A Nine-point Algorithm for Estimating Para-Catadioptric Fundamental Matrices

Christopher Geyer[†] and Henrik Stewenius[‡] cgeyer@ri.cmu.edu, stewe@vis.uky.edu † Robotics Institute, Carnegie Mellon University ‡ Center for Visualization and Virtual Environments, University of Kentucky

Abstract

We present a minimal-point algorithm for finding fundamental matrices for catadioptric cameras of the parabolic type. Central catadioptric cameras—an optical combination of a mirror and a lens that yields an imaging device equivalent within hemispheres to perspective cameras—have found wide application in robotics, tele-immersion and providing enhanced situational awareness for remote operation. We use an uncalibrated structure-from-motion framework developed for these cameras to consider the problem of estimating the fundamental matrix for such cameras. We present a solution that can compute the para-catadioptirc fundamental matrix with nine point correspondences, the smallest number possible. We compare this algorithm to alternatives and show some results of using the algorithm in conjunction with random sample consensus (RANSAC).

1. Introduction

In this paper we present an algorithm that improves the state-of-the-art in structure-from-motion (SfM) for omnidirectional cameras, specifically concentrating on a type of catadioptric¹ camera that uses a parabolic mirror and an orthographic lens. We present a minimal-point solver of fundamental matrices that when used in a RANSAC framework improves reconstruction for uncalibrated para-catadioptric cameras. Because of their wide field of view, both mirrorand fisheye-based omnidirectional cameras have found a number of applications in robotics, where an omnidirectional camera can provide the situational awareness that a tele-operator or an autonomous navigation system needs to determine that there are no threats to a vehicle. Omnidirectional cameras have even found uses in realtors' offices, where they are used to give virtual tours to prospective home buyers.

Mirror systems offer a number of advantages: an inexpensive way to obtain a $180^{\circ} \times 360^{\circ}$, i.e., hemispherical, field-of-view (FOV), making them ideal for robotic appli-

cations. For example, some catadioptric cameras provide a 210° vertical FOV, or 105° from the optical axis, that can provide a robot with 15° of coverage below the horizon, whereas most fisheye lenses yield at most a 180° vertical FOV. An important class of catadioptric systems are the *central catadioptric* cameras [1]. Central catadioptric systems have a *single effective viewpoint*, meaning that within a hemisphere they can be exactly warped into a perspective image taken from the same viewpoint without suffering from parallax. The para-catadioptric combination, or parabolic mirror plus orthographic lens, is one such case and para-catadioptric cameras are commercially available. In this paper we focus on structure-from-motion for the para-catadioptric case.

Structure-from-motion for omnidirectional cameras has been an active area of research, starting with T. Svoboda et al. [19] who first determined the form of the epipolar geometry for central catadioptric cameras. S.B. Kang [11], and at the same time Chang and Hebert [3], first presented discrete SfM estimators which apply traditional perspective methods for SfM to calibrated catadioptric images. Later, Geyer and Daniilidis [7] solved the uncalibrated structurefrom-motion problem for para-catadioptric systems by a non-linear embedding image coordinates assuming both image center and focal length are unknown. Simultaneously, Fitzgibbon [6] considered self-calibration of a radial distortion parameter from two views using a non-linear embedding followed by a quadratic eigenvalue solver. Micusik and Pajdla [13] extended Fitzgibbon's approach, instead solving a polynomial eigenvalue problem, to obtain results for more highly distorted fisheye lenses. More recently Micusik et al. have continued this line of research in [14]. In [14], the authors provide a solution to uncalibrated SfM assuming that the image center is known. In this paper we provide a minimal point solver for the uncalibrated case including unknown image center by applying tools from algebraic geometry.

Recently researchers have applied methods from algebraic geometry to solve long-standing practical problems in computer vision. Nistér provided an efficient minimal 5-

¹*Cata-di-op-tric*: mirror (cata) plus lens (dioptric).

point solver for essential matrices [16] that inspired some of the work here. These techniques are generalized by Stewénius, et al. in [17, 18] using methods such as Gröbner trace algorithms [20]. These solvers work by transforming the problem to an eigenproblem by first computing a Gröbner basis and then finding a so-called action matrix for multiplication by a monomial on the quotient ring defined by the Gröbner basis.

