})\,{}_{IJ}e^{1}\,{}_{IK}e^{1} = ({}_{IJ}e^{2} - {}_{IJ}e^{1}t^{121})({}_{IK}e^{2} - {}_{IK}e^{1}t^{112}), $$
$$ t^{222}\,{}_{IK}e^{1}\,{}_{IJ}e^{1} = ({}_{IJ}e^{1}t^{221}\,{}_{IK}e^{2} + {}_{IK}e^{1}\,{}_{IJ}e^{2}t^{212} - t^{211}\,{}_{IJ}e^{2}\,{}_{IK}e^{2}). $$

Comparing these identities to the constraints we derived previously, we see that the second identity corresponds exactly to our first constraint (7.8). The first identity has a different form, and turns out to be incorrect in that the two sides have opposite sign, so the first

identity of [KH98] should be replaced by our second constraint (7.9), as confirmed by the authors of that paper [Kah99].

7.5 ML Estimation of Tensors from Point Matches

This section considers the problem of estimating the multi-view tensors from point correspondences across affine views.

7.5.1 Estimation of the Unifocal Tensor

The affine unifocal tensors, or epipoles, ${}^{JI}e$, ${}^{IJ}e$, in a pair of views defined in Section 6.4.1 are exactly the coefficients of the registered bifocal tensor ${}_{IJ}F$ between the views, as seen from (6.11). The estimation of the unifocal tensors follows directly from the estimation of the bifocal tensor.

7.5.2 Estimation of the Bifocal Tensor

The affine bifocal tensor ${}_{IJ}F$ defined in Section 6.4.2 is estimated from point correspondences across two views using the bifocal incidence relation (6.24) (or (6.25)). The affine bifocal incidence relation, or epipolar constraint, is a *linear* constraint on the coordinates of a point in two views. It follows that the bifocal incidence relation is a constraint on the $4 \times N$ measurement matrix of two views, reducing its rank by one, from rank four to rank three. Hence, the affine bifocal incidence relation exactly encodes Tomasi and Kanade's [TK92] *rank three constraint* on a registered measurement matrix, in the two view case.

Estimation from Registered Points

Rewriting the registered bifocal incidence relation (6.24) as the matrix relation

$$\begin{bmatrix} {}^{JI}e_1 & {}^{JI}e_2 & {}^{IJ}e_1 & {}^{IJ}e_2 \end{bmatrix} \begin{bmatrix} x_I^1 \\ x_I^2 \\ x_J^1 \\ x_J^2 \end{bmatrix} = 0$$

we notice that the form is that of the restriction model (3.9) of a linear functional relationship

$$\mathsf{B}_1 \cdot \boldsymbol{\Delta}\mathbf{x}_p = 0, \quad p \in \{1, \ldots, N\}, \tag{7.13}$$

with the particular restriction matrix

$$\mathsf{B}_1 \stackrel{\text{def}}{=} \begin{bmatrix} {}^{JI}e_1 & {}^{JI}e_2 & {}^{IJ}e_1 & {}^{IJ}e_2 \end{bmatrix} \tag{7.14}$$

defined up to scale and consisting of the components of the registered bifocal tensor.

Using the maximum likelihood estimators of the parametric model derived in Section 3.4.2 and the equivalence of the parametric and the restrictions models from Section 3.3.2, we now obtain the maximum likelihood estimator of B_1 under the assumption of normal feature errors with a covariance matrix $\Psi = \sigma^2\Psi_0$ known up to scale. From the equivalence relation (3.11) between the two models we have the constraint

$$\widehat{\mathsf{B}}_1\widehat{\mathsf{M}}_3 \stackrel{!}{=} \mathbf{0}_3^T,$$

relating the estimators of the two models. If $\widehat{\mathsf{M}}_3$ is the ML estimator of the joint projection matrix of two views, as defined in the isotropic factorization (3.31), we obtain the ML estimator of the bifocal tensor

$$\boxed{\widehat{\mathsf{B}}_1 = \left(\Psi_0^{-\frac{1}{2}}\mathsf{U}_1'^*\right)^T,} \tag{7.15}$$

where $\mathsf{U}_1'^*$ is the least[2] left singular vector of the Mahalanobis transformed measurement matrix $\Delta\mathsf{W}'$, defined in (3.23). This can be seen by partitioning the unitary matrix U' of left singular vectors as follows and substituting (3.31) and (7.15) to obtain

$$\mathsf{U}' = \begin{bmatrix}\mathsf{U}_3' | \mathsf{U}_1'^*\end{bmatrix} = \left[\Psi_0^{-\frac{1}{2}}\widehat{\mathsf{M}}_3 | \Psi_0^{\frac{1}{2}}\widehat{\mathsf{B}}_1^T\right],$$

which, by the unitary property of U', gives

$$\mathbf{0}_3 = \mathsf{U}_1'^{*T}\mathsf{U}_3' = \widehat{\mathsf{B}}_1\Psi_0^{\frac{1}{2}}\Psi_0^{-\frac{1}{2}}\widehat{\mathsf{M}}_3 = \widehat{\mathsf{B}}_1\widehat{\mathsf{M}}_3,$$

as required. The ML estimate (7.15) can be computed from $N \geq 4$ point correspondences in the two views, assuming that the depicted scene points are non-coplanar. We note that the information contained in the registered bifocal tensor ${}_{IJ}F$ and the registered projection matrices M_I, M_J is equivalent, and that the optimal estimate of the tensor is obtained by the same factorization procedure that produces the well known optimal estimator of the projection matrices (see Section 3.4). This is no coincidence, but is a direct consequence of the functional relationship (6.10) between the tensor and the projection matrices, by the *invariance principle* of maximum likelihood estimators (see e.g. [BNB96]).

[2]By the *least* left singular vector we mean the left singular vector corresponding to the smallest singular value.

Estimation from Unregistered Points

The unregistered bifocal incidence relation (6.25) can also be written in the form of the restriction model (3.9) as

$$\mathsf{B}_1 \cdot \mathbf{x}_p = b, \quad p \in \{1, \ldots, N\},$$

with $b \stackrel{\text{def}}{=} -_{IJ}F_{33}$. From the equivalence relations (3.11) and (3.12) we obtain exactly the same ML estimator (7.15) for B_1 as before, and from (3.12) and (3.16) we obtain the ML estimator for the additional parameter b as

$$\widehat{b} = \widehat{\mathsf{B}}_1 \widehat{\mathbf{t}} = \widehat{\mathsf{B}}_1 \overline{\mathbf{w}}. \tag{7.16}$$

From this we conclude that working with unregistered measurements does not bring any advantages in terms of the accuracy of the estimation of the bifocal tensor, irrespective of the number of correspondences. On the contrary, it has been shown by Hartley [Har95a, Har97, Har98] that pre-normalization of the image data, by registration to the centroid $\overline{\mathbf{w}}$ followed by scaling, is essential for the numerical stability of many related estimation algorithms.

When the feature errors are isotropic, i.e. $\Sigma_0 = \mathsf{I}_4$, the ML estimator (7.15) is simply

$$\widehat{\mathsf{B}}_1 = (\mathsf{U}_1^*)^T, \tag{7.17}$$

where U_1^* is the least left singular vector of the registered measurement matrix $\Delta\mathsf{W}$. It is easily shown that (7.17) and (7.16) are the solutions of the affine epipolar equation obtained in [SZB95], when minimizing the sum of the squared reprojection error.

Detecting Degenerate Data Configurations

Torr et al. [TZM98] provided methods to detect degenerate data configurations while estimating the fundamental matrix. They provide a catalog of non-degenerate and degenerate models of the projective fundamental matrix, together with a scoring function to select the model best described by the data. Torr et al. term the data degenerate with respect to a model *if the underlying set of noise free or true correspondences does not admit to a unique solution with respect to that model*. The hierarchy of models (of degree 1) includes both the affine fundamental matrix, as well as the affine image transformation (or *affinity*) that

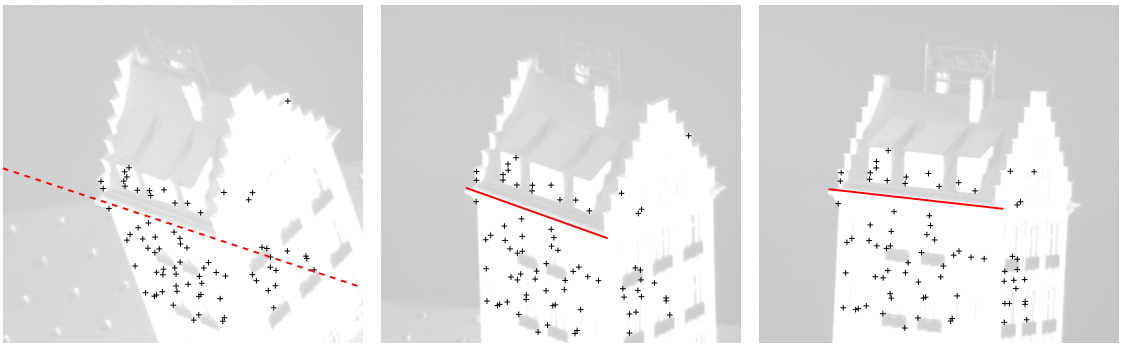


Figure 7.2: Line transfer into the first view from the second and third, using the affine trifocal tensor, as described in the transfer equations (6.27). The trifocal tensor was estimated by the ML estimator from the 74 corresponding image points, shown by '+'.

may be an appropriate model when the data do not uniquely determine the affine fundamental matrix. The affine fundamental matrix has four degrees of freedom and is uniquely determined by four corresponding points in general position. The image-to-image affinity has six degrees of freedom and is uniquely determined by three points in a general (and necessarily planar) position. In addition they provide a model of *image translation* with one degree of freedom, requiring one correspondence and a *no motion* model with zero degrees of freedom requiring no correspondences.

We notice in the above hierarchy of degenerate models, that the image-to-image affinity corresponds to our linear restriction model (3.9) of codimension two with an unknown restriction matrix B_2. The image translation and no motion models correspond to the same model with B_2 fixed and $\mathbf{b}$ unknown or fixed, respectively. The counterparts to our codimension three and four models are not considered as they do not correspond to a general configuration of image points.

The scoring function proposed by Torr et al. is devised for use with a robust estimator such as RANSAC. The function is based on minimum description length and is takes into account the dimension and degrees of freedom in the model, as well as the number of inliers and outliers to the estimated solution.

