HAL
open science

Critical Motions in Euclidean Structure from Motion

Fredrik Kahl, Bill Triggs

To cite this version:

Fredrik Kahl, Bill Triggs. Critical Motions in Euclidean Structure from Motion. International Conference on Computer Vision & Pattern Recognition (CVPR '99), Jun 1999, Fort Collins, United States. pp.366–372, 10.1109/CVPR.1999.784661 . inria-00548315

HAL Id: inria-00548315

https://inria.hal.science/inria-00548315

Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Critical Motions in Euclidean Structure from Motion

Fredrik Kahl
Department of Mathematics
Lund University
Box 118, S-221 00 Lund, Sweden
fredrik@maths.lth.se

Bill Triggs
INRIA Rhône-Alpes
655 ave. de l'Europe, 38330 Montbonnot, France
http://www.inrialpes.fr/movi/people/Triggs
Bill.Triggs@inrialpes.fr

Abstract

We investigate the motions that lead to ambiguous Euclidean scene reconstructions under several common calibration constraints, giving a complete description of such *critical motions* for: (*i*) internally calibrated orthographic and perspective cameras; (*ii*) in two images, for cameras with unknown focal lengths, either different or equal. One aim of the work was to evaluate the potential of modern algebraic geometry tools for rigorously proving properties of vision algorithms, so we use ideal-theoretic calculations as well as classical algebra and geometry. We also present numerical experiments showing the effects of near-critical configurations for the varying and fixed focal length methods.

Keywords: structure from motion, critical motions, autocalibration, algebraic geometry.

1 Introduction

'Structure from Motion' (SFM) is the problem of recovering 3D scene geometry from several images. Using projective image measurements, it is only possible to recover structure, camera poses ('motion') and camera internal parameters ('calibrations') up to an unknown 3D projectivity [8, 5]. With additional scene, motion or calibration constraints, one can reduce the ambiguity to a Euclidean similarity [13, 4, 12, 7]. **Autocalibration** is the recovery of Euclidean structure, motion and calibration using partial (often qualitative) constraints on the camera calibrations, *e.g.* vanishing skew or equal focal lengths between images. It is useful because cameras often obey such constraints rather well, whereas — especially for hand-held cameras viewing unknown scenes — motion or structure assumptions are often rather dubious. Unfortunately, most autocalibration methods have situations in which they fail or are exceptionally weak. Practically, it is important to characterize and avoid these **critical sets**. Criticality is often independent of the specific camera calibrations, in which case we speak of **critical motions**.

'Classical' autocalibration assumes a moving projective camera with constant but unknown intrinsic parameters [4, 18, 1, 23, 17]. Sturm [19, 20] categorizes both the intrinsic and some algorithm-specific critical motions for this. The uniformity of the constraints makes this case relatively simple to analyze. But it is also somewhat unrealistic: it is often reasonable, *e.g.* to assume that the constant skew actually vanishes (a stronger constraint), whereas focal length often varies between images (a weaker constraint). Also, although he characterizes the degeneracies fully, Sturm only manages to give a rather implicit description of the corresponding critical motions. For practical purposes a more explicit description would be useful.

This paper derives explicit critical motions for Euclidean SFM under several simple two image 'unknown focal length' calibration constraints [6, 16, 24, 2, 9]. However, we start by giving a complete description of criticality for *known* calibrations, for both perspective and orthographic cameras in multiple images. Although this analysis does not result in any new ambiguities, it rules out the possibility of any further unknown ones.

A second goal of our work — one aspect of our European project CUMULI — was to investigate the use of formal algebraic reasoning tools to deduce rigorous properties of vision algorithms. Sturm [19] relies mainly on geometric intuition. This is unreliable in our less symmetrical situation and we have used a mixture of geometry, classical algebra, and ideal-theoretic algebraic geometry calculations (Gröbner bases, ideal quotient, radical and decomposition) in MAPLE and MACAULAY 2. However we will focus on giving geometric interpretations of our algebraic results whenever possible.

We consider only autocalibration degeneracies: scene and motion constraints are explicitly excluded from consideration. Also, for both projective and Euclidean reconstruction there are certain scene geometries for which SFM is inherently ambiguous [12, 15, 11, 10]. We exclude such **critical surfaces** by assuming that the scene is

Submitted to CVPR'99. This work was supported by Esprit LTR project CUMULI. [30/Nov/98]

generic enough to allow unambiguous recovery of projective structure. Hence, *criticality occurs iff the calibration constraints admit alternative Euclidean 'interpretations' of the given projective structure.*

2 Background

Image projection: We assume familiarity with the modern projective formulation of vision geometry [3, 12, 23]. A **perspective (pinhole) camera** is modeled in homogeneous coordinates by the projection equation $\boldsymbol{x} \simeq \mathbf{P}\mathbf{X}$ where $\mathbf{X} = (X, Y, Z, W)^\top$ is a 3D world point, $\boldsymbol{x} = (x, y, z)^\top$ is its 2D image and $\mathbf{P}$ is the 3×4 camera **projection matrix**. In a Euclidean frame $\mathbf{P}$ can be decomposed

$$\mathbf{P} = \boldsymbol{K}\boldsymbol{R}\,(\boldsymbol{I}_{3\times 3} | -\boldsymbol{t}) \qquad \boldsymbol{K} = \begin{pmatrix} f & fs & u_0 \\ 0 & fa & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

into a rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ encoding the camera's 3D pose (**extrinsic parameters**), and a 3×3 upper triangular **calibration matrix** $\boldsymbol{K}$ encoding its internal geometry. Here, f is the **focal length**, a the **aspect ratio**, s the **skew** and (u_0, v_0) the **principal point**.