Our goal is to efficiently perform robust reconstructions with para-catadioptric systems with little or no user supervision, and this process must by necessity be robust to outliers. Random sample consensus (RANSAC) [5] is typically the method of choice for finding the inlier correspondences in many structure-from-motion problems. See Hartley and Zisserman [10] for a review of RANSAC from a SfM perspective. RANSAC is simple: randomly choose K of your data points, generate a hypothesis from the K datapoints, see how many inliers that model has, if this exceeds the number of inliers for any previous model keep it, and repeat M times. It does not have high storage requirements; it just needs time. If N' of the original N datapoints are inliers, then to maintain a minimum probability p of finding the right model requires that M, the number of RANSAC iterations, be greater than $\log(1-p)/\log[1-(N'/N)^K]$. Therefore, in a RANSAC implementation it is generally desirable to find a hypothesis generator for which K is as small as possible. We provide a 9-point solver for generating a hypothesis for a para-catadioptric fundamental matrix from 9 point correspondences. Taking into account the average number of real roots generated by the solver, we have to test about 33% fewer hypotheses to achieve the same level of reliability from the alternative-a 15-point solverin datasets, of which 50% are outliers.

Why not use the mirror's silhouette to calibrate at least the image center? Most of the time the mirror's silhouette can be very useful, however, it often occurs that the edge of the mirror silhouette is blurred or has lens flares both of which hamper its detection by automatic means. Furthermore, optical elements on robotic platforms often move during operation. It is both useful and interesting to develop an approach that is robust to changes in calibration, including the image center. We hope that some day an optical device might be constructed from lenses or clever catadioptric designs that achieve the same projection model while minimizing obstruction by the camera's reflection, or which do not have a mirror silhouette.

The novelty in this paper is the development of a minimal 9-point solver for para-catadioptric fundamental matrices. In addition we characterize for the first time the constraint on para-catadioptric fundamental matrices in a set of algebraic equations. We also provide a novel reconstruction technique for para-catadioptric cameras that is able to automatically calibrate focal length and image center. In the last section we present experiments in simulation, which show that in cases the 9-point algorithm provides motion estimates up to three times better than the next alternative, as well as show reconstructions obtained from images taken with a para-catadioptric camera.

2. Background

In the perspective structure-from-motion problem, we use homogeneous coordinates to linearize two and multi-view structure-from-motion problems. Geyer and Daniilidis [8] showed how to achieve a similar result for para-catadioptric projections by embedding the uncalibrated image plane on a sphere. One applies the inverse of stereographic projection, taking the image plane to the unit sphere via projection from the sphere's north pole. The two-view problem then becomes linear: there exists a 4×4 matrix F, called the *para-catadioptric fundamental matrix*, that is linear in the known point correspondences. It encodes the epipolar geometry and can be decomposed into the motion between the two cameras.

The difficulty with the theory provided in [8] and earlier in [7], however, was that it lacked the capability to get a good initial estimate on the space of fundamental matrices. The manifold of para-catadioptric fundamental matrices is not invariant to 4×4 orthogonal transformations, and so there is no corresponding projection theorem like there is for the manifold of perspective essential matrices, or perspective fundamental matrices. Furthermore, until now there had been no way to construct a para-catadioptric fundamental matrix from a minimum number of point correspondences. We rectify this situation in section 4, but first we give the necessary background material.

We start with the para-catadioptric projection. We assume that a convex parabolic mirror is placed in front of an orthographically projecting camera so that the axes of the two optics are parallel. The parabolic projection of a point x = (x, y, z), the derivation of which we refer the reader to [1], is given by:

$$q_{\xi}(\mathbf{x}) = \frac{f}{z - \sqrt{x^2 + y^2 + z^2}} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}$$
(1)

where (c_x, c_y) is the image center, f is the focal length, and $\xi = (c_x, c_y, f)$. We define the image center to be the intersection of the mirror's optical axis with the imaging sensor, and f is the mirror's focal length as measured by the orthographic camera, i.e., measured in pixels. We assume that the aspect ratio is 1.

There are two special cases. First, we do not allow x to be 0. Second, there is a single point at infinity corresponding to the optical axis above the focal point of the mirror. We let $q_{\xi}(0,0,z) = p_{\infty}$ for all positive z, and let $\mathbb{R}^{2^*} = \mathbb{R}^2 \cup \{p_{\infty}\}$ be q's range.