7.5.3 Estimation of the Trifocal Tensor

The estimation of the affine trifocal tensor from point correspondences follows analogously to the estimation of the bifocal tensor. As in the binocular case, we can write the geometric

constraints between three views of the same scene point using a restriction model

$$\mathsf{B}_3 \cdot \boldsymbol{\Delta}\mathbf{x}_p = \mathbf{0}_3, \quad p \in \{1, \ldots, N\}, \tag{7.18}$$

where B_3 is the 3×6 restriction matrix of the linear functional relationship. This equation underlines the fact, that since the measurement space of three views has dimension $M = 6$, the subspace spanned by the true measurements of a generic 3D point set must have codimension $q = 6 - 3 = 3$. Consequently, there exist at most three independent linear constraints on the measurements. This implies that the four affine trifocal constraints (6.41) cannot be linearly independent, as hitherto assumed (see e.g. [Tor95, QOM98]).

While the linear functional relationship (7.18) provides a minimal set of constraints on the points $\boldsymbol{\Delta}\mathbf{x}_p$ in the joint image, we have from Section 3.3.2 (p. 40) that the 18 parameters of the restriction matrix B_3 are defined only up to a 3×3 ambiguity representing an arbitrary choice of basis in the dual space. The remaining $18-9 = 9$ degrees of freedom equal those of the joint projection matrix of three views as given in Table 7.1 (p. 170). It follows that a given ML estmate $\widehat{\mathsf{B}_3}$ of the restriction matrix corresponds to one particular choice of basis in the dual space, and that other equivalent ML estimates are obtained by left multiplication of $\widehat{\mathsf{B}_3}$ by an invertible 3×3 matrix.

The four trifocal constraints for registered cameras (6.41) can be written together in the single matrix equation

$$\begin{bmatrix} t_1^{22} & t_2^{22} & 0 & {}^{IK}e_2 & 0 & -{}^{IJ}e_2 \\ t_1^{21} & t_2^{21} & 0 & {}^{IK}e_1 & -{}^{IJ}e_2 & 0 \\ t_1^{12} & t_2^{12} & {}^{IK}e_2 & 0 & 0 & -{}^{IJ}e_1 \\ t_1^{11} & t_2^{11} & {}^{IK}e_1 & 0 & -{}^{IJ}e_1 & 0 \end{bmatrix} \cdot \boldsymbol{\Delta}\mathbf{x}_p = \mathbf{0}_4, \quad p \in \{1, \ldots, N\}, \tag{7.19}$$

exposing their commonality with the restriction model (7.18).

The three bifocal incidence relations or epipolar constraints (7.20) existing between the three views can similarly be represented by a single matrix equation

$$\begin{bmatrix} {}^{JI}e_1 & {}^{JI}e_2 & {}^{IJ}e_1 & {}^{IJ}e_2 & 0 & 0 \\ {}^{KI}e_1 & {}^{KI}e_2 & 0 & 0 & {}^{IK}e_1 & {}^{IK}e_2 \\ 0 & 0 & {}^{KJ}e_1 & {}^{KJ}e_2 & {}^{JK}e_1 & {}^{JK}e_2 \end{bmatrix} \cdot \boldsymbol{\Delta}\mathbf{x}_p = \mathbf{0}_3, \quad p \in \{1, \ldots, N\}, \tag{7.20}$$

which for non-collinear camera centres is of rank 3 and hence provide the same constraints. For three collinear camera centres, however, the rank of the bifocal incidence relations (7.20) drops to two and the three bifocal incidence relations fail to provide the same constraints as the trifocal incidence relations.

Constraints on the Trifocal Tensor and Point Incidence Relations

We have just shown that in the case of point matches, three bifocal or trifocal incidence relations suffice to characterize the geometry of three affine views. We will now show that in the case of point matches the constraints between the components of the trifocal tensor, derived in Section 7.4, are indeed implicit in the trifocal incidence relations.

We contract the trifocal incidence relation for points (6.41) with the tensors ${}^{IJ}e_j$ and ${}^{IK}e_k$, to obtain

$$\left({}_I^{JK}t_i^{jk}x_I^i + {}_{IK}e^k x_J^j - {}_{IJ}e^j x_K^k\right) \cdot {}^{IJ}e_j \cdot {}^{IK}e_k = 0$$

or

$${}_I^{JK}t_i^{jk} \cdot {}^{IJ}e_j \cdot {}^{IK}e_k \cdot x_I^i + {}^{IK}e_k \cdot {}_{IK}e^k \cdot {}^{IJ}e_j \cdot x_J^j - {}^{IJ}e_j \cdot {}_{IJ}e^j \cdot {}^{IK}e_k \cdot x_K^k = 0.$$

Since the contraction of a tensor with its dual is zero ${}^{IK}e_k \cdot {}_{IK}e^k = {}^{IJ}e_j \cdot {}_{IJ}e^j = 0$, we are left with only the first term of the above expression

$${}_I^{JK}t_i^{jk} \cdot {}^{IJ}e_j \cdot {}^{IK}e_k \cdot x_I^i = 0.$$

This equation must hold for any x_I^i, giving the two constraints

$${}_I^{JK}t_i^{jk} \cdot {}^{IJ}e_j \cdot {}^{IK}e_k \cdot = 0_i, \qquad i \in \{1, 2\},$$

which are the constraints (7.12) between the components of the registered trifocal tensor. Hence we have shown that the constraints between the components of the trifocal tensor are implicit in the trifocal incidence relations for points.

7.5.4 Estimation of the Quadrifocal Tensor

The estimation of the affine quadrifocal tensor from point correspondences follows analogously to the estimation of the bifocal tensor. As in the bifocal case, we can write the geometric constraints between four views of the same scene point using a restriction model

$$\mathsf{B}_5 \cdot \mathbf{\Delta x}_p = \mathbf{0}_5, \quad p \in \{1, \ldots, N\},$$

where B_5 is the 5×8 restriction matrix of the linear functional relationship. This equation underlines the fact, that since the measurement space of four views has dimension $M =$

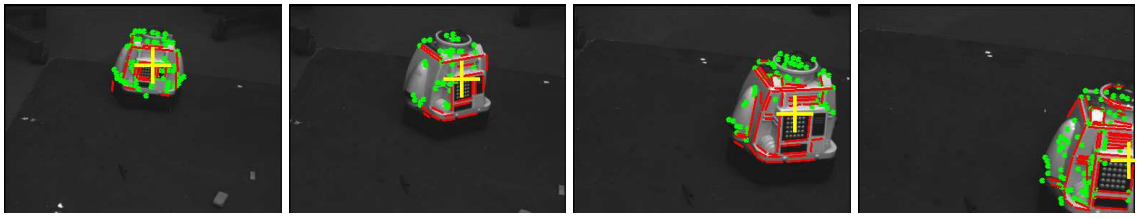


Figure 7.3: Successful tracking of an object by matching points and lines in successive frames using the constrained trifocal tensor. The yellow cross shows the "fixation point", a virtual point transferred across the entire sequence. Shown is every 12th frame. Reproduced from [HTM99, Hay00] courtesy of Eric Hayman.

8, the subspace spanned by the true measurements of a generic 3D point set must have codimension $q = 8 - 3 = 5$. As a consequence there exist at most five independent linear constraints on the measurements.

7.6 Estimation of a Constrained Trifocal Tensor using Lines

Estimation of the affine trifocal tensor from line correspondences subject to the constraints (7.12) is a harder problem than the estimation from points. Two common but very different methods observing the constraints in the projective case are Hartley's method [Har97] and the method of Torr and Zisserman [TZ97]. A further linear method given by Quan et al. [QOM98] uses only the directions of lines, thus giving only one relation per line. The affine case is simpler in that considerably fewer parameters need to be estimated. Each triplet of corresponding lines produces two relations (6.33) by selecting two points on the line in the first view. Since the constraints between the tensor elements (7.12) are trilinear in the tensor components, this suggests iterative estimation methods. To compute all 16 components of the trifocal tensor up to global scale, while enforcing constraints between tensor elements, we can use any combination of n_p point and n_l line correspondences, provided the the following inequality is satisfied

$$3n_p + 2n_l + 3 \geq 15,$$

hence we need a minimum of 4 points or 6 lines. The corresponding condition for the estimation of the registered tensor is

$$3(n_p - 1) + 2n_l + 2 \geq 11.$$

The emergence of the explicit constraints on the trifocal tensor (7.12) does therefore not suggest major computational advantages over an affine adaptation [BL98, MC98] of Hartley's non-linear method [Har97]. The new constraints do however offer an alternative parameterization to Hartley's in that it is now possible to solve for the entries of the trifocal tensor without resorting to a parameterization of the tensor in terms of projection matrices. Further investigation of this and related methods has been carried out by Hayman [HTM99, Hay00]. An example of the results obtained is shown in Figure 7.3.

7.7 Affine Triangulation

Given two images of a scene point, we wish to determine the point of intersection of their back-projection in the scene. This problem is known as *triangulation* [HS97]. Due to imperfect measurements, the back-projected rays of corresponding image points do not necessarily intersect. Given two observed images of a point, their back-projections intersect if and only if the image points satisfy the epipolar constraint. Given the probability distribution of the image errors, and error free knowledge of the epipolar geometry between the views, we seek two image locations, satisfying the epipolar constraint, that are most likely of causing the observed image points. In the case of projective cameras this problem is nonlinear, and its solution requires the solving a sixth order polynomial equation for each correspondence [HS97]. In the affine case, this problem has a linear solution.