Absolute Conic: Projective geometry encodes only collinearity and incidence. Affine structure (parallelism) is encoded projectively by singling out a **plane at infinity** Π_∞ of **direction vectors** or **points at infinity**, and Euclidean (similarity) structure by a proper virtual conic on Π_∞. This **absolute conic** Ω_∞ gives dot products between direction vectors. Its dual, the **dual absolute conic** Ω^*_∞, gives those between plane normals. Ω^*_∞ is a 4×4 symmetric rank 3 positive semidefinite contravariant matrix. $\Omega^*_\infty = \mathrm{diag}(1, 1, 1, 0)$ in any Euclidean frame. Π_∞ is Ω^*_∞'s unique null vector: $\Omega^*_\infty \Pi_\infty = 0$. Ω^*_∞'s image projection is $\omega^*_\infty \equiv \mathbf{P}\Omega^*_\infty \mathbf{P}^\top = \boldsymbol{K}\boldsymbol{K}^\top$, a dual image conic that encodes the camera calibration. $\boldsymbol{K}$ is recoverable from ω^*_∞ or its dual image point conic $\omega_\infty = {\omega^*_\infty}^{-1}$ by Cholesky factorization. ω^*_∞ and ω_∞ are proper virtual (positive definite) so long as the camera centre is finite. In calibrated image coordinates $\boldsymbol{K} = \boldsymbol{I}$, $\omega^*_\infty = \omega_\infty = \boldsymbol{I}$. We often use the abbreviations (D)(I)AC for (*D*ual)(*I*mage) *A*bsolute *C*onic.

False absolute conics: Given only a 3D projective reconstruction derived from uncalibrated images, the true absolute conic Ω_∞ is not distinguished in any way from any other proper virtual planar conic in projective space. In fact, given any such conic Ω^*, it is easy to find a 'rectifying' projective transformation that converts it to the Euclidean DAC form $\Omega^*_\infty = \mathrm{diag}(1, 1, 1, 0)$ and hence defines a false Euclidean structure. To recover the true structure, we need constraints that single out the true Ω_∞ and Π_∞ from all possible 'false' ones. In this paper we will constrain only the camera intrinsic parameters $\boldsymbol{K}_i$, or equivalently the images of the true absolute conic $\omega^*_{\infty i} = \boldsymbol{K}_i \boldsymbol{K}_i^\top$.

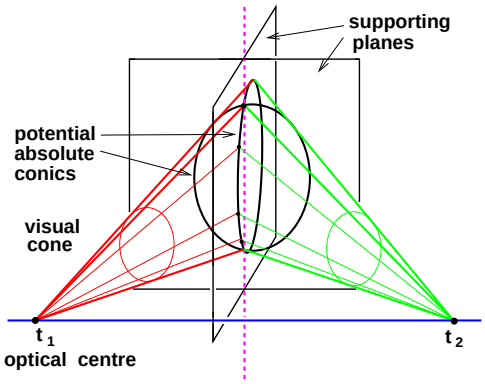

Figure 1: Intersecting the visual cones of two image conics satisfying the Kruppa constraints generates a pair of 3D conics.

The constraints may apply to individual image conics (*e.g.* vanishing skew $s = 0$), or link them as a group (*e.g.* equal but unknown focal lengths $f_i = f$ for all i). Ambiguity arises only if some non-absolute conic and its images satisfy the constraints. We call such conics **potential** or **false absolute conics**. They correspond one-to-one with possible false Euclidean structures for the scene. Ω denotes a potential 3D absolute conic, Ω^* its dual, ω its image and ω^* its dual image. True absolute conics are denoted $\Omega_\infty, \Omega^*_\infty, \omega_\infty, \omega^*_\infty$.

Affine camera: A camera whose optical plane coincides with Π_∞ is **affine** [14]. This is a good approximation for distant (and therefore large focal length) cameras viewing small objects. All visual rays except those on Π_∞ become parallel and the dual image absolute conic ω^*_∞ degenerates to rank 2. An **orthographic** camera is a calibrated affine one and has $\omega^*_\infty = \mathrm{diag}(1, 1, 0)$.

Kruppa constraints: Given image conics in several images, there may or may not be a 3D quadric having them as image projections. Constraints which guarantee this in two images are called **Kruppa constraints**. Any proper image conic is tangent to exactly two epipolar lines (possibly complex and/or coincident). It turns out [12, 3, 24] that there is a corresponding 3D quadric iff the tangent lines in the two images are in epipolar correspondence (see fig. 1). In fact, for non-coincident image centres and proper image conics satisfying the Kruppa constraints, there is always a linear one parameter family of 3D dual quadrics with these images. This family contains exactly two planar (rank 3) dual quadrics, and also the rank 2 one defined by (the symmetric outer product of) the two camera centres. If the image conics are virtual, the planar 3D quadrics are too and hence can serve as potential absolute conics. Thus: *In two images with distinct finite centres, a pair of proper virtual conics defines a potential 3D absolute conic iff it satisfies the Kruppa*

constraints, and in this case it always defines exactly two potential 3D absolute conics[1].