If the camera is calibrated, that is, we know ξ , we can invert formula (1) up to scale and end up with a calibrated unit vector, which we can substitute in the perspective epipolar constraint $p^T Eq = 0$. In general the camera may be uncalibrated. Fortunately, though, the para-catadioptric projection admits an exact linearization to account for unknown calibration parameters. We embed the image coordinates in \mathbb{P}^3 using the inverse of stereographic projection, given by the operator $\sim :\mathbb{R}^{2^*} \to \mathbb{P}^3$, which we define as follows:

$$\widetilde{(u,v)} \quad \equiv \quad \begin{bmatrix} 2u & 2v & -1+u^2+v^2 & 1+u^2+v^2 \end{bmatrix}^{\mathrm{T}}$$

and where $\tilde{p}_{\infty} = (0, 0, 1, 1)$. Each lifting lies on the unit sphere in \mathbb{P}^3 , and so satisfies $\tilde{p}^T Q \tilde{p} = 0$, where Q = diag(1, 1, 1, -1).

The power of the embedding is the following law that allows a commutation of a translation and scale in the plane with a linear transformation of the projective unit sphere. For each $\xi = (c_x, c_y, f)$ there exists a 4×4 matrix K_{ξ} such that for all x:

$$\widetilde{q_{\xi}(\mathbf{x})} = \lambda \,\mathsf{K}_{\xi} \, \widetilde{q_{(0,0,1)}(\mathbf{x})}, \qquad (2)$$

for some $\lambda(\mathbf{x})$, and where $q_{(0,0,1)}(\mathbf{x})$ is the canonical paracatadioptric projection analogous to the canonical perspective projection (x/z, y/z). The matrix K_{ξ} is given below:

$$\mathsf{K}_{\xi} = \begin{bmatrix} 1 & 0 & c_x & -c_x \\ 0 & 1 & c_y & -c_y \\ -\frac{c_x}{f} & -\frac{c_y}{f} & -\frac{c_x^2 + c_y^2 - f^2 - 1}{2f} & \frac{c_x^2 + c_y^2 - f^2 + 1}{2f} \\ -\frac{c_x}{f} & -\frac{c_y}{f} & -\frac{c_x^2 + c_y^2 + f^2 - 1}{2f} & \frac{c_x^2 + c_y^2 + f^2 + 1}{2f} \end{bmatrix}$$

Each K_{ξ} is an element of the larger *Lorentz group*, defined as:

$$O(3,1) = \{A : A^T Q A = Q\}$$

Each Lorentz transformation A preserves the unit sphere set-wise: if x satisfies $x^TQx = 0$, then because $A \in O(3,1)$, $x^TA^TQAx = 0$. The Lorentz group is a six-dimensional Lie group having properties similar to that of the group of rotations, O(3), and has a Rodrigues formula with which we can parameterize O(3,1) by its tangent space.

The final tool is a substitution law. The substitution law allows us to take any constraint for perspective projections expressed in homogeneous coordinates, and turn it into a constraint on liftings of para-catadioptric projections. For every para-catadioptric projection $p = q_{\xi}(x)$ there exists a λ such that:

$$\lambda \mathbf{x} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathsf{K}_{\boldsymbol{\xi}}^{-1} \widetilde{\mathbf{p}} \,. \tag{3}$$

Since this says that the right hand side is parallel to the direction of x, of which p is the projection, we can treat the entire right-hand side of the equation as the homogeneous coordinates of the perspective projection of x. If we had tried to invert q_{ξ} directly, while not knowing ξ , we would have obtain a non-linear constraint in p.

Now for the climax. We can write a bilinear epipolar constraint in the lifted coordinates. For perspective views, the essential matrix $E = \widehat{t}R$ gives a bilinear epipolar constraint for a camera pair separated by a relative motion (R, t), where $R \in O(3)$, $t \in \mathbb{R}^3$ with $t \neq 0$, and $\widehat{a}b = a \times b$ for all a and b. If we let:

$$p = q_{\xi_1}(x)$$
 and $q = q_{\xi_2}(Rx+t)$,

then using equation (3) we obtain the *para-catadioptric* epipolar constraint:

$$\widetilde{\mathbf{q}}^{\mathrm{T}} \underbrace{\mathsf{K}_{\xi_{2}}^{-T} \mathsf{P}^{\mathrm{T}} \mathsf{E} \mathsf{P} \, \mathsf{K}_{\xi_{1}}^{-1}}_{para-catadioptric} \widetilde{\mathbf{p}} = 0 \tag{4}$$

where $P = \begin{bmatrix} I & 0 \end{bmatrix}$ is the canonical perspective projection from \mathbb{P}^3 to \mathbb{P}^2 .

Like the perspective fundamental and essential matrices, the 4×4 para-catadioptric fundamental matrix is rank 2. In \mathbb{P}^3 the span of the two-dimensional nullspace of F is a line through the unit sphere where the intersections are the liftings of the two epipoles for the first view. Similarly F^T yield the epipoles for the second view.