Given the measured point $\mathbf{w}$ in the joint image with error covariance Ψ we seek the point $\widehat{\mathbf{w}}$ satisfying the epipolar constraint, which is closest to $\mathbf{w}$ in the Mahalanobis sense. It is the solution of the constrained minimization problem

$$(\mathbf{w}-\widehat{\mathbf{w}})^T\Psi^{-1}(\mathbf{w}-\widehat{\mathbf{w}}) \stackrel{!}{=} \min, \quad \text{subject to} \quad \mathsf{B}_1\widehat{\mathbf{w}} = 0,$$

where the row matrix B_1 is the bifocal restriction matrix (7.14). The solution is given by the orthogonal projection in Mahalanobis space of $\mathbf{w}$ onto the complement of the subspace spanned by the column of the adjoint matrix $\mathsf{B}_1^* = \mathsf{B}_1^T$, as

$$\begin{aligned}\widehat{\mathbf{w}} &= \mathbf{w} - \Psi^{\frac{1}{2}}\left(\Psi^{\frac{1}{2}}\mathsf{B}_1^T(\mathsf{B}_1\Psi\mathsf{B}_1^T)^{-1}\mathsf{B}_1\Psi^{\frac{1}{2}}\right)\Psi^{-\frac{1}{2}}\mathbf{w} \\ &= \left(\mathsf{I}_4 - \Psi\mathsf{B}_1^T(\mathsf{B}_1\Psi\mathsf{B}_1^T)^{-1}\mathsf{B}_1\right)\mathbf{w} \\ &\stackrel{\text{def}}{=} \mathsf{C}\mathbf{w}. \end{aligned} \tag{7.21}$$

The triangulation matrix C is the resulting projection operator. We notice that C depends only on the known epipolar geometry and on the covariance of the point $\mathbf{w}$. When all points in the joint image have equal covariance, $\Psi = \Psi_0$, the matrix C is fixed, and needs only be computed once for each pair of views. In this case the optimal estimate of the epipolar geometry represented by B_1 is obtained by factorization, as given by (7.15), and C takes the particular form

$$\begin{aligned}\mathsf{C} &= \mathsf{I}_4 - \Psi_0 \mathsf{B}_1^T (\mathsf{B}_1 \Psi_0 \mathsf{B}_1^T)^{-1} \mathsf{B}_1 \\ &= \mathsf{I}_4 - \Psi_0^{\frac{1}{2}} \mathsf{U}_1'^{*} (\mathsf{U}_1'^{*T} \mathsf{U}_1'^{*})^{-1} \mathsf{U}_1'^{*T} \Psi_0^{-\frac{1}{2}} \\ &= \mathsf{I}_4 - \Psi_0^{\frac{1}{2}} \mathsf{U}_1'^{*} \mathsf{U}_1'^{*T} \Psi_0^{-\frac{1}{2}}\end{aligned} \tag{7.22}$$

This simple linear solution should be compared to the solution in the case of projective cameras given in [HS97], which involves solving a separate sixth order polynomial equation for each pair of corresponding points.

This result is easily extended to $F > 2$ views. For an $M = 2F$ dimensional joint image we have $\widehat{\mathbf{w}} = \mathsf{C}_M \mathbf{w}$ with the matrix

$$\boxed{\mathsf{C}_M = \mathsf{I}_M - \Psi \mathsf{B}_q^T (\mathsf{B}_q \Psi \mathsf{B}_q^T)^{-1} \mathsf{B}_q,} \tag{7.23}$$

and in the case of equal covariance of points in the joint image, and the F-focal geometry obtained by factorization as $\mathsf{B}_q = \mathsf{U}_q'^{*T} \Psi_0^{-\frac{1}{2}}$, we have

$$\boxed{\mathsf{C}_M = \mathsf{I}_M - \Psi_0^{\frac{1}{2}} \mathsf{U}_q'^{*} \mathsf{U}_q'^{*T} \Psi_0^{-\frac{1}{2}}.} \tag{7.24}$$

7.8 Conclusions

In this chapter we considered aspects of the problem of matching image features between affine views by estimating the multifocal tensors. The contributions of this chapter are as follows.

1. We have specified the number of constraints that exist between the components of each of the four affine tensors (Table 7.3 on page 171).

2. In the case of the affine trifocal tensor, we derived the minimal set of constraints existing between the entries of the tensor (7.11), an important novel result. These constraints can be interpreted as encoding the fact that the epipoles ${}_{IJ}e$ and ${}_{IK}e$ are the images of the centre of camera I.

3. We derived the maximum likelihood estimators of the affine tensors from point correspondences, in terms of the restriction form of the linear functional relationship, and showed this to be obtained by factorization. This hammers home the fact that the machinery of the multifocal tensors is only required when dealing with line correspondences.

4. The constraints on the trifocal tensor derived in this chapter open the possibility of estimating the tensor directly from line correspondences without resorting to a parameterization in terms of projection matrices. This is not likely to lead to a significant computational gain, however.

5. Finally, we gave the linear solution to the affine triangulation problem, simply stated in terms of factorization (7.24).

8

Conclusions

This thesis has been concerned with the utilization of object symmety as a cue for segmentation and object recognition.

Summary

The main objective of this thesis was to investigate means to detect 3D bilateral symmetry from affine views. Towards this goal we have made the following contributions:

- We have extended the notion of *skewed symmetry* to three dimensions, and given a definition in terms of degenerate structure that applies equally to 3D affine structure and to multiple affine views.
- We have developed an optimal method to detect degenerate structure, and through that also 3D skewed symmetry, in the presence of non-isotropic uncertainties in point locations.
- We have investigated simple techniques of searching for symmetric objects within a scene, and we have implemented and demonstrated fully automatic detection of 3D skewed symmetry from unconnected point features in affine views of uncluttered scenes.

- Further, we have presented constraints along with optimal tests which allow us to test for bilateral rather than skewed symmetry, by checking consistency between three skewed symmetric sets. We have also given constraints on smaller sets, by making use of knowledge about the internal parameters of the affine camera.

- In an effort to increase our understanding of the matching relations that exist between affine views, we specialized the four projective multifocal tensors to the affine case, and used these to carry the bulk of all known projective multi-view relations into affine views, unearthing some new relations in the process.

- Having done that, we addressed the problem of estimating the tensors. In particular, we provided for the first time a minimal set of constraints on the affine trifocal tensor, and investigated ways of estimating the tensors from point and line correspondences.

Future Research

As for future research directions prompted by this thesis:

- As for direct improvements to the detection of 3D bilateral symmetry from point features, means to reduce the complexity of the search would incur the greatest practical advantages. In this regard, the use of either connected or more descriptive point features could be expected to be beneficial. Of particular interest are features that incorporate a view invariant description of the local image structure.

- The approach could be extended to projective rather than affine views, in order to make it applicable to larger scene structure. This is expected to be a computationally harder problem due to weaker constraints. On a different note, recent advances in self-calibration may make it less objectionable to incorporate Euclidean information at the symmetry detection stage.

- The approch could also be extended to include other types of symmetry, although few are as generic as bilateral symmetry.

- Extending the approach to other types of input features, such as straight lines and curved contours, which are arguably more descriptive of the scene content, and bring constraints of their own to the symmetry detection problem.

- The problem of matching features between views is still open to improvement, especially in the case of lines and curves. The role of the affine quadrifocal tensor introduced in this thesis, and the constraints on its elements also demand further investigation.

- Finally, investigating how the information gained from the knowledge about symmetry in the scene is best used to further image understanding.

Part III

Appendices

A

The Affine Camera and Projective Geometry

The affine camera model was suggested by Mundy and Zisserman [MZ92a] as a special case of the then newly rediscovered projective camera, thereby generalizing the linear camera approximations that were in common use at the time. During the last decade the geometry of multiple projective views has been the subject on intense study, resulting in a greater understanding of the relations between multiple views, as well as giving rise to many practical new algorithms. Despite many advances in the understanding of the affine camera, a comparable effort to that of understanding the projective camera has not been undertaken. In light of the deeper understanding of the projective camera we feel that it is necessary to take a fresh look at the affine camera and draw its relation to the projective camera more clearly. We begin this section by giving an overview over the properties of the projective camera, before turning to the affine camera model.

A.1 The Projective Camera

The *projective camera* P is defined as the linear projective projection represented by the matrix

$$\mathsf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \tag{A.1}$$

mapping scene points to image points in a *projective view* by the equation

$$\mathbf{x} = \mathsf{P}\mathbf{X}.$$

The projective camera is usually taken to be uncalibrated, i.e. that the parameters are unknown, that nothing is known about the camera apart from the projection matrix having the above form.

Representation and Bases

The particular representation of P is determined by the choice of basis in the scene and in the image. A change of basis in the image transforms the coordinates of image points by

$$\mathbf{x} \mapsto \mathsf{K}\mathbf{x}$$

and those of image lines by

$$\mathbf{l}^T \mapsto \mathbf{l}^T\mathsf{K}^{-1},$$

conserving the scalar product between points and lines

$$\mathbf{l}^T\mathbf{x} \mapsto \mathbf{l}^T\mathsf{K}^{-1}\mathsf{K}\mathbf{x} = \mathbf{l}^T\mathbf{x}.$$

A change of basis in the scene transforms point coordinates by

$$\mathbf{X} \mapsto \mathsf{T}\mathbf{X}$$

and those of planes by

$$\pi^T \mapsto \pi^T\mathsf{T}^{-1},$$

conserving the scalar product between scene points and planes

$$\pi^T\mathbf{X} \mapsto \pi^T\mathsf{T}^{-1}\mathsf{T}\mathbf{X} = \pi^T\mathbf{X}.$$

Under the two basis changes, the representation of the projective camera changes

$$\mathsf{P} \mapsto \mathsf{K}\mathsf{P}\mathsf{T}^{-1}.$$

Significance of the Null Space

A characteristic property of a projective view is the unique null space $\mathbf{C}$ of the camera, defined by

$$\mathsf{P}\mathbf{C} = \mathbf{0}_3,$$

since it is invariant under any change of coordinates K in the image, as seen by

$$\mathsf{K}\mathsf{P}\mathbf{C} = \mathsf{K}\mathbf{0}_3 = \mathbf{0}_3.$$

It follows that two views of a non-degenerate scene from cameras with a different null space $\mathbf{C} \neq \mathbf{C}'$ cannot be transformed into each other, and conversely that two views from cameras with the same null space can always be transformed into each other, and are thus projectively equivalent.

A.1.1 Geometric Interpretation of the Projective Camera

The projective camera defines three planes of reference, whose coefficients appear in the rows of the projection matrix [Fau93]

$$\begin{aligned} \mathbf{p}_1^T &= \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \end{bmatrix} \\ \mathbf{p}_2^T &= \begin{bmatrix} p_{21} & p_{22} & p_{23} & p_{24} \end{bmatrix} \\ \mathbf{p}_3^T &= \begin{bmatrix} p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}, \end{aligned}$$

and whose common intersection is the centre of projection (the null space) $\mathbf{C}$. The convention to map homogeneous to inhomogeneous image coordinates by

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_1/x_3 & x_2/x_3 \end{bmatrix}$$

leads to the interpretation of $\mathbf{p}_3$ as the plane at zero distance or the *focal plane* [Fau93] (the plane through the centre of projection parallel to the image plane), while the planes $\mathbf{p}_1$ and $\mathbf{p}_2$ are defined by the centre of projection and the image axes $x = 0$ and $y = 0$ respectively.