The Kruppa constraints have several algebraic formulations [12, 3, 24]. Below we will use the following 3×3 symmetric rank 2 matrix version linking the two dual image conics, the fundamental matrix and one epipole:

$$\boldsymbol{F}^\top \omega_2^* \boldsymbol{F} \;\simeq\; [\boldsymbol{e}]_\times \, \omega_1^* \, [\boldsymbol{e}]_\times^\top$$

This vanishes when dotted with the epipole and only holds up to scale, so it gives only two independent constraints.

3 Approach

We want to explicitly characterize the **critical motions** (relative camera placements) for which particular calibration constraints are insufficient to uniquely determine Euclidean 3D structure. We assume that projective structure is available. Alternative Euclidean structures correspond one-to-one with possible locations for the absolute conic in the projective reconstruction. Any proper virtual projective plane conic is potentially absolute, so we look for such conics Ω whose images also satisfy the given calibration constraints. There is ambiguity iff more than one such conic exists. We want *Euclidean* critical motions, so we work in a Euclidean frame where the true absolute conic Ω_∞ has its standard coordinates.

Several general properties help to simplify the problem:
Calibration invariance: The calibration constraints we use assert either equality between images, or that certain parameters have their 'calibrated' values $(f,a,s,u,v) = (1,1,0,0,0)$. They are satisfied for a set of cameras iff they are also satisfied when each image is premultiplied by its true inverse calibration $\boldsymbol{K}_i^{-1}$. Hence, we are free to assume that each camera is actually calibrated, $\boldsymbol{K}_i = \boldsymbol{I}$. The only difference from the fully calibrated case is that our weaker knowledge does not allow every false conic with $\omega_i^* \neq \boldsymbol{I}$ to be excluded outright.
Rotation invariance: For known-calibrated cameras $\omega_\infty^* = \boldsymbol{I}$, the image of any false AC must be identical to the image of the true one which is invariant to camera rotations. Hence, *criticality depends only on the camera centres, not on their orientations*. More generally, any camera rotation that leaves the calibration constraints intact is irrelevant. For example, arbitrary rotations about the optical axis and 180° flips about any axis in the optical plane are irrelevant if (a,s) is either $(1,0)$ or unconstrained, and (u_0,v_0) is either $(0,0)$ or unconstrained.
Translation invariance: For true or false absolute conics on the plane at infinity, translations are irrelevant so criticality depends only on camera orientation.

[1] With more than two images the situation is more delicate and the pairwise Kruppa constraints are *not* always sufficient to guarantee the existence of a corresponding 3D quadric.

In essence, Euclidean structure recovery in projective space is a matter of parametrizing all of the possible proper virtual plane conics, then using the calibration constraints on their images to algebraically eliminate parameters until only the unique true absolute conic remains. More abstractly, if $\mathbf{C}$ parametrizes the possible conics and $\mathbf{X}$ the camera geometries, the constraints cut out some algebraic variety in $(\mathbf{C},\mathbf{X})$ space. A constraint set is useful for Euclidean SFM only if this variety generically intersects the subspaces $\mathbf{X} = \mathbf{X}_0$ in one (or at most a few) points $(\mathbf{C},\mathbf{X}_0)$, as each such intersection represents an alternative Euclidean structure for the reconstruction from that camera geometry. A set of camera poses $\mathbf{X}$ is **critical** for the constraints if it has exceptionally (*e.g.* infinitely) many intersections.

For elimination calculations, algebraic varieties are described by **ideals** (the sets of polynomials that vanish on them), which in turn are characterized by certain 'exhaustive' polynomial sets called **Gröbner bases**. Varieties can also be **decomposed** into irreducible components — a generalization of polynomial factorization that we often use as an aid to interpreting results. These are all 'standard' algebraic geometry calculations available in specialized tools like MACAULAY 2 (http://www.math.uiuc.edu/Macaulay2/) and SINGULAR, and in slightly less powerful form in general systems like MAPLE.

Potential absolute conics can be represented in several ways. The following parametrizations have all proven relatively tractable:

(i) Choose a Euclidean frame in which Ω^* is diagonal, and express all camera poses w.r.t. this [19, 20]. This is symmetrical w.r.t. all the images and usually gives the simplest equations, but in a frame that changes as Ω^* does. To find explicit critical motions, one must revert to camera-based coordinates which is sometimes delicate. The finite and Π_∞ cases must also be treated separately, *e.g.* $\Omega^* = \mathrm{diag}(c_1,c_2,c_3,c_4)$ with either c_3 or c_4 zero.