3. The Manifold of Para-catadioptric Fundamental Matrices

Our goal is to take a minimum number of point pairs (p_i, q_i) , which we will show is nine, and from them generate a hypothesis—a fundamental matrix F for which eqn. (4) is satisfied by each point pair. Nister [16] solved the similar problem of finding an essential matrix from five points by finding the roots of a set of algebraic equations, which are derived from the cubic constraints on \mathcal{E} , the manifold of essential matrices. The space of para-catadioptric fundamental matrices is defined as follows:

$$\mathcal{F} \equiv \left\{\mathsf{K}_{\xi_2}^{-T}\mathsf{P}^{\mathsf{T}}\mathsf{E}\,\mathsf{P}\,\mathsf{K}_{\xi_1}^{-1}:\mathsf{E}\in\mathcal{E}, \xi_i = (c_x^{(i)}, c_y^{(i)}, f^{(i)})\right\}.$$

Recall that $\mathcal{E} = \{R\hat{t} : R \in O(3), t \in \mathbb{R}^3, t \neq 0\}$. We derive the analogous algebraic constraint for this space.

The defining characteristic of the set of para-catadioptric fundamental matrices is similar to that for the set of essential matrices. A matrix E is an essential matrix iff $E = U^T \operatorname{diag}(1, 1, 0)V$ for some $U, V \in O(3)$. Even though it is for an uncalibrated model, the constraint on \mathcal{F} is similar. Geyer and Daniilidis [9] showed that $F \in \mathcal{F}$ iff F can be written as

$$\mathsf{F} = \mathsf{U}^{\mathrm{T}} \operatorname{diag}(1, 1, 0, 0) \mathsf{V},$$

where U, V \in O(3, 1). This constraint is equivalent to a constraint on the roots of a characteristic equation, which in this case is $|FQF^{T}Q - \lambda I| = 0$. We give here an equivalent algebraic constraint whose proof we omit because of space constraints.

Theorem 1. A real non-zero 4×4 matrix F is a catadioptric fundamental matrix (can be decomposed into the form of eqn. (4)) iff rank F = 2,

$$\mathsf{F} \mathsf{Q} \mathsf{F}^{\mathrm{T}} \mathsf{Q} \mathsf{F} = \frac{1}{2} \operatorname{tr} \left(\mathsf{F} \mathsf{Q} \mathsf{F}^{\mathrm{T}} \mathsf{Q} \right) \mathsf{F}, \qquad (5)$$

and the eigenvalues of both FQF^{T} and $F^{T}QF$ are positive.

This theorem's three conditions yield sixteen cubic equations from the sixteen 3×3 minors that must be zero to constrain F to be rank 2, and sixteen cubic equations from eqn. (5). The eigenvalue conditions yield two inequalities. In the next section we derive a solver based on the 32 cubic equations. The solver ignores the inequality constraints, other than to check each root afterwards. Before continuing, we determine the number of points needed for the solver. We cite the following theorem [9] which tells us that \mathcal{F} is a 9-dimensional manifold, implying that nine points are sufficient to determine F (up to a discrete number of homogeneous roots).

Theorem 2. [9] The space of para-catadioptric fundamental matrices, \mathcal{F} , is isomorphic to a quotient of two groups:

$$\mathcal{F} \simeq \left. O(3,1) \times O(3,1) \right. / \mathcal{H}_{\mathsf{F}} \,,$$

where

$$\mathcal{H}_{\mathsf{F}} = \{ (\mathsf{A}, \mathsf{B}) \in \mathcal{O}(3, 1) \times \mathcal{O}(3, 1) : \mathsf{A}^{\mathrm{T}}\mathsf{F}\mathsf{B} = \mathsf{F} \},\$$

for any $F \in \mathcal{F}$. Furthermore, \mathcal{F} is a manifold of dimension:

$$\dim \mathcal{F} = \dim O(3,1) \times O(3,1) - \dim \mathcal{H}_{\mathsf{F}} = 9$$

This theorem uses the observation that the decomposition (5) implies that \mathcal{F} is in one-to-one correspondence with a partition of $O(3, 1) \times O(3, 1)$. Thus, the dimension of \mathcal{F} is the dimension of $O(3, 1) \times O(3, 1)$ minus the dimension of any one partition, which is the redundancy. The redundancy is characterized by the dimension of the *stabilizer* \mathcal{H}_F of any one F. We can choose any F, so we choose diag(1, 1, 0, 0)and calculate that dim $\mathcal{H}_F = 3$.