A.2 Fixed Sets and Calibration

This section discusses the sets of points in projective space and planes in the dual space, which remain fixed under Euclidean motion of the camera. By definition these sets com-

pletely determine the Euclidean space, and hence the Euclidean structure of that space. Knowledge of the fixed sets amounts to calibration of both the camera and the structure.

A.2.1 The Absolute Disk and its Image

The set of planes in $\mathcal{P}^{*3}$ fixed set-wise under a general Euclidean motion is the *absolute disk*, introduced by [Tri97] under the name *absolute quadric*. It is the dual (or plane) quadric

$$\pi^T \mathsf{Q}^* \pi = 0,$$

represented in every Euclidean frame by the matrix

$$\mathsf{Q}^* = \mathsf{Q}_0^* \stackrel{\text{def}}{=} \begin{bmatrix} \mathsf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3^T & 0 \end{bmatrix}. \tag{A.2}$$

Since Q^* has rank 3, the absolute quadric is degenerate, more specifically it is the *disk quadric* [SK52], consisting of the plane pencils defined by each of the (tangent) lines of the *absolute dual conic* Ω^* in π_∞, whose representation in every Euclidean frame is

$$\Omega^* = \Omega_0^* \stackrel{\text{def}}{=} \mathsf{I}_3.$$

In a projective frame Q^* becomes an arbitrary symmetric positive semidefinite rank 3 matrix [Tri97]

$$\mathsf{Q}^* = \mathsf{T}\mathsf{Q}_0^*\mathsf{T}^T.$$

The point $\pi^T\mathsf{Q}^*$ is the *pole* of π with respect to Q^* [SK52]. It is the 3D point on π_∞ representing the normal direction of π.

The absolute disk provides, via the scalar product of plane normals

$$\mathbf{n}_1^T \mathbf{n}_2 = \pi_1^T \mathsf{Q}^* \pi_2,$$

a basis independent measure of the angle θ between the planes π_1 and π_2 [Tri97]

$$\cos\theta = \frac{\pi_1^T \mathsf{Q}^* \pi_2}{\sqrt{(\pi_1^T \mathsf{Q}^* \pi_1)(\pi_2^T \mathsf{Q}^* \pi_2)}}.$$

The absolute disk projects onto the dual image of the absolute conic ω^* by [Tri97]

$$\omega^* = \mathsf{P}\mathsf{Q}^*\mathsf{P}^T.$$

This equation describes the relationship between scene and camera calibration, providing a basis for self-calibration [Tri97, PKV98].

In particular this makes it possible to measure the angle between two plane normals from the two vanishing lines in one image via [Tri97]

$$\mathbf{l}_1^T \omega^* \mathbf{l}_2 = \mathbf{l}_1^T \mathsf{P}\mathsf{Q}^*\mathsf{P}^T \mathbf{l}_2 = \pi_1^T \mathsf{Q}^* \pi_2.$$

Given two views, the angle between two epipolar planes must measure the same from either view [Tri97]. This is the *Kruppa constraint* [FLM92] used for self-calibration.

A.2.2 The Camera Cone and its Image

The dual of the absolute disk is not invariant to translation, so that the point equivalent to the absolute disk cannot be defined [Tri97]. However, the point quadric is absolute under pure rotation.

The Dual of the Absolute Disk

The dual of the disk quadric is the *quadric cone*, a point quadric of rank 3 [SK52]. The point quadric

$$\mathbf{X}^T \mathsf{Q} \mathbf{X} = 0,$$

represented in a particular Euclidean frame by the matrix

$$\mathsf{Q} = \mathsf{Q}_0 \stackrel{\text{def}}{=} \begin{bmatrix} \mathsf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3^T & 0 \end{bmatrix},$$

is the quadric cone with vertex at the origin $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T$, intersecting π_∞ in the absolute conic Ω, whose representation in every Euclidean frame is

$$\Omega = \Omega_0 \stackrel{\text{def}}{=} \mathsf{I}_3.$$

The cone Q_0 is invariant to a pure rotation of the Euclidean frame but, under a general Euclidean coordinate transformation

$$\mathsf{T}_e = \begin{bmatrix} \mathsf{R} & -\mathsf{R}\mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix},$$

Q_0 is mapped to

$$\mathsf{Q} = \mathsf{T}_e^{-T} \mathsf{Q}_0 \mathsf{T}_e^{-1} = \begin{bmatrix} \mathsf{I}_3 & \mathbf{t} \\ \mathbf{t}^T & \mathbf{t}^T\mathbf{t} \end{bmatrix}$$

reflecting the translation of the origin of the frame away from the vertex of the cone.

In a projective frame Q becomes an arbitrary symmetric positive semidefinite rank 3 matrix

$$\mathsf{Q} = \mathsf{T}^{-T}\mathsf{Q}_0\mathsf{T}^{-1}.$$

The plane $\mathbf{X}^T\mathsf{Q}$ is the *polar plane* of $\mathbf{X}$ with respect to Q [SK52]. It is the plane through the vertex of Q with a normal in the direction of to $\mathbf{X}$.

The absolute cone allows us to measure the angle between two scene rays, passing through its vertex and the points $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively via

$$\mathbf{X}_1^T\mathsf{Q}\mathbf{X}_2.$$

The Camera Cone

A camera P with a centre of projection (null space) C, which maps the absolute conic to ω, defines the unique quadric cone

$$\mathsf{Q}_C = \mathsf{P}^T\omega\mathsf{P},$$

with a vertex at the centre of projection (null space) C. The camera cone has 8 d.o.f. (10, less scale, less rank deficiency is 8), three of which are determined by the null space of P, obviously preserved by

$$\mathsf{Q}_C\mathbf{C} = \mathsf{P}^T\omega\mathsf{P}\mathbf{C} = \mathsf{P}^T\omega\mathbf{0}_3 = \mathbf{0}_4,$$

the remaining 5 d.o.f. being those of the absolute conic, Ω.

In particular the *camera cone* Q_C makes it possible to measure the angle between two projection rays directly from the image of the two points via

$$\mathbf{X}_1^T\mathsf{Q}_C\mathbf{X}_2 = \mathbf{X}_1^T\mathsf{P}^T\omega\mathsf{P}\mathbf{X}_2 = \mathbf{x}_1^T\omega\mathbf{x}_2.$$

For points at infinity (directions) $\mathbf{X}_1 = \mathbf{D}_1$ and $\mathbf{X}_2 = \mathbf{D}_2$, this becomes

$$\mathbf{D}_1^T\mathsf{Q}_C\mathbf{D}_2 = \mathbf{v}_1^T\omega\mathbf{v}_2,$$

meaning that we can measure the angle between any two directions in space from only the two vanishing points in one image, as shown in [ZLA98].

A.3 The Affine Camera

The affine camera P_a was defined in [MZ92a] as the special case $p_{31} = p_{32} = p_{33} = 0$ of the projective projection matrix (A.1), or

$$\mathsf{P}_a = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{bmatrix}.$$

This restricted form of the projection matrix forms a subclass of projection matrices, which includes well known approximations to the perspective camera: orthography, weak perspective (or scaled orthography), and para-perspective. The set of parameters of the affine and the para-perspective cameras are in a one-to-one correspondence [Bas96], but the affine camera is generally taken to be uncalibrated, i.e. the nine parameters (independent up to a common scale factor) are not assumed to be known.

The affine camera is an affine projection, in that points at infinity are mapped to infinity. The projection P_a maps the plane at infinity π_∞ to the line at infinity l_∞ in the image. The consequence is that parallel lines in space (intersecting in π_∞) are mapped to parallel lines in the image (intersecting in l_∞), and the midpoint between two points in space (harmonic with the intersection with π_∞) is mapped to the midpoint between the images of the two points (harmonic with the intersection with l_∞).

A.3.1 Geometric Interpretation of the Affine Camera

The focal plane of the affine camera is given by the plane equation

$$\mathbf{p}_3^T \mathbf{X} = p_{34} x_4 = 0,$$

whose unique solution is the plane at infinity π_∞ ($x_4 = 0$). The centre of projection is given by the common intersection of the three reference planes, which is the null space of the projection P_a

$$\mathsf{P}_a \mathbf{C} = \mathbf{0}_3.$$

The bottom equation gives $c_4 = 0$, and substituting this into the remaining equations we obtain

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \mathbf{0}_2$$

giving the direction of projection

$$\mathbf{d}_c \stackrel{\text{def}}{=} \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}^T = \begin{bmatrix} p_{11} & p_{12} & p_{13} \end{bmatrix} \times \begin{bmatrix} p_{21} & p_{22} & p_{23} \end{bmatrix} = \mathbf{m}_1^T \times \mathbf{m}_2^T,$$

where $\mathbf{m}_1^T$ and $\mathbf{m}_2^T$ are the top two rows of P_∞ defining the two lines on π_∞, where π_∞ is intersected by the reference planes $\mathbf{p}_1^T$ and $\mathbf{p}_2^T$. Hence the centre of projection of the affine camera is the point

$$\mathbf{C} = \begin{bmatrix} \mathbf{m}_1^T \times \mathbf{m}_2^T \\ 0 \end{bmatrix}, \tag{A.3}$$

lying on the plane at infinity. Note that the centre of projection only depends on the basis of space, and is invariant to a change of basis in the image.

A.3.2 The Affine Image of the Absolute Disk

Under an affine point transformation of space T_a the Euclidean representation of the absolute disk (A.2) is mapped to

$$\mathsf{Q}^* = \mathsf{T}_a \mathsf{Q}_0^* \mathsf{T}_a^T = \begin{bmatrix} \Omega^* & \mathbf{0}_3 \\ \mathbf{0}_3^T & 0 \end{bmatrix},$$

where the symmetric positive definite matrix Ω^* represents the dual of the absolute conic on the plane at infinity. Under an affine projection by the camera P_a, Q^* maps to the dual image of the absolute conic (DIAC)

$$\omega^* = \mathsf{P}_a \mathsf{Q}^* \mathsf{P}_a^T = \mathsf{P}_a \begin{bmatrix} \Omega^* & \mathbf{0}_3 \\ \mathbf{0}_3^T & 0 \end{bmatrix} \mathsf{P}_a^T = \begin{bmatrix} \mathsf{M}\Omega^*\mathsf{M}^T & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{bmatrix} = \begin{bmatrix} \mathsf{A}\mathsf{A}^T & \mathbf{0}_2 \\ \mathbf{0}_2^T & 0 \end{bmatrix}, \tag{A.4}$$

which in the affine case is a degenerate line conic of rank 2, consisting of the two pencils of lines defined by each of the circular points $\mathbf{i}, \mathbf{j}$ on the line at infinity $\mathbf{l}_\infty^T$ in the image.