(ii) Work in the first camera frame, encoding Ω^* by its first image ω_1^* and supporting plane $(\boldsymbol{n}^\top,1)$. Subsequent images $\omega_i^* \simeq \boldsymbol{H}_i \, \omega_1^* \, \boldsymbol{H}_i^\top$ are given by the inter-image homographies $\boldsymbol{H}_i = \boldsymbol{R}_i + \boldsymbol{t}_i \boldsymbol{n}^\top$ where $(\boldsymbol{R}_i | -\boldsymbol{t}_i)$ is the ith camera pose. The output is in the first camera frame and remains well-defined even if the conic tends to infinity, but the algebra required is significantly heavier.

(iii) Parametrize Ω^* implicitly by two images ω_1^*, ω_2^* subject to the Kruppa constraints. In the 2 image case this approach is both relatively simple and rigorous — as above, two proper virtual dual image conics satisfy the Kruppa constraints iff they define a (pair of) corresponding 3D potential absolute conics — but it does not extend so easily to multiple images.

4 Calibrated Cameras

We start with fully calibrated perspective cameras:

Theorem 4.1 *Given projective structure and calibrated perspective cameras at $m \geq 3$ distinct finite camera centres, Euclidean structure can always be recovered uniquely. With $m = 2$ distinct centres there is always exactly a 2-fold ambiguity corresponding to a 'twisted pair'.*

Proof: The camera orientations are irrelevant because any false absolute conic has the same rotation invariant images as the true one. Assuming that $\boldsymbol{K} = \boldsymbol{I}$ does not change the critical motions. Calibrated cameras never admit false absolute conics on Π_∞, as the (known) visual cone of each camera intersects Π_∞ in a unique conic, which is the true AC. Given a finite false AC, work in a frame in which it is diagonal and supported on the $z = 0$ plane: $\Omega^* \equiv \mathrm{diag}(c_1, c_2, 0, c_4)$. Since the cameras are calibrated and their orientations are irrelevant, the conic projection in each camera becomes $(\boldsymbol{I}|-\boldsymbol{t})\,\Omega^*\,(\boldsymbol{I}|-\boldsymbol{t})^\top \simeq \boldsymbol{I}$. It is easy to show that the only solutions to this are $\Omega^* \simeq \mathrm{diag}(1,1,0,1/z^2)$ and $\boldsymbol{t}_\pm = (0,0,\pm z)^\top$ for some $z > 0$. Hence, ambiguity implies that there are at most two camera centres, and the false AC is a circle of imaginary radius iz, centred in the plane bisecting the two centres.

This two-fold ambiguity corresponds exactly to the well-known **twisted pair** duality [11, 10, 15], where one of the cameras is rotated by 180° around the axis joining their two centres. The improper self-inverse projective transformation

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & z \\ 0 & 0 & 1/z & 0 \end{pmatrix}$$

interchanges the true and false DACs $\mathbf{T}\Omega^*\mathbf{T}^\top \simeq \Omega^*_\infty$ and takes the projection matrices $\mathbf{P}_\pm = \boldsymbol{R}_\pm(\boldsymbol{I}|-\boldsymbol{t}_\pm)$ to $\mathbf{P}_-\mathbf{T}^{-1} = \mathbf{P}_-$ and $\mathbf{P}_+\mathbf{T}^{-1} = -\mathbf{P}_+\boldsymbol{U}$ where $\boldsymbol{U} = \mathrm{diag}(-1,-1,1,1)$ is a 180° twisted pair rotation about the z axis. The 'twist' $\mathbf{T}$ represents a very strong projective deformation which cuts the scene in half, moving the plane between the cameras to infinity. By considering twisted *vs.* non-twisted optical ray intersections, one can also show that it reverses the relative signs of the projective depths [21] of each correspondence, *e.g.* as recovered by the equation $\lambda_1 \boldsymbol{F}\boldsymbol{x}_1 = \lambda_2(\boldsymbol{e} \wedge \boldsymbol{x}_2)$. Moreover, *any* proper virtual Kruppa geometry (fig. 1) has such a 'twisted pair' projective involution symmetry, so *calibrated or not, two image Euclidean structures always occur in twisted pairs*. However the twist is a simple 180° rotation only for axisymmetric DIACs.

Theorem 4.2 *Given projective structure and $m \geq 3$ scaled orthographic cameras with distinct projective centres (*i.e. *viewing directions, with diametrically opposite ones identified), Euclidean structure can always be recovered uniquely. With only $m = 2$ distinct centres there is a one parameter family of possible structures corresponding to the bas relief ambiguity [11, 10, 15, 22].*

Proof: Choose coordinates in which camera 1 has orientation $\boldsymbol{R}_1 = \boldsymbol{I}$. Orthographic and affine cameras have Π_∞ as their optical planes, so Π_∞ is known and any potential AC must lie on it. Potential DACs have the form $\Omega^* = \begin{pmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 \end{pmatrix}$ for symmetric 3×3 $\boldsymbol{C}$. The orthographic calibration constraint is that $\boldsymbol{U}\boldsymbol{C}\boldsymbol{U}^\top \simeq \mathrm{diag}(1,1)$ where $\boldsymbol{U}$ is the first two rows of $\boldsymbol{R}$. In image 1 this gives $\boldsymbol{C}_{11} - \boldsymbol{C}_{22} = \boldsymbol{C}_{12} = 0$ and two analogous constraints in image 2. Representing $\boldsymbol{R}_2$ by a quaternion $\mathbf{q}$ and eliminating $\boldsymbol{C}_{11}$ between these constraints gives

$$(q_0^2+q_3^2)(q_1^2+q_2^2)\left((q_0q_1+q_2q_3)\boldsymbol{C}_{13} + (q_0q_2-q_1q_3)\boldsymbol{C}_{23}\right) = 0$$