In summary, 9 points constrain are sufficient to solve for a para-cadiotpric fundamental matrix, and Theorem 1 gives 32 cubic equations that we can use as the constraint with which to solve for F.

4. A 9-point Algorithm

The 9-point algorithm takes nine point correspondences of the form $p_i = (u_{1,i}, v_{1,i})$ and $q_i = (u_{2,i}, v_{2,i})$, and gives a para-catadioptric fundamental matrix that defines an epipolar geometry that all nine points satisfy exactly. The 9point algorithm has three major steps: (1) determine a basis for the solution from the nullspace of a coefficient matrix; (2) extract the coefficients of 32 polynomials obtained by substituting the unknown linear combination into the fundamental matrix constraints; and (3) solve the system by constructing a Gröbner basis.

We start by taking equation 4 and writing the following linear constraint:

$$\underbrace{((u_{1,i}, v_{1,i}) \otimes (u_{2,i}, v_{2,i}))}_{a_i} \operatorname{vec} \mathsf{F} = 0, \tag{6}$$

where \otimes is the Kronecker product so that a_i is a row-vector containing all the information from observation pair *i*, and vec F is F reshaped to a column matrix. We stack the vectors a_i for nine observations into a 9 × 16 matrix A. We compute the 7 vectors vec F₁ through vec F₇ that form an orthonormal basis for the nullspace of A, i.e., A vec F_i = 0 and vec F_i^T vec F_j = $\delta_{i,j}$ for all *i* and *j*. The nullspace vectors correspond directly to seven 4 × 4 matrices F₁ to F₇ and the solution F must be of the form

$$\mathsf{F} = x_0 \mathsf{F}_0 + \sum_{i=1}^6 x_i \mathsf{F}_i \tag{7}$$

for some scalars x_0 through x_6 . Since the nine equations generated by (4) are homogeneous in F, we set $x_0 = 1$.

To solve for F, we need to substitute the expression for F given in (7) into the algebraic equations characterizing \mathcal{F} : equation (5) and the rank 2 constraint. The substitution yields a system of 32 degree three polynomial equations in x_1 through x_6 . The problem will be how to solve for these six unknown variables.

Substituting F from (7) into equation (5) yields 16 of the 32 third order polynomial equations. That F is rank 2 provides the remaining 16 equations. In particular, if F is rank 2 then all 3×3 sub-matrices of F have zero determinants. There are 16 such sub-matrices and each determinant is a degree three polynomial. In general the coefficients of these 32 polynomials are linearly independent.

Our next step is to compute a Gröbner basis for the ideal defined by this set of equations. We will use the graded reverse lexicographic order on the monomials. To make it easier to describe this process we will use the following

• A set of polynomial equations is always represented on the form MX = 0 where M is a matrix of scalars and X is a vector of monomials ordered in Graded Reverse Lexicographic order. We will let M stand for the equations MX = 0.

- Gauss-Jordan elimination with a prerecorded sequence of pivot columns is denoted GJ. Effectively we are using a Gröbner Trace [20].
- To make notation easier we permit ourselves to write xM to mean "Take the polynomial equations represented by M and multiply these equations by x and represent the result as a scalar matrix."
- To select equations from a set of equations we select rows of a matrix. We let M(a : b) mean rows a to b of M.

The start system is represented by a 32×84 matrix M₀ which is of full rank. A Gröbner basis can be found through the following four steps:

$$M_{1} = GJ(M_{0}), \qquad (8)$$

$$M_{2} = GJ\begin{pmatrix} x_{1}M_{1}(12:32) \\ x_{2}M_{1}(12:32) \\ x_{3}M_{1}(12:32) \\ x_{4}M_{1}(12:32) \\ x_{5}M_{1}(12:32) \\ \end{bmatrix}, \qquad (9)$$

$$M_{3} = GJ \begin{pmatrix} x_{4}M_{2}(90:98) \\ x_{5}M_{2}(90:98) \\ x_{6}M_{2}(90:98) \\ x_{6}^{2}M_{1}(13:31) \end{pmatrix},$$
(10)

$$\mathsf{M}_4 = \qquad \mathrm{GJ}\left(\begin{bmatrix} \mathsf{M}_1 \\ \mathsf{M}_2 \\ \mathsf{M}_3 \end{bmatrix} \right). \tag{11}$$

We have now computed a Gröbner basis. There are several ways to go from a Gröbner basis to the solutions, we choose to first compute the 64×64 action matrix A_{x_1} for multiplication by x_1 in the quotient ring defined by M_4 . The action matrix [4] is a generalization of the compainion matrix for univariate polynomials to multivariate polynomials and describe the effect of multiplication with a polynomial modulo a Gröbner basis. In practice the action matrix is computed by copying certain partial rows from M_4 . The left eigenvectors of A_{x_1} then encode the 64 solutions for (x_1, \ldots, x_6) and we can compute the solutions for F by inserting into Equation 7.