B

Notes on Affine Multifocal Relations

B.1 Expansion of the Irreducible Tensor Components in Terms of the Registered Tensors

Applying a Laplace expansion along the fourth column of the irreducible minors of section 6.4, gives the following relation to the minors of the joint registered projection matrix of section 6.3

$$
\begin{aligned}
{}_{IJ}F_{33} &= m_J^{24} \cdot {}_{IJ}e^1 - m_J^{14} \cdot {}_{IJ}e^2 + m_I^{24} \cdot {}_{JI}e^1 - m_I^{14} \cdot {}_{JI}e^2 \\
&= -m_J^{j4} \cdot {}^{IJ}e_j - m_I^{i4} \cdot {}^{JI}e_i, \qquad \text{(B.1)} \\
{}_I^{JK}T_3^{jk} &= m_K^{k4} \cdot {}_{IJ}e^j - m_J^{j4} \cdot {}_{IK}e^k + m_I^{24} \cdot {}^{IJK}t^{1jk} - m_I^{14} \cdot {}^{IJK}t^{2jk} \\
&= m_K^{k4} \cdot {}_{IJ}e^j - m_J^{j4} \cdot {}_{IK}e^k - m_I^{i4} \cdot {}_I^{JK}t_i^{jk}, \qquad j,k \in \{1,2\}, \\
{}^{IJKL}Q^{ijkl} &= m_L^{l4} \cdot {}^{IJK}t^{ijk} + m_K^{k4} \cdot {}^{LIJ}t^{lij} + m_J^{j4} \cdot {}^{KLI}t^{kli} + m_I^{i4} \cdot {}^{JKL}t^{jkl}, \qquad i,j,k,l \in \{1,2\}.
\end{aligned}
$$

B.2 Relations to Known Trilinear Constraints

B.2.1 Torr's Affine Trifocal Constraint

Comparison between the coefficients of the affine trifocal constraint equations defined by [Tor95], and the affine trifocal tensor (section 6.4.3) establishes the following one to one correspondence between the coefficients $p_1, \ldots, p_{16}$ and the non-vanishing components of the unregistered trifocal tensor

$$\left\{T_1^{jk}\right\} = \begin{bmatrix} -p_5 & -p_6 & 0 \\ -p_7 & -p_8 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\left\{T_2^{jk}\right\} = \begin{bmatrix} -p_9 & -p_{10} & 0 \\ -p_{11} & -p_{12} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\left\{T_3^{jk}\right\} = \begin{bmatrix} -p_{13} & -p_{15} & -p_1 \\ -p_{14} & -p_{16} & -p_2 \\ -p_3 & -p_4 & 0 \end{bmatrix}.$$

Torr [Tor95] gives the *three view constraint for the affine camera* between three images of a single point, generalizing a result obtained by [UB91] for orthography

$$\begin{aligned} p_1 x_K^1 + p_5 x_I^1 + p_9 x_I^2 + p_3 x_J^1 + p_{13} &= 0 \\ p_2 x_K^1 + p_7 x_I^1 + p_{11} x_I^2 + p_3 x_J^2 + p_{14} &= 0 \\ p_1 x_K^2 + p_6 x_I^1 + p_{10} x_I^2 + p_4 x_J^1 + p_{15} &= 0 \\ p_2 x_K^2 + p_8 x_I^1 + p_{12} x_I^2 + p_4 x_J^2 + p_{16} &= 0, \end{aligned}$$

for inhomogeneous image coordinates. Given the definition of the coefficients p_i, as they appear in [Tor95], it is easily verified that the above relations contain errors of sign, and should read

$$\begin{aligned} -p_1 x_K^1 + p_5 x_I^1 + p_9 x_I^2 - p_3 x_J^1 + p_{13} &= 0 \\ -p_2 x_K^1 + p_7 x_I^1 + p_{11} x_I^2 - p_3 x_J^2 + p_{14} &= 0 \\ -p_1 x_K^2 + p_6 x_I^1 + p_{10} x_I^2 - p_4 x_J^1 + p_{15} &= 0 \\ -p_2 x_K^2 + p_8 x_I^1 + p_{12} x_I^2 - p_4 x_J^2 + p_{16} &= 0 \end{aligned}$$

or

$$\begin{aligned}
T_1^{11}x_I^1 + T_2^{11}x_I^2 - T_3^{31}x_J^1 - T_3^{13}x_K^1 + T_3^{11} &= 0 \\
T_1^{21}x_I^1 + T_2^{21}x_I^2 - T_3^{31}x_J^2 - T_3^{23}x_K^1 + T_3^{21} &= 0 \\
T_1^{12}x_I^1 + T_2^{12}x_I^2 - T_3^{32}x_J^1 - T_3^{13}x_K^2 + T_3^{12} &= 0 \\
T_1^{22}x_I^1 + T_2^{22}x_I^2 - T_3^{32}x_J^2 - T_3^{23}x_K^2 + T_3^{22} &= 0.
\end{aligned}$$

Since the inhomogenous image coordinates imply $x_I^3 = 1$, the four equations can be written together as

$$T_i^{jk}x_I^i - T_3^{3k}x_J^j - T_3^{j3}x_K^k = 0^{jk}, \qquad j,k \in \{1,2\},$$

which is exactly (6.39).

We will now proceed in a manner analogous to [QOM98] to obtain a single constraint in the registered trifocal tensor. Multiplying the four equations by $x_J^1x_K^1$, $-x_J^1x_K^2$, $-x_J^2x_K^1$, and $x_J^2x_K^2$ respectively, and adding all four equations,

$$\begin{aligned}
T_1^{22}x_I^1x_J^1x_K^1 + T_2^{22}x_I^2x_J^1x_K^1 + T_3^{32}x_J^2x_J^1x_K^1 + T_3^{23}x_K^2x_J^1x_K^1 + T_3^{22}x_J^1x_K^1 &+ \\
-T_1^{21}x_I^1x_J^1x_K^2 - T_2^{21}x_I^2x_J^1x_K^2 - T_3^{31}x_J^2x_J^1x_K^2 - T_3^{23}x_K^1x_J^1x_K^2 - T_3^{21}x_J^1x_K^2 &+ \\
-T_1^{12}x_I^1x_J^2x_K^1 - T_2^{12}x_I^2x_J^2x_K^1 - T_3^{32}x_J^1x_J^2x_K^1 - T_3^{13}x_K^2x_J^2x_K^1 - T_3^{12}x_J^2x_K^1 &+ \\
T_1^{11}x_I^1x_J^2x_K^2 + T_2^{11}x_I^2x_J^2x_K^2 + T_3^{31}x_J^1x_J^2x_K^2 + T_3^{13}x_K^1x_J^2x_K^2 + T_3^{11}x_J^2x_K^2 &= 0,
\end{aligned}$$

cancels all terms involving T_3^{j3} and T_3^{3k}, leaving us with

$$\begin{aligned}
T_1^{22}x_I^1x_J^1x_K^1 + T_2^{22}x_I^2x_J^1x_K^1 + T_3^{22}x_J^1x_K^1 &+ \\
-T_1^{21}x_I^1x_J^1x_K^2 - T_2^{21}x_I^2x_J^1x_K^2 - T_3^{21}x_J^1x_K^2 &+ \\
-T_1^{12}x_I^1x_J^2x_K^1 - T_2^{12}x_I^2x_J^2x_K^1 - T_3^{12}x_J^2x_K^1 &+ \\
T_1^{11}x_I^1x_J^2x_K^2 + T_2^{11}x_I^2x_J^2x_K^2 + T_3^{11}x_J^2x_K^2 &= 0,
\end{aligned}$$

which by dualizing the trifocal tensor in the indices J and K, can be written more compactly as

$$_{IJK}t_{ijk}x_I^ix_J^jx_K^k + {}_{IJK}T_{3jk}x_J^jx_K^k = 0,$$

observing that indices are in the range $\{1,2\}$. For registered cameras this relation simplifies to

$$_{IJK}t_{ijk}x_I^ix_J^jx_K^k = 0.$$

The last result was obtained by [QOM98]. Here the relation to the tensor to ${}_I^{JK}T$ is made explicit. Torr [Tor95] also gives the affine line transfer equations and two constraints for line correspondence over three affine views.

B.2.2 Quan's Affine Trilinear Form

Quan et al. [QOM98] give the registered constraint equations

$$\begin{aligned} t_4x_I^1 + t_8x_I^2 + t_{11}x_J^1 + t_9x_K^1 &= 0 \\ t_2x_I^1 + t_6x_I^2 + t_{11}x_J^2 + t_{10}x_K^1 &= 0 \\ t_3x_I^1 + t_7x_I^2 + t_{12}x_J^1 + t_9x_K^2 &= 0 \\ t_1x_I^1 + t_5x_I^2 + t_{12}x_J^2 + t_{10}x_K^2 &= 0. \end{aligned}$$

They definine a tensor T_{ijk} by the first eight coefficients $t_1, \ldots, t_8$ to obtain

$$\begin{aligned} T_{122}x_I^1 + T_{222}x_I^2 + t_{11}x_J^1 + t_9x_K^1 &= 0 \\ T_{121}x_I^1 + T_{212}x_I^2 + t_{11}x_J^2 + t_{10}x_K^1 &= 0 \\ T_{112}x_I^1 + T_{221}x_I^2 + t_{12}x_J^1 + t_9x_K^2 &= 0 \\ T_{111}x_I^1 + T_{211}x_I^2 + t_{12}x_J^2 + t_{10}x_K^2 &= 0. \end{aligned}$$

By eliminating the remaining coefficients $t_9, \ldots, t_{12}$, they arrive at the trilinear form [QOM98]

$$T_{ijk}x_I^i x_J^j x_K^k = 0.$$

C

The Quadratic p-relations of Four Views

C.1 The Quadratic p-relations

The fact that the quadrifocal tensor for registered views consists of exactly the four registered trifocal tensors between the four views (6.17), prompts us to look for constraints in the common information shared between the trifocal tensors.