This must hold for any motion satisfying the constraints. The first two terms correspond to optical axis rotations and 180° flips that leave the optical centre fixed, and are therefore excluded by the statement. Solving for $\boldsymbol{C}$ in terms of $\mathbf{q}$ using the final term gives a linear family of solutions $\boldsymbol{C} \simeq \alpha\boldsymbol{I} + \beta(\mathbf{o}_1\mathbf{o}_2^\top + \mathbf{o}_2\mathbf{o}_1^\top)$ where $\mathbf{o}_1 = (0,0,1)^\top$ and $\mathbf{o}_2 =$ (the third row of $\boldsymbol{R}$) are the optical centres, and (α,β) are arbitrary parameters. Given $\boldsymbol{I}$ and any false DAC $\boldsymbol{C} \not\simeq \boldsymbol{I}$, we can uniquely recover the family and its two camera centres (the three rank 2 members of the family each decompose into point pairs, but only one of these is real). Since each family encodes its centres, families with distinct centres never coincide. By linearity, they therefore intersect in at most one conic. All families intersect in the true DAC $\boldsymbol{C} = \boldsymbol{I}$, so no other intersection is possible. *I.e.* false structures are impossible for orthographic images from ≥ 3 distinct centres. That the one parameter ambiguity for two cameras corresponds to the *bas relief* 'flattening' is well known [11, 10, 15, 22].

Two image orthographic absolute conic geometry is easily understood in terms of the Kruppa constraints. These are well behaved as the cameras tend to infinity, and hence still define a one parameter family of dual quadrics. However as the cameras recede and their focal length increases, their DIACs become progressively flatter and this constrains the 3D family to be flatter too, until in the limit all members of the family become infinitely flat rank 3 disk quadrics squashed onto Π_∞.

5 Focal Lengths from 2 Images

For two cameras, projective geometry is encapsulated in the 7 d.o.f. fundamental matrix, and Euclidean geometry in the 5 d.o.f. essential matrix. Hence, from 2 projective images we might hope to estimate Euclidean structure plus two additional calibration parameters. Hartley

[6] gave a method for the case where the only unknown calibration parameters are the focal lengths of the two cameras. This was later elaborated by Newsam *et.al.* [16], and Zeller & Faugeras and Bougnoux [24, 2]. Hippisley-Cox & Porrill [9] give a related method for equal but unknown focal lengths and aspect ratios. All of these methods are Kruppa-based. We will give a unified presentation and derive the critical motions for the Hartley-Newsam-Bougnoux (unequal f) and Newsam (equal f) case.

Suppose that we can write all pairs of dual image conics satisfying the calibration constraints as a parametric family $(\omega_1^*(\lambda), \omega_2^*(\lambda))$. As they already obey the calibration constraints, pairs of nonsingular conics in this family represent possible 3D absolute conics iff they also satisfy the Kruppa constraints, $\boldsymbol{F}^\top \omega_2^*(\lambda)\boldsymbol{F} = \mu[\boldsymbol{e}]_\times \omega_1^*(\lambda)[\boldsymbol{e}]_\times^\top$ for some scalar μ. Solving these equations for λ, μ gives the possible image DIACs and hence 3D absolute conics. If $\omega_i^*(\lambda)$ are linear in their parameters λ, the system is bilinear in λ, μ. In particular, for zero skew and known principal point $\boldsymbol{p}_i$, $\omega_i^*(\lambda)$ is linear in f_i^2 and $(a_i f_i)^2$. For known a_i and unconstrained f_i, this gives fully linear equations in μf_1^2, μ and f_2^2:

$$\boldsymbol{F}^\top \left(f_2^2 \boldsymbol{D} + \boldsymbol{p}_2\boldsymbol{p}_2^\top \right) \boldsymbol{F} = [\boldsymbol{e}]_\times \left((\mu f_2^2)\boldsymbol{D} + \mu \boldsymbol{p}_1\boldsymbol{p}_1^\top \right) [\boldsymbol{e}]_\times^\top$$

where $\boldsymbol{D} \equiv \mathrm{diag}(1,1,0)$. Writing the 3×3 symmetric rank 2 matrices $\boldsymbol{F}^\top\boldsymbol{D}\boldsymbol{F}, \ldots, [\boldsymbol{e}]_\times \boldsymbol{p}_1\boldsymbol{p}_1^\top [\boldsymbol{e}]_\times^\top$ as 6 vectors gives a 6×4 rank 3 homogeneous linear system $\mathbf{M}_{6\times 4}\,(f_2^2, 1, \mu f_1^2, \mu)^\top = 0$. This can easily be solved for μ, f_1, f_2. There are multiple solutions for f_i — and hence ambiguous Euclidean structures — iff the coefficient matrix $\mathbf{M}_{6\times 4}$ has rank ≤ 2. We will study this case below. Newsam *et.al.* [16] use the SVD of $\boldsymbol{F}$ to project 3 independent rows out of this system. Bougnoux [2] uses properties of fundamental matrices to solve it in closed form:

$$f_2^2 = -\frac{(\boldsymbol{p}_2^\top \boldsymbol{F}\boldsymbol{D}[\boldsymbol{e}]_\times \boldsymbol{p}_1)\,(\boldsymbol{p}_2^\top \boldsymbol{F}\boldsymbol{p}_1)}{\boldsymbol{p}_1^\top \boldsymbol{F}^\top\boldsymbol{D}\boldsymbol{F}\boldsymbol{D}[\boldsymbol{e}]_\times \boldsymbol{p}_1}$$

If the focal lengths are known to be equal, $f_1 = f_2 = f$, the system takes the form $\mathbf{M}_{6\times 2}(\mu)\begin{pmatrix} f^2 \\ 1 \end{pmatrix} = 0$ where $\mathbf{M}_{6\times 2}(\mu)$ is linear in μ and generically has rank 2. This system has a nontrivial solution iff all of its 2×2 minors vanish — a set of quadratic constraints on μ. If the focal lengths really are equal, at most two of these quadratics are linearly independent and we can generically eliminate the μ^2 term between them, solve linearly for μ, substitute into $\mathbf{M}_{6\times 2}$ (which then has rank 1) and solve uniquely for f^2. This fails iff all of the quadratics are: (i) proportional — in which case the single quadratic gives exactly two possible solutions for μ and f; (ii) zero — in which case $\mathbf{M}_{6\times 2} = 0$ and any f is possible. We will return to these cases below. Finally (*c.f.* [9]), equal but unknown aspect ratios and focal lengths $a_1 = a_2 = a$, $f_1 = f_2 = f$, give a 6×3 rank 3 system $\mathbf{M}_{6\times 3}(\mu)\,(f^2, (a\,f)^2, 1)^\top = 0$, which has a solution iff the determinant of any of its nontrivial 3×3 minors vanishes — a single cubic in μ, giving at most 3 solutions for μ, f, a.

Now consider the critical motions of the above methods. Assume finite a, f and $\boldsymbol{t} \neq 0$.

Theorem 5.1 *For the known a, unequal f problem, the critical motions for the Hartley, Newsam and Bougnoux methods are all identical and intrinsic to any method for this problem. In fact, they are exactly the two evident singularities of Bougnoux' equations:* (i) $\boldsymbol{p}_2^\top \boldsymbol{F}\boldsymbol{p}_1 = 0$ *and* (ii) $\boldsymbol{p}_2^\top \boldsymbol{F}\boldsymbol{D}[\boldsymbol{e}]_\times \boldsymbol{p}_1 = 0$.

Case (i) occurs when the principal points are in epipolar correspondence, *i.e.* the optical axes intersect. (ii) occurs whenever the point $\boldsymbol{D}[\boldsymbol{e}]_\times \boldsymbol{p}_1$ on the line at infinity in the first camera lies on the epipolar line $\boldsymbol{F}^\top \boldsymbol{p}_2$ of the other principal point. This condition is actually symmetric between the images. If $\boldsymbol{p}_1 = \boldsymbol{p}_2 = (0,0,1)^\top$, ($ii$) occurs whenever $\boldsymbol{F}^\top \boldsymbol{p}_2$ contains the direction orthogonal to the epipolar line $[\boldsymbol{e}]_\times \boldsymbol{p}_1$, *i.e.* whenever the epipolar plane of optical axis $\boldsymbol{p}_1$ is *orthogonal* to that of axis $\boldsymbol{p}_2$ [16]. If either principal point coincides with an epipole, both (i) and (ii) apply and a second order singularity occurs.

Theorem 5.2 *For the known a equal f problem, there is a unique solution for f everywhere outside the critical variety of the unequal f method. On this variety there are generically exactly two solutions corresponding to the two roots of the single surviving quadratic in μ. Both solutions may be real, or one may be imaginary ($f^2 < 0$). There are more than two real solutions (in fact any f is possible) only on the following subvarieties of the corresponding-principal-point variety $\boldsymbol{p}_2^\top \boldsymbol{F}\boldsymbol{p}_1 = 0$, where $(\boldsymbol{R}(\mathbf{q}), \boldsymbol{t})$ is the relative pose of the second camera with quaternion $\mathbf{q}$:*
(i) $t_3q_2 - t_2q_3 + t_1q_0 = 0$ *and* $t_3q_1 - t_1q_3 - t_2q_0 = 0$
(ii) $t_1q_1 + t_2q_2 + t_3q_3 = 0$ *and* $t_2q_1 - t_1q_2 + t_3q_0 = 0$
(iii) $q_1 = 0$ *and* $q_2 = 0$
(iv) $q_3 = 0$ *and* $q_0 = 0$

Each of these subvarieties has codimension 2 in the space of all motions, and codimension 1 in the corresponding principal point variety. (iii) and (iv) correspond to parallel optical axes (axis rotations, and 180° flips about any axis in the optical plane, plus arbitrary translation). (ii) requires both planar motion $\mathbf{q}\cdot\boldsymbol{t} = 0$ and corresponding principal points. The intersection of these two varieties has two components: (a) arbitrary planar motions when the optical axes lie in the plane (*e.g.* a driving car with forwards-pointing camera), and (b) 'turntable rotations' about the intersection point of the two optical axes, when these do not lie in the plane. Subvariety (ii) corresponds to case (b). Case (a) has two solutions for f but is generically nonsingular.