These steps might seem complicated but we will on acceptance of the paper provide Matlab source-code online. The solver runs in 0.062s on a 2.4GHz Athlon CPU. Most of the time is spent on eliminations, mainly on the last matrix.

5. Calibration and the 15-Point Algorithm

Para-catadioptric cameras, and in fact many omnidirectional cameras, are very easy to calibrate. Almost every construct—conics, images of lines, radial distortion reveal information about the location of the image center, and depending on the camera type, the focal length or scale parameters. In the para-catadioptric camera, the intrinsic parameters can be very easily recovered from a fundamental matrix involving the same camera taken at two locations, i.e., the intrinsic parameters are the same and $K_{\xi_1} = K_{\xi_2}$. The intrinsic parameters can be recovered almost directly from the nullspaces of F. This gives an alternative method for estimating fundamental matrices which we call the 15point algorithm.

Since we only encode image center and focal length, $\xi = (c_x, c_y, f)$ is in one-to-one correspondence with the homogeneous vector $K_{\xi}O$, where $O = (0, 0, 0, 1)^T$. We call the resulting transformation of the origin $\tilde{\xi}$:

$$\widetilde{\xi} = \begin{bmatrix} 2fc_x & 2fc_y & -1 + c_x^2 + c_y^2 + f^2 & 1 + c_x^2 + c_y^2 + f^2 \end{bmatrix}^T.$$

As it turns out, one can show that this point represents an imaginary circle representing the image of the absolute conic [8]. What is important from the point of view of calibration, is that $PK_{\xi}^{-1}\tilde{\xi} = PO = 0$. Consequently $F\tilde{\xi}_1 = 0$ and $F^T\tilde{\xi}_2 = 0$. Furthermore, if $\xi_1 = \xi_2$, then

$$\widetilde{\xi} = \mathcal{N} \begin{bmatrix} \mathbf{F} \\ \mathbf{F}^T \end{bmatrix} \,,$$

where \mathcal{N} is the nullspace operator. This is true in the absence of noise and as long as $\mathcal{N}(F) \neq \mathcal{N}(F^T)$ (i.e., as long as the rotation axis is not parallel to the translation axis).

This allows us to combine a 15-point linear algorithm with the calibration invariant above to compare with the 9point algorithm. Like in the 8-point algorithm for fundamental matrices, we construct the coefficient matrix from equations (6), and determine the 1D nullspace that gives us an estimate vec F_0 . In general $F_0 \notin \mathcal{F}$ and there is no projection mechanism by which we can average singular values. Instead we choose any point on $\mathcal{N}(\mathsf{F})$ within the sphere (both F_0 's nullspaces must intersect the sphere otherwise this method fails) and let this be $\tilde{\xi}_1$. Similarly we choose a $\tilde{\xi}_2$ inside the sphere and on $\mathcal{N}(\mathsf{F}^T)$. We then construct an estimate of an essential matrix as follows:

$$\mathsf{E}_0 = \mathsf{P}\mathsf{K}_{\xi_2}^{-\mathrm{T}}\mathsf{F}_0\mathsf{K}_{\xi_1}^{-1}\mathsf{P}^{\mathrm{T}}.$$

where again $\mathsf{P} = \begin{bmatrix} I & 0 \end{bmatrix}$. If F_0 were a para-catadioptric fundamental matrix, then E_0 would be an essential matrix. In general it is not, so we project $\mathsf{E}_0 = \mathsf{U}^{\rm T} \operatorname{diag}(\sigma_1, \sigma_2, \sigma_3)\mathsf{V}$ to $\mathsf{E}_1 = \mathsf{U}^{\rm T} \operatorname{diag}(1, 1, 0)\mathsf{V}$ on the essential manifold. Then,

$$\mathsf{F}_0 = \mathsf{K}_{\xi_2}^{\mathrm{T}} \mathsf{P}^{\mathrm{T}} \mathsf{E}_1 \mathsf{P} \mathsf{K}_{\xi_1} \,.$$

is the estimate obtained from the 15-point algorithm.