Noting that the registered unifocal and trifocal tensors obtained from four views are the *Grassmann-coordinates* [HP47, Tri95b, Hey98] of the 8×3 inhomogeneous joint projection matrix of four views

$$\begin{bmatrix} \mathsf{M}_I \\ \mathsf{M}_J \\ \mathsf{M}_K \\ \mathsf{M}_L \end{bmatrix},$$

we start by looking at the *quadratic p-relations* [HP47, Hey98] existing between Grassmann-coordinates (also known as the *Grassmann simplicity relations* [Tri95b] or the *Plücker relations* (?) [GH78]). These relations restrict the $\binom{8}{3} = 56$ dimensional space of 3-index skew tensors to a 15 dimensional quadratic sub-variety that exactly parameterizes the possible subspaces [Tri95b]. The 56 Grassmann coordinates are exactly the $4 \cdot 8 = 32$ components

of the four reduced trifocal tensors along with the $12 \cdot 2 = 24$ components of the $4 \cdot 3 = 12$ registered unifocal tensors (or epipoles) between the four views.

For a three-index tensor the quadratic p-relations are given by [HP47]

$$p_{i_1 i_2 j_1} p_{j_2 j_3 j_4} - p_{i_1 i_2 j_2} p_{j_1 j_3 j_4} + p_{i_1 i_2 j_3} p_{j_1 j_2 j_4} - p_{i_1 i_2 j_4} p_{j_1 j_2 j_3} = 0.$$

Choosing the six indices blindly results in the staggering number of 8^6 relations. As we will now show, most of these relations are equivalent.

How Many Quadratic p-relations?

As made explicit in the form of the quadratic p-relations given in [Tri95b],

$$u^{i_1 i_2 [j_1} u^{j_2 j_3 j_4]} = 0,$$

the quadratic p-relations involve fixed i's and anti-symmetrization [Tri95b] over the j's. Hence changing the order of either the i's or the j's does not affect the relation. Repeated i's, i.e. $i_1 = i_2$, make the constraint vanish trivially. Repeated j's, say $j_1 = j_2$ also reduce the constraint to a triviality

$$p_{i_1 i_2 j_1} p_{j_1 j_3 j_4} - p_{i_1 i_2 j_1} p_{j_1 j_3 j_4} = 0.$$

Sharing one index between the two sets, i.e. $j_1 = i_1$, results in a three term constraint (see also [HP47]).

$$-p_{i_1 i_2 j_2} p_{i_1 j_3 j_4} + p_{i_1 i_2 j_3} p_{i_1 j_2 j_4} - p_{i_1 i_2 j_4} p_{i_1 j_2 j_3} = 0,$$

whereas sharing two indices $j_1 = i_1$, $j_2 = i_2$, reduces the constraint to a triviality

$$p_{i_1 i_2 j_3} p_{i_1 i_2 j_4} - p_{i_1 i_2 j_4} p_{i_1 i_2 j_3} = 0.$$

From these observations we have cut down the number of p-relations we need to examine to $\binom{n}{2} \cdot \binom{n-2}{4} = \binom{8}{2} \cdot \binom{6}{4} = 28 \cdot 15 = 420$ four term relations and $\binom{n}{2} \cdot \binom{2}{1} \cdot \binom{n-2}{3} = \binom{8}{2} \cdot \binom{2}{1} \cdot \binom{6}{3} = 28 \cdot 2 \cdot 20 = 1120$ three term relations.

In addition to the above observations, it is easy to show that the following identity holds

$$i_1 i_2 [j_1 j_2 j_3 j_4] = j_1 j_2 [i_1 i_2 j_3 j_4] + j_3 j_4 [i_1 i_2 j_1 j_2].$$

For a shared index, this identity reduces to

$$i_1 i_2 [i_1 j_2 j_3 j_4] = i_1 j_2 [i_1 i_2 j_3 j_4] + j_3 j_4 [i_1 i_2 i_1 j_2],$$

with the rightmost term vanishing, giving

$$i_1 i_2 [i_1 j_2 j_3 j_4] = i_1 j_2 [i_1 i_2 j_3 j_4].$$

It follows, that we can count i_2 as one of the j's, and rewrite the shared index relations as

$$u^{ij_1[i} u^{j_2 j_3 j_4]} = 0,$$

where the order of j's does not affect the relation. Hence the number of three term relations we need to examine is down to $\binom{n}{1} \cdot \binom{n-1}{4} = \binom{8}{1} \cdot \binom{7}{4} = 8 \cdot 35 = 280$.

C.2 The Quadratic p-relations of Four Registered Views

Using the above observations, we attempt to list the classes of equivalent quadratic p-relations, that exist between the Grassmann coordinates of four registered views, modulo a permutation of the images. This is the approach taken by [Tri95a], in the case of three projective views. We divide the constraints into three and four view constraints, depending on whether they involve rows from three or four projection matrices.

Quadratic p-relations of Three Views

The quadratic p-relations of three views can be divided into three types (I - III)

$$I_1 I_2 [J_1 J_2 K_1 K_2], \quad I_1 I_2 [I J_1 J_2 K], \quad I J_1 [I J_2 K_1 K_2].$$

Written in terms of the registered tensors, they are

$$\begin{aligned} {}_{IJ}e^1{}_{KJ}e^2 - {}_{IJ}\, e^2{}_{KJ}e^1 + {}_{IK}\, e^1{}_{JK}e^2 - {}_{IK}\, e^2{}_{JK}e^1 &= 0, \\ -{}_{IJ}e^1 \cdot t^{i2k} + {}_{IJ}\, e^2 \cdot t^{i1k} - {}_{IK}\, e^k \cdot {}_{JI}\, e^i &= 0^{ik}, \\ -{}_{JI}e^i{}_{KI}e^i + t^{i11} \cdot t^{i22} - t^{i12} \cdot t^{i21} &= 0. \end{aligned}$$

Quadratic p-relations of Four Views

In addition to the quadratic p-relations of three views, we have the following six types (I - VI)

$$I_1I_2[J_1J_2KL], \quad I_1J_1[I_2J_2KL], \quad I_1J[I_2K_1K_2L],$$
$$IJ[K_1K_2L_1L_2], \quad I_1I_2[IJKL], \quad IJ_1[IJ_2KL].$$

Written in terms of the registered tensors, they are

$$\begin{array}{rrcrrcrrcrrcl}
{}_{IJ}e^1 & {}^{JKL}t^{2kl} & - & {}_{IJ}e^2 & {}^{JKL}t^{1kl} & + & {}_{IK}e^k & {}_{JL}e^l & - & {}_{IL}e^l & {}_{JK}e^k & = & 0^{kl}, \\
-{}_{IJ}e^1 & {}^{JKL}t^{2kl} & - & {}_{JI}e^1 & {}^{IKL}t^{2kl} & + & {}^{IJK}t^{11k} & {}^{IJL}t^{22l} & - & {}^{IJL}t^{11l} & {}^{IJK}t^{22k} & = & 0^{kl}, \\
-{}_{IJ}e^j & {}_{KL}e^l & - & {}^{IJK}t^{1j1} & {}^{IKL}t^{22l} & + & {}^{IJK}t^{1j2} & {}^{IKL}t^{21l} & - & {}^{IJL}t^{1jl} & {}_{KI}e^2 & = & 0^{jl}, \\
{}^{IJK}t^{ij1} & {}_{LK}e^2 & - & {}^{IJK}t^{ij2} & {}_{LK}e^1 & + & {}^{IJL}t^{ij1} & {}_{KL}e^2 & - & {}^{IJL}t^{ij2} & {}_{KL}e^1 & = & 0^{ij}, \\
 & & & -{}_{IJ}e^j & {}^{IKL}t^{ikl} & + & {}_{IK}e^k & {}^{IJL}t^{ijl} & - & {}_{IL}e^l & {}^{IJK}t^{ijk} & = & 0^{ijkl}, \\
 & & & -{}_{JI}e^I & {}^{IKL}t^{Ikl} & + & {}^{IJK}t^{I1k} & {}^{IJL}t^{I2l} & - & {}^{IJL}t^{I1l} & {}^{IJK}t^{I2k} & = & 0^{Ikl},
\end{array}$$

where the capital index in the last relation indicates that summation is not implied over the repeated index.

Bibliography

[ABB+99] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, third edition, 1999.

[ÅHKO99] Kalle Åström, Anders Heyden, Fredrik Kahl, and Magnus Oskarsson. Structure and motion from lines under affine projections. In *Proc. 7th Int'l Conf. on Computer Vision* [IEE99].

[Aka74] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, 19(6):716–723, 1974.

[And63] T. W. Anderson. Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, 34:122–148, 1963.

[And84a] T. W. Anderson. Estimating linear statistical relationships. *Annals of Statistics*, 12:1–45, 1984.

[And84b] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley series in probability and mathematical statistics. Wiley, New York, 2nd edition, 1984.

[AR56] T. W. Anderson and Herman Rubin. Statistical inference in factor analysis. In Jerzy Neyman, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume V, pages 111–150, Berkeley and Los Angeles, 1956. University of California.

[AS98] Shai Avidan and Amnon Shashua. Threading fundamental matrices. In Burkhardt and Neumann [BN98], pages 124–140.

[AZH96] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In Buxton and Cipolla [BC96], pages 3–16.

[BA75] Edwin H. Bryant and William R. Atchley, editors. *Multivariate Statistical Methods: Within-Groups Covariation*, volume 2 of *Benchmark Papers in Systematic and Evolutionary Biology*. Dowden, Hutchinson and Ross; Halsted Press, Stroudsburg, Pa; New York, 1975.

[Bas96] Ronen Basri. Paraperspective $\equiv$ affine. *International Journal of Computer Vision*, 19(2):169–179, 1996.

[Bau00] Adam Baumberg. Reliable feature matching across widely separated views. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 774–781, Hilton Head Island, South Carolina, June 2000. IEEE Computer Society Press, Los Alamitos, CA.

[BC96] B. Buxton and R. Cipolla, editors. *Proc. 4th European Conf. on Computer Vision*, LNCS 1064/1065, Cambridge, April 1996. Springer-Verlag.

[BL98] Lars Bretzner and Tony Lindeberg. Use your hand as a 3-d mouse, or, relative orientation from extended sequences of sparse point and line correspondences using the affine trifocal tensor. In Burkhardt and Neumann [BN98], pages 141–157.

[Blo79] D. B. Bloom. *Linear Algebra and Geometry*. Cambridge University Press, 1979.

[BN98] H. Burkhardt and B. Neumann, editors. *Proc. 5th European Conf. on Computer Vision*, LNCS 1406/1407, Freiburg, Germany, June 1998. Springer-Verlag.