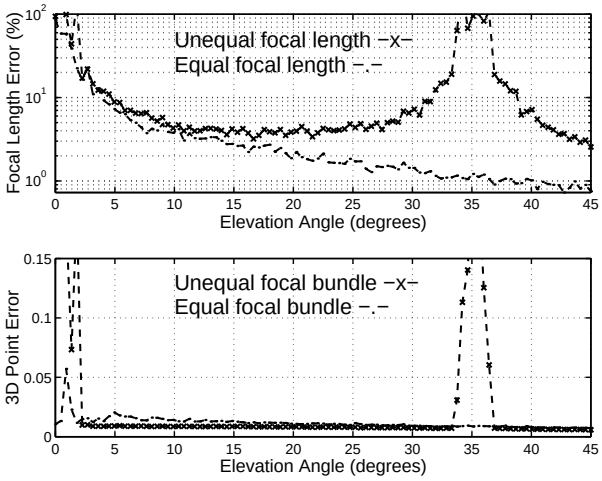

Figure 2: Relative errors in quasi-linear f and bundle-based 3D structure *vs.* camera elevation, for unequal and equal f methods.

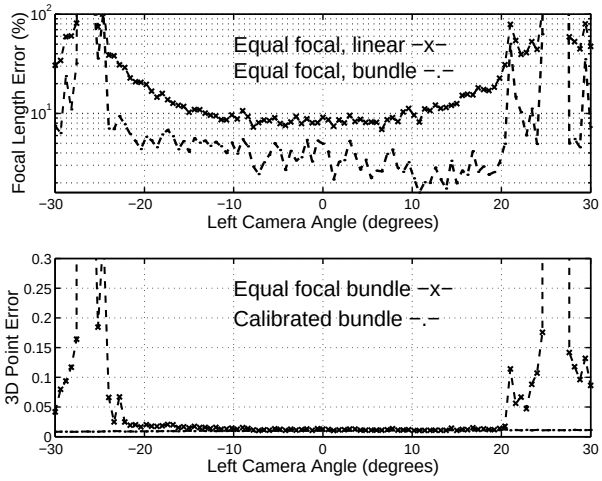

Figure 3: Errors in quasi-linear and bundle-based f, and 3D structure with unknown and known f, for equal f methods.

The above results are straightforward but fairly heavy to prove using the automated algebraic tools we are studying here. (Newsam *et.al.* [16] — a reference we were unaware of while completing this work — give a fairly simple SVD-based proof for their unequal f method, but an incomplete result for the equal f one). Since we were initially sceptical that the general Kruppa approach and Bougnoux' detailed manipulations [2] introduced no spurious ambiguities, we proved the results twice: once in a fundamental matrix / Kruppa constraint based parametrization, and once in an image conic / plane homography based one. In each case, given the parametrization we can more or less mechanically calculate and decompose the variety on which the constraints degenerate using MACAULAY 2. The calculations are 'routine', although the homography based ones are near the limits of the current system.

6 Experiments

We have performed some synthetic experiments to evaluate the effects of critical motions. We will focus on the question of how far from critical two cameras must be to get reasonable estimates of focal length and Euclidean 3D structure. The first experiment studies the unequal f case, the second the equal f one. For both experiments, two unit focal length perspective cameras view 25 points distributed uniformly within the unit sphere. Gaussian noise of 1 pixel standard deviation was added to the 512×512 images. For each pose, an optimal projective structure and fundamental matrix is estimated by projective bundle adjustment, the focal length(s) are estimated quasi-linearly as above, Euclidean bundle adjustment is applied to get Euclidean structure, and the resulting 3D error is calculated by Euclidean alignment. Means over 100 trials are shown. The Bougnoux and Newsam unequal f methods give essentially identical results: only the latter is plotted.

In the first experiment, cameras at $(-2,-2,0)$ and $(2,-2,0)$ focus on the origin. Their elevation angles are then varied, upwards for the left camera and downwards for the right one, so that their optical axes are skewed and no longer meet. Quasi-linear focal lengths and bundle adjusted Euclidean structures are estimated, both with and without the equal f constraint. Fig. 2 shows the resulting RMS errors as a function of elevation angle. At zero elevation, the optical axes intersect and the cameras are equidistant from this intersection, so both equal and unequal f methods are critical. This can be seen clearly in the graphs. The unequal f method also breaks down when the epipolar planes of the optical axes become orthogonal at around 35° elevation — the second component of the unequal f critical variety, but non-critical for the equal f method. For geometries more than about 5-10° from criticality, the unequal and equal f bundles both give results very similar to the optimal 3D structure obtained with *known* calibration.