Figure 2 shows an example of the nullspaces of fundamental matrices recovered from a sequence of images. The nullspaces nearly intersect a single point in space. This intersection point is $\tilde{\xi}$, which we can recover from the nullspaces via usage of an SVD operation.

6. Optimization and Reconstructions

The purpose of the 9-point algorithm is to obtain any initial estimate at all. We will see in the next section that the 15-point algorithm gives estimates which are poor initializers for motion estimation. Furthermore, when we do not know what data points are inliers, the 9-point algorithm allows us to efficiently apply RANSAC to find the inliers. In either case, we must apply a non-linear optimization method to improve motion and structure estimates. We use the following normalized cost function suggested in [9]:

$$c(\mathsf{F}) = \sum_{i=1}^{n} \frac{\left(\widetilde{\mathsf{q}}_{i}^{\mathrm{T}} \mathsf{F} \widetilde{\mathsf{p}}_{i}\right)^{2}}{\left\|\frac{\partial}{\partial(\mathsf{p}_{i},\mathsf{q}_{i})}\widetilde{\mathsf{q}}_{i}^{\mathrm{T}} \mathsf{F} \widetilde{\mathsf{p}}_{i}\right\|^{2}}$$
(12)

This cost function is an analog to the normalized epipolar cost for perspective cameras.

For reconstruction we used a combination of: (1) 9-point and 15-point algorithms implemented in C and Matlab, and combined with a vanilla RANSAC implementation; (2) the affine-invariant Harris detection and descriptor implementations provided by Mikolajczyk et al. [15]; (3) and the Sparse Bundle Adjustment (SBA) package by Lourakis and Argyros [12]; and (4) a custom SBA driver written in Mathematica.

To generate the reconstruction in the next section, we use the following steps:

- For every pair of frames, apply the 9-point-based RANSAC algorithm to a list of best matched affineinvariant Harris features as measured by their SIFT descriptors.
- 2. Improve fundamental matrix estimates by minimizing (12) using Levenberg-Marquardt, or a robust version of Levenberg-Marquardt using the kernel $\sqrt{\varepsilon^2 + x}$ where x is a summand in (12).
- 3. Calibrate the camera by intersecting the nullspaces of single fundamental matrices to obtain points, and then apply mean-shift to isolate the peak of the distribution; or, apply RANSAC to find nullspaces close to coincident to a single point.
- 4. Build a model by using a pair of frames to start, and then building on to the model a frame at a time using a linear 6-point algorithm for registering a 3D point set to image points in a calibrated camera. Use SBA to refine estimates.

7. Experiments

In our implementation the 9-point algorithm performs at a tenth of the speed of the 15-point algorithm. These times may be improved with greater attention to optimization in the implementation; but we are hampered by the number of roots generated, for example, see the histogram of the number of real roots in Figure 3. A benefit of the 9-point algorithm that we see in simulation results, is its ability to provide an estimate that is on average closer to a basin of attraction closer to the true value.

Figure 1 shows images and a reconstruction from the images. Here we used twelve images for which the calibration was initially unknown. A video of this model will be a part of the supplemental material.

In Figures 4 and 5 we show the results of simulations in which we apply RANSAC followed by non-linear optimization to fit a fundamental matrix to noisy point correspondences. We tested four variants: Levenberg-Marquardt minimization initialized by RANSAC based on either the 9-point algorithm or the 15-point algorithm; and a robust version of the Levenberg-Marquardt algorithm mentioned in section 6, again either initialized by the 9-point or the 15-point algorithms.

The errors as a function of pixel noise are shown in Figure 5. The image size was 1000×1000 and the camera had a 210° field-of-view. The noise was varied between 1 and 16 pixels stepping in factors of 2. The intrinsic parameters do not show large variation across algorithms, however the motion estimates do. Across the board the two 9-point variants outperform the 15-point variants, in some cases almost by a factor of 4. As noise increases the marginal benefit decreases.

Because of the high-dimension of the optimization space, fitting the epipolar geometry can become illconditioned when the baseline is small. This will affect calibration to some degree, and so as to try guage the effects of these factors we plot in Figure 4 a scatter-plot of the parallax in rotation corrected images vs. calibration error. Variance in image center decreases with increasing median parallax.

8. Conclusion

We have demonstrated a minimal point solver for paracatadioptric fundamental matrices that we show is useful in a RANSAC framework, and which provides an initialization that is more accurate than a linear algorithm. A number of interesting problems remain open: How can we take into account prior calibration information, such as incorporating a constraint that the intrinsic parameters be equal? This would reduce the number of points required, and also lower the dimension of the problem, possibly improving performance at small baselines.