[BNB96] Robert Bartoszyński and Magdalena Niewiadomska-Bugaj. *Probability and Statistical Inference*. Wiley series in probability and statistics. Wiley-Interscience, New York, 1996.

[Bow66] T. G. R. Bower. The visual world of infants. *Scientific American*, 215:80–92, December 1966.

[Bow74] T. G. R. Bower. *Development in Infancy*. W. H. Freeman, San Francisco, 1974.

[Boz87] Hamparsum Bozdogan. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

[BR79] H. B. Barlow and B. C. Reeves. The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19:783–793, 1979.

[BSF88] Yaakov Bar-Shalom and Thomas E. Fortmann. *Tracking and Data Association*. Academic Press, San Diego, London, 1988.

[BZM95] P.A. Beardsley, A.P. Zisserman, and D.W. Murray. Sequential update of projective and affine structure from motion. Technical report OUEL 2070/95, Dept of Eng Science, University of Oxford, 1995.

[Can00] Nikos Canterakis. A minimal set of constraints for the trifocal tensor. In Vernon [Ver00], pages I.84–99.

[CVP97] *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, June 1997. IEEE Computer Society Press, Los Alamitos, CA.

[dAHR98] Lourdes de Agapito, Eric Hayman, and Ian Reid. Self-calibration of a rotating camera with varying intrinsic parameters. In Nixon and Carter [NC98], pages 105–114.

[Ekl94] J. O. Eklundh, editor. *Proc. 3rd European Conf. on Computer Vision*, LNCS 800/801, Stockholm, May 1994. Springer-Verlag.

[Fau92] Olivier D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In Sandini [San92], pages 563–578.

[Fau93] O. D. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, Cambridge, MA, 1993.

[FB81] M.A. Fischler and R.C. Bolles. Random sample concensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.

[Fis38] R. A. Fisher. The statistical utilization of multiple measurement. *Annals of Eugenics*, 8:376–386, 1938.

[FLM92] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In Sandini [San92], pages 321–334.

[FM95] Olivier Faugeras and Bernard Mourrain. On the geometry and algebra of the point and line correspondences between N images. In *Proc. 5th Int'l Conf. on Computer Vision* [IEE95], pages 951–956.

[FP98] Olivier Faugeras and Théodore Papadopoulo. Grassmann-Cayley algebra for modelling systems of cameras and the algebraic equations of the manifold of trifocal tensors. *Philosophical Transactions of the Royal Society of London. Series A*, 356(1740):1123–1152, 1998.

[FZB94] R. Fawcett, A. Zisserman, and J. M. Brady. Extracting structure from an affine view of a 3D point set with one or two bilateral symmetries. *Image and Vision Computing*, 12(9):615–622, November 1994.

[GH78] Phillip Griffiths and Joseph Harris. *Principles of Algebraic Geometry*. A Wiley-Interscience publication. Wiley, New York, 1978.

[GLP84] W. E. L. Grimson and T. Lozano-Pérez. Model-based recognition and localization from sparse range or tactile data. *International Journal of Robotics Research*, 3(3):3–35, 1984.

[Gri90] W. E. L. Grimson. *Object Recognition by Computer: The Role of Geometric Constraints*. MIT Press, Cambridge, MA, 1990.

[Har92] Richard I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In Sandini [San92], pages 579–587.

[Har94a] R. M. Haralick. Performance characterization in computer vision. *CVGIP: Image Understanding*, 60(2):245–249, September 1994.

[Har94b] R. I. Hartley. Self-calibration from multiple views with a rotating camera. In Eklundh [Ekl94], pages 471–478.

[Har95a] R.I. Hartley. In defence of the 8-point algorithm. In *Proc. 5th Int'l Conf. on Computer Vision* [IEE95], pages 1064–1070.

[Har95b] R.I. Hartley. A linear method for reconstruction from lines and point s. In *Proc. 5th Int'l Conf. on Computer Vision* [IEE95], pages 882–887.

[Har97] Richard I. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1997.

[Har98] Richard I. Hartley. Minimizing algebraic error. In *Proc. 6th Int'l Conf. on Computer Vision* [IEE98], pages 469–476.

[Hay00] Eric Hayman. *The Use of Zoom within Active Vision*. PhD thesis, Department of Engineering Science, University of Oxford, 2000.

[Hey98] Anders Heyden. A common framework for multiple view tensors. In Burkhardt and Neumann [BN98], pages 3–19.

[HGC92] R. I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 761–764, Champaign, June 1992. IEEE Computer Society Press, Los Alamitos, CA.

[Hor86] Berthold Klaus Paul Horn. *Robot Vision*. MIT Press, Cambridge, Mass., 1986.

[Hot33] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24(6):417–441, 1933. First of two parts. Reprinted in [BA75].

[HP47] W. V. D. Hodge and D. Pedoe. *Methods of Algebraic Geometry*, volume 1. Cambridge University Press, Cambridge, 1947.

[HS88] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conf., Manchester*, pages 147–151, 1988.

[HS97] Richard I. Hartley and Peter Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[Hsu41] P. L. Hsu. On the problem of rank and the limiting distribution of Fisher's test function. *Annals of Eugenics*, 11:39–41, 1941.

[HTM99] Eric Hayman, Torfi Thórhallsson, and David W. Murray. Zoom-invariant tracking using points and lines in affine views - an application of the multifocal tensors. In *Proc. 7th Int'l Conf. on Computer Vision* [IEE99], pages 269–276.

[HY38] A. S. Householder and Gale Young. Matrix approximation and latent roots. *American Mathematical Monthly*, 45:165–171, 1938.

[HZ00] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2000.

[IA00] Michal Irani and P. Anandan. Factorization with uncertainty. In Vernon [Ver00], pages I.539–553.

[IEE93] IEEE. *Proc. 4th Int'l Conf. on Computer Vision*, Berlin, June 1993. IEEE Computer Society Press, Los Alamitos, CA.

[IEE95] IEEE. *Proc. 5th Int'l Conf. on Computer Vision*, Cambridge, MA, June 1995. IEEE Computer Society Press, Los Alamitos, CA.

[IEE98] IEEE. *Proc. 6th Int'l Conf. on Computer Vision*, Bombay, India, January 1998. Narosa Publishing House.

[IEE99] IEEE. *Proc. 7th Int'l Conf. on Computer Vision*, Kerkyra, Greece, September 1999. IEEE Computer Society Press, Los Alamitos, CA.

[JB91] Y. Jiang and H. Bunke. Determination of the symmetries of polyhedra and an application to object recognition. In *Proc. Int. Workshop on Computational Geometry, Bern*, pages 113–121, March 1991.

[Jen83] B. Jenkins. Component processes in the perception of bilaterally symmetric dot textures. *Psychophysiology*, 34:433–440, 1983.

[Joh57] M. Johansen. The experienced continuations of some three-dimensional forms. *Acta Psychologica*, 13:1–26, 1957.

[Kah99] Fredrik Kahl. Private communication, June 1999.

[Kan81] T. Kanade. Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17:409–460, 1981.

[Kan96] Kenichi Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, Amsterdam, 1996.

[Kar46] K. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae, ser. A1, Math. Phys.*, 37, 1946.

[Kep96] Johannes Kepler. *Mysterium Cosmographicum*. Tübingen, 1596.

[KG98] Nahum Kiryati and Yossi Gofman. Detecting symmetry in grey level images: The global optimization approach. *International Journal of Computer Vision*, 29(1):29–45, 1998.

[KH95] Tapas Kanungo and Robert M. Haralick. Multivariate hypothesis testing for gaussian data: Theory and software. Tech Report ISL-TR-95-05, Intelligent Systems Laboratory, Department of Electrical Engineering, University of Washington, Seattle, WA, October 1995.

[KH98] Fredrik Kahl and Anders Heyden. Structure and motion from points, lines and conics with affine cameras. In Burkhardt and Neumann [BN98], pages 327–341.

[KJPH93] T. Kanungo, M. Y. Jaisimha, J. Palmer, and R. M. Haralick. A quantitative methodology for analyzing the performance of detection algorithms. In *Proc. 4th Int'l Conf. on Computer Vision* [IEE93], pages 247–252.

[KM94] C. Kohl and J. Mundy. The development of the image understanding environment. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1994.

[KS73] Maurice G. Kendall and Alan Stuart. *The Advanced Theory of Statistics*, volume 2: Inference and Relationship. Griffin, London, third edition, 1973.

[Kv91] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *Journal of the Optical Society of America. A, Optics and image science*, 8(2):377–385, February 1991.

[Law40] D. N. Lawley. The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, Sec. A*, 60:64–82, 1940.

[LH81] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[LM63] D. N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. Butterworths Mathematical Texts. Butterworths, London, 1963.

[LM71] D. N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. Butterworths, London, 2nd edition, 1971.

[Loè55] M. M. Loève. *Probability Theory*. Van Nostrand, Princeton, NJ, 1955.

[Low85] D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, MA, 1985.

[LZ99] David Liebowitz and Andrew Zisserman. Combining scene and auto-calibration constraints. In *Proc. 7th Int'l Conf. on Computer Vision* [IEE99], pages 293–300.

[Mac86] Ernst Mach. *Beiträge zur Analyse der Empfindungen*. Gustav Fischer, Jena, 1886.

[Man89] Joseph Manning. Computational complexity of geometric symmetry detection in graphs. In *Proc. of the First Great Lakes Comp. Sci. Conf. Kalamazoo, Mich.*, volume 507 of *Lecture Notes in Comp. Sci.*, pages 1–7, New York, Berlin, Heidelberg, oct 1989. Springer-Verlag.

[MC98] P. R. S. Mendonça and R. Cipolla. Analysis and computation of an affine trifocal tensor. In Nixon and Carter [NC98], pages 125–133.

[McL94] Philip F. McLauchlan. Horatio: Libraries for vision applications. Technical report OUEL 1967/92 (revised), Dept. of Engineering Science, University of Oxford, 1994.

[McL99] Philip F. McLauchlan. The variable state dimension filter applied to surface-based structure from motion. CVSSP Technical Report VSSP-TR-4/99, School of Electrical Engineering, Information Technology and Mathematics, University of Surrey, Guildford, 1999.

[MKB79] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, San Diego, 1979.

[MM95] Philip F. McLauchlan and David W. Murray. A unifying framework for structure and motion recovery from image sequences. In *Proc. 5th Int'l Conf. on Computer Vision* [IEE95].

[MM96] P.F. McLauchlan and D.W. Murray. Active camera calibration for a head-eye platform using the variable state-dimension filter. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(1):15–22, January 1996.