In the second experiment, cameras at $(-1,-2,0)$ and $(1,-2,0)$ focus on the origin, then the left camera is rotated so that its optical axis sweeps the world plane $z = 0$. This is always critical for the unequal f method and the equal f one always gives two possible solutions. But in these trials, one is always tiny or imaginary and can safely be discarded. In fig. 3, the upper graph compares the quasi-linear equal f result with that obtained after optimal equal f bundle adjustment. The lower graph compares the

structures obtained with equal f and known-calibration bundle adjustments. At rotation angles of around $-27°$ the camera axes are parallel, and at around $+27°$ their intersection is equidistant from both cameras. These are intrinsic equal f degeneracies, clearly visible in the graphs. Moving about 5-10° from criticality suffices to ensure reasonably accurate focal lengths and Euclidean structure.

7 Conclusions

We have explicitly described the critical motions for a number of simple calibration constraints, ranging from unknown focal lengths to fully calibrated cameras. Numerical experiments studying the effects of near-critical configurations were also presented.

One of our aims was to see what could be achieved in vision with formal ideal-theoretic calculations. It is clear that although automated tools for this (MACAULAY 2, SINGULAR, COCOA) have progressed significantly in recent years, they can not yet replace geometric intuition. Even when a calculation terminates — and the 'ceiling' for this is still frustratingly low — the geometric interpretation of the results remains a difficult 'inverse problem'. However when it comes to rigorously proving formal properties of systems of equations we have found these tools a powerful computational aid and a good deal more reliable than 'proof by intuition'. Hence, we feel that these methods do have a place in vision, particularly for studying singularities of simple algebraic (auto)calibration and camera pose methods.

We are currently investigating critical motions where even less is known about the calibration, *e.g.* cameras having zero skew and unit aspect ratio, but with the other parameters unknown and possibly varying.

References

[1] K. Åström and A. Heyden. Euclidean reconstruction from constant intrinsic parameters. In *Int. Conf. Pattern Recognition*, pages 339–343, 1996.

[2] S. Bougnoux. From projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Int. Conf. Computer Vision*, 1998.

[3] O. Faugeras. *Three-Dimensional Computer Vision.* MIT Press, Cambridge, Mass, 1993.

[4] O. Faugeras, Q.-T. Luong, and S. Maybank. Camera self-calibration: Theory and experiments. In *European Conf. Computer Vision*, pages 321–334. Springer-Verlag, 1992.

[5] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *European Conf. Computer Vision*, pages 563–578. Springer-Verlag, 1992.

[6] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *European Conf. Computer Vision*, pages 579–87. Springer-Verlag, 1992.

[7] R. Hartley. Euclidean reconstruction from multiple views. In 2nd *Europe-U.S. Workshop on Invariance*, pages 237–56, Ponta Delgada, Azores, October 1993.

[8] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Int. Conf. Computer Vision & Pattern Recognition*, pages 761–4, 1992.

[9] S. Hippisley-Cox and J. Porrill. Auto-calibration, Kruppa's equations and the intrinsic parameters of a camera. In *British Machine Vision Conference*, pages 771–7, 1994.

[10] B. K. P. Horn. Motion fields are hardly ever ambiguous. *Int. J. Computer Vision*, 1, 1987.

[11] H. C. Longuet-Higgins. Multiple interpretations of a pair of images of a surface. *Proc. Roy. Soc. London*, A:418, 1988.

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

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

[14] J. L. Mundy and A. Zisserman, editors. *Geometric invariance in Computer Vision.* MIT Press, Cambridge Ma, USA, 1992.

[15] S. Negahdaripour. Multiple interpretations of the shape and motion of objects from two perspective images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(11):1025–1039, 1990.

[16] G. Newsam, D.Q. Huynh, M. Brooks, and H.-P. Pan. Recovering unknown focal lengths in self-calibration: An essentially linear algorithm and degenerate configurations. In *Int. Arch. Photogrammetry & Remote Sensing*, volume XXXI-B3, pages 575–80, Vienna, 1996.

[17] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Int. Conf. Computer Vision*, 1998.

[18] M. Pollefeys, L. Van Gool, and M. Oosterlinck. The modulus constraint: A new constraint for self-calibration. In *Int. Conf. Pattern Recognition*, pages 349–353, 1996.

[19] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. Conf. Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1100–1105, 1997.

[20] P. Sturm. *Vision 3D Non Calibrée: Contributions à la Reconstruction Projective et Étude des Mouvements Critiques pour l'Auto-Calibrage.* PhD thesis, Institut National Polytechnique de Grenoble, 1997.

[21] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *European Conf. Computer Vision*, pages 709–720, 1996.

[22] R. Szeliski and S. B. Kang. Shape ambiguities in structure from motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5), May 1997.

[23] B. Triggs. Autocalibration and the absolute quadric. In *Int. Conf. Computer Vision & Pattern Recognition*, 1997.

[24] C. Zeller and O. Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. Technical Report 2793, INRIA, Sophia-Antipolis, France, 1996.