Acknowledgements

Christopher Geyer thanks Honeywell Research Labs for their generous support. The authors wish to thank the anonymous reviewers for their many helpful comments.



Figure 1. Top: Subset of twelve images used to perform 3D reconstruction. Bottom: Point cloud of reconstructed area near building.



Figure 2. If a fundamental matrix F is known to have $K_1 = K_2$ then, in the absence of noise, the nullspaces of F and F^T intersect in a single point that encodes the intrinsic parameters. Here we show the spans of the nullspaces for a set of estimated fundamental matrices, which nearly intersect in a single point; we use the intersections to calibrate the camera. Note that the nullspaces intersect the sphere at the two epipoles represented on the sphere.

References

- S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision (IJCV)*, 35(2):1 – 22, 1999.
- [2] J. Barreto and K. Daniilidis. A minimal solution for relative pose with unknown focal length. In *IEEE Conference* on Computer Vision and Pattern Recognition (CVPR), pages 1258–1265, Washington D.C., USA, June 2006.
- [3] P. Chang and M. Hebert. Omni-directional structure from motion. In Workshop on Omnidirectional Vision (OMNIVIS), 2000.
- [4] D. Cox, J. Little, and D. O'Shea. Using Algebraic Geometry. Springer Verlag, 1998.



Figure 3. . A histogram of the frequency of real roots during runs of the 9 point algorithm. On average there are 17.1 real roots that need to be tested out of a total of 64 roots.



Figure 4. A scatter-plot of errors in image center in pixels vs. the median of the parallax (distance between points) in rotationcorrected images. With greater parallax, variance in image center decreases.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications*-



Figure 5. (a) - (d): Errors in image center, focal length, rotation, and translation as a function of noise as a result of simulation. Lev-Marq9 denotes Levenberg-Marquardt initialized with the 9-point algorithm. Robust15 denotes the robust Levenberg-Marquardt initialized with the 15-point algorithm.

of-the-ACM, 24(6):381-95, 1981.

- [6] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *IEEE Conference* on Computer Vision and Pattern Recognition (CVPR), pages 125–132, 2001.
- [7] C. Geyer and K. Daniilidis. Structure and motion from uncalibrated catadioptric views. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 279– 286, 2001.
- [8] C. Geyer and K. Daniilidis. Properties of the catadioptric fundamental matrix. In *European Conference on Computer Vision (ECCV)*, pages 140–154, London, UK, 2002. Springer-Verlag.
- [9] C. Geyer and K. Daniilidis. Mirrors in motion: Epipolar geometry and motion estimation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 766–773, 2003.
- [10] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge Univ. Press, 2000.
- [11] S. Kang. Catadioptric self-calibration. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000.
- [12] M. Lourakis and A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Technical Report 340, Institute of Computer Science - FORTH, Heraklion, Crete, Greece, Aug. 2004. Available from http://www.ics.forth.gr/~lourakis/sba.
- [13] B. Micusik and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. In *IEEE Conference* on Computer Vision and Pattern Recognition (CVPR), 2003.
- [14] B. Micusik and T. Pajdla. Para-catadioptric camera autocalibration from epipolar geometry. In Asian Conference on Computer Vision (ACCV), 2004.
- [15] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *International Journal of Computer Vision (IJCV)*, 65:43 – 72, 2005.
- [16] D. Nistér. An efficient solution to the five-point relative pose problem. *PAMI*, 26(6):756–777, June 2004.
- [17] H. Stewénius, F. Kahl, D. Nistér, and F. Schaffalitzky. A minimal solution for relative pose with unknown focal length. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 789–794, San-Diego, USA, June 2005.
- [18] H. Stewénius, D. Nistér, M. Oskarsson, and K. Åström. Solutions to minimal generalized relative pose problems. In *Workshop on Omnidirectional Vision*, Beijing China, Oct. 2005.
- [19] T. Svoboda, T. Pajdla, and V. Hlavac. Epipolar geometry for panoramic cameras. In *European Conference on Computer Vision (ECCV)*, 1998.
- [20] C. Traverso. Gröbner trace algorithms. In Lecture Notes in Computer Science, ISAAC '88: Proceedings of the International Symposium ISSAC'88 on Symbolic and Algebraic Computation, pages 125–138, London, UK, 1989. Springer-Verlag.