[MRM94] P. F. McLauchlan, I. D. Reid, and D. W. Murray. Recursive affine structure and motion from image sequences. In Eklundh [Ekl94], pages 217–224.

[Mui82] Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley series in probability and mathematical statistics. Wiley, New York, 1982.

[MZ92a] J. L. Mundy and A. P. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, MA, 1992.

[MZ92b] J. L. Mundy and A. P. Zisserman. Towards a new framework for vision. In *Geometric Invariance in Computer Vision* [MZ92a], pages 1–39.

[MZB95] Dipti Prasad Mukherjee, Andrew Zisserman, and Michael Brady. Shape from symmetry: detecting and exploiting symmetry in affine images. *Philosophical Transactions of the Royal Society of London. Series A*, 351:77–106, 1995.

[Näh93] S. Näher. LEDA manual. Technical Report MPI-I-93-109, Max-Planck-Institut für Informatik, Saarbrücken, 1993.

[NC98] M. Nixon and J. Carter, editors. *Proc. 9th British Machine Vision Conf.*, Southampton, September 1998. BMVA.

[Nom66] Katsumi Nomizu. *Fundamentals of Linear Algebra*. McGraw-Hill, 1st edition, 1966.

[Pea01] Karl Pearson. On lines and planes of closest fit to systems of points in space. *Phil. Mag., Ser. 6*, 2(11):559–572, 1901. Reprinted in [BA75].

[PF98] Théodore Papadopoulo and Olivier Faugeras. New characterization of the trifocal tensor. In Burkhardt and Neumann [BN98], pages 109–123.

[PKV98] Marc Pollefeys, Reinhard Koch, and Luc Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th Int'l Conf. on Computer Vision* [IEE98], pages 90–95.

[Pon89] J. Ponce. On characterizing ribbons and finding skewed symmetries. In *Proc. IEEE Int'l Conf. on Robotics and Automation*, pages 49–54, 1989.

[PTVP92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P.Flannery. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2nd edition, 1992.

[PZ98a] P. Pritchett and A. Zisserman. Matching and reconstruction from widely separated views. In R. Koch and L. Van Gool, editors, *3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506*, pages 78–92. Springer-Verlag, June 1998.

[PZ98b] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 754–760, January 1998.

[QK97] Long Quan and Takeo Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(8):834–845, 1997.

[QOM98] Long Quan, Yuichi Ohta, and Roger Mohr. Geometry of multiple affine views. In *Proc. European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE'98)*, volume 1506 of *Lecture Notes in Computer Science*, pages 32–46, Freiburg, Germany, June 1998. Springer-Verlag.

[Qua96] Long Quan. Self-calibration of an affine camera from multiple views. *International Journal of Computer Vision*, 19(1):93–105, 1996.

[RFZM93] C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J. L. Mundy. Extracting projective structure from single perspective views of 3D point sets. In *Proc. 4th Int'l Conf. on Computer Vision* [IEE93], pages 573–582.

[RM96] I. D. Reid and D. W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18:41–60, 1996.

[RN95] S. J. Russel and P. Norvig. *Artificial Intelligence : a modern approach*. Prentice-Hall, Englewood Cliffs, NJ., 1995.

[SA91] M. E. Spetsakis and J. Aloimonos. A multi-frame approach to visual motion perception. *International Journal of Computer Vision*, 16(3):245–255, 1991.

[San92] G. Sandini, editor. *Proc. 2nd European Conf. on Computer Vision*, LNCS 588, Santa Margharita Ligure, Italy, May 1992. Springer-Verlag.

[SC96] Bernt Schiele and James L. Crowley. Object recognition using multidimensional receptive field histograms. In Buxton and Cipolla [BC96], pages I.610–619.

[SG92] Ian Stewart and Martin Golubitsky. *Fearful symmetry : is God a geometer?* Blackwell, Oxford, 1992.

[Sha94] Amnon Shashua. Trilinearity in visual recognition by alignment. In Eklundh [Ekl94], pages 479–484.

[SK52] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry*. Clarendon Press, Oxford, 1952.

[SK79] M. S. Srivastava and C. G. Khatri. *An Introduction to Multivariate Statistics*. North Holland, New York, 1979.

[SM97] Cordelia Schmid and Roger Mohr. Local greyvalue invariants for image retrieval. *PAMI*, 19(5):x, 1997.

[SMB00] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):1–22, 2000.

[Spe04] Charles Spearman. "General-intelligence", objectively determined and measured. *American Journal of Psychology*, 15:201–293, 1904.

[Stu97] Peter Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In CVPR [CVP97], pages 1100–1105.

[SW00] Amnon Shashua and Lior Wolf. On the structure and properties of the quadrifocal tensor. In Vernon [Ver00], pages I.710–724.

[SZB95] L. S. Shapiro, A. Zisserman, and M. Brady. 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision*, 16(2):147–182, 1995.

[Thó96] T. Thórhallsson. Detecting bilateral symmetry of 3D point sets from affine views. In *Proc. 7th British Machine Vision Conf.*, pages 73–82, Edinburgh, September 1996. BMVA.

[Thó00] Torfi Thórhallsson. Optimal rectification of an affine structure from hypothesized orthogonality relations. In *Proc. 15th Int'l Conf. on Pattern Recognition*, pages I.396–401, Barcelona, September 2000. IEEE, IEEE Computer Society Press, Los Alamitos, CA.

[Thu31] L. L. Thurstone. Multiple factor analysis. *Psychol. Rev.*, 38:406–427, 1931. Reprinted in [BA75].

[Tin45] Gerhard Tintner. A note on rank, multicollinearity and multiple regression. *Annals of Mathematical Statistics*, 16:304–308, 1945.

[Tin46] Gerhard Tintner. Multiple regression for systems of equations. *Econometrica*, 14:5–36, 1946.

[TK92] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[TM97] P H S Torr and D W Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[TM99] Torfi Thórhallsson and David W. Murray. The tensors of three affine views. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 450–456, Fort Collins, June 1999. IEEE Computer Society Press, Los Alamitos, CA.

[Tor95] P.H.S. Torr. *Motion segmentation and outlier detection*. PhD thesis, Department of Engineering Science, University of Oxford, 1995.

[Tor98] P. H. S. Torr. Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society of London. Series A*, 356(1740):1321–1338, 1998.

[Tri95a] Bill Triggs. The geometry of projective reconstruction I: Matching constraints and the joint image. Unpublished report, July 1995.

[Tri95b] Bill Triggs. Matching constraints and the joint image. In *Proc. 5th Int'l Conf. on Computer Vision* [IEE95], pages 338–343.

[Tri97] Bill Triggs. Autocalibration and the absolute quadric. In CVPR [CVP97], pages 609–614.

[TV00] Tinne Tuytelaars and Luc Van Gool. Wide baseline stereo matching based on local, affine invariant regions. In Majid Mirmehdi and Barry Thomas, editors, *Proc. 11th British Machine Vision Conf.*, pages 412–425, Bristol, September 2000. BMVA.

[Tyl96a] C. W. Tyler, editor. *Human Symmetry Perception and its Computational Analysis*. VSP, Utrecht, The Netherlands, 1996.

[Tyl96b] Christopher W. Tyler. Human symmetry perception. In Tyler [Tyl96a], pages 3–22.

[TZ97] P H S Torr and A Zisserman. Robust parametrization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997.

[TZM98] P. H. S. Torr, A. Zisserman, and S. J. Maybank. Robust detection of degenerate configurations while estimating the fundamental matrix. *Computer Vision and Image Understanding*, 71(3):312–333, 1998.

[UB91] Shimon Ullman and Ronen Basri. Recognition by linear combinations of models. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(10):992–1006, 1991.

[Ver00] D. Vernon, editor. *Proc. 6th European Conf. on Computer Vision*, LNCS 1842/1843, Dublin, June 2000. Springer-Verlag.

[VFL96] Thierry Viéville, Olivier Faugeras, and Quang-Tuan Luong. Motion of points and lines in the uncalibrated case. *International Journal of Computer Vision*, 17(1):7–41, 1996.

[VL93] T. Viéville and Q.-T. Luong. Motion of points and lines in the uncalibrated case. Technical Report 2054, I.N.R.I.A., Sophia-Antipolis, 1993.

[VMUO95] L. Van Gool, T. Moons, D. Ungureanu, and A. Oosterlinck. The characterization and detection of skewed symmetry. *Computer Vision and Image Understanding*, 61(1):138–150, January 1995.

[VP96] Thomas Vetter and Tomaso Poggio. Symmetric 3D objects are an easy case for 2D object recognition. In Tyler [Tyl96a], pages 349–359.

[VT71] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory*. Wiley, New York, 1968-71.

[Wag95] Johan Wagemans. Detection of visual symmetries. *Spatial Vision*, 9:9–32, 1995. Reprinted in [Tyl96a], pages 25-48.

[Wat65] S. Watanabe. Karhunen-loève expansion and factor analysis theoretical remarks and applications. In *Proc. 4th Prague Conf. Inform. Theory*, 1965.

[Wey52] H. Weyl. *Symmetry*. Princeton University Press, Princeton, NJ, 1952.

[WVd91] J. Wagemans, L. Van Gool, and G. d'Ydewalle. Detection of symmetry in tachistoscopically presented dot patterns: Effects of muliple axes and skewing. *Perception and Psychophysics*, 50:413–427, 1991.

[WVd92] J. Wagemans, L. Van Gool, and G. d'Ydewalle. Orientational effects and component processes in symmetry detection. *The Quarterly Journal of Experimental Psychology*, 44A(3):475–508, 1992.

[Zis96] Andrew Zisserman. A users guide to the trifocal tensor. Unpublished report, 1996.

[ZLA98] Andrew Zisserman, David Liebowitz, and Martin Armstrong. Resolving ambiguities in auto-calibration. *Philosophical Transactions of the Royal Society of London. Series A*, 356(1740):1193–1211, 1998.

[ZPA95] H. Zabrodsky, S. Peleg, and D. Avnir. Symmetry as a continuous feature. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(12):1154–1166, 1995.

[ZW94] H. Zabrodsky and D. Weinshall. Utilizing symmetry in the reconstruction of three-dimensional shape from noisy images. In Eklundh [Ekl94], pages 403–410.

[ZW97] Hagit Zabrodsky and Daphna Weinshall. Using bilateral symmetry to improve 3D reconstruction from image sequences. *Computer Vision and Image Understanding*, 67:48–57, 1997. Previously published as [ZW94].