Mirrors in motion: Epipolar geometry and motion estimation

Christopher Geyer
University of California, Berkeley
Berkeley, CA 94720
cgeyer@eecs.berkeley.edu

Kostas Daniilidis
University of Pennsylvania
Philadelphia, PA 19104
kostas@cis.upenn.edu

Abstract

In this paper we consider the images taken from pairs of parabolic catadioptric cameras separated by discrete motions. Despite the nonlinearity of the projection model, the epipolar geometry arising from such a system, like the perspective case, can be encoded in a bilinear form, the catadioptric fundamental matrix. *We show that all such matrices have equal* Lorentzian singular values, *and they define a nine-dimensional manifold in the space of* 4×4 *matrices. Furthermore, this manifold can be identified with a quotient of two Lie groups. We present a method to estimate a matrix in this space, so as to obtain an estimate of the motion. We show that the estimation procedures are robust to modest deviations from the ideal assumptions.*

1. Introduction

The widespread use of omnidirectional cameras in panoramic visualization and robot navigation motivated many of us to explore how such systems could be used beyond simple display or localization from vertical edges. For example it has been shown that the class of catadioptric sensors with a single viewpoint [3], i.e., equivalent to a perspective camera up to an *unknown* distortion, obey a simple projection model [8]. In this model we can describe the projections induced by these cameras as a composition of two central projections. The first is a central projection of the point in space to the sphere; the second is a central projection from a point lying between the sphere center and its north pole to a plane. This model puts traditional off-the-shelf and catadioptric cameras in the same context: the former is just a special case in which the two projection centers coincide.

It has long been suspected, and in places proven, that increasing the field of view improves the conditioning of motion estimation. What else can we expect from "mirrors in motion" regarding motion estimation? This is the question we study in this paper. A resulting algorithm would have significant impact, both on visualization and navigation: By taking few panoramic pictures with a high resolution catadioptric device we would be able not only to unwarp to any direction from the recording viewpoints but also from any other viewpoint up to occlusion. Regarding navigation, such an algorithm would be of great help in map building: several scene scans, even obtained by other sensors, could be registered correctly in the same coordinate system if we had accurate motion estimates.

One of the first issues we address is the appropriate representation of image features. Models of catadioptric projection are still non-linear, so what are efficient ways of decoupling motion and structure estimation? This has been achieved for perspective cameras using homogeneous coordinates, but what can we say in other cases? For parabolic cameras, does a space exist where the epipolar constraint can be written bilinearly? If so, what are the properties of the bilinear forms and what kind of space do all such matrices form? Then, how do we estimate optimally and robustly?

Recently we have been able to give answers which lead to simple and effective algorithms in all these questions. In [9, 10] we gave a representation of image features with which one can express the epipolar constraint and general multiview relations for parabolic systems as linear constraints. We showed that the set of transformations of this feature space is the group $O(3,1)$ of Lorentz transformations. Furthermore, the left and right nullspaces of the new *parabolic catadioptric fundamental matrix* encode the intrinsic parameters of the two cameras so that self-calibration reduces to kernel estimation.

In this paper our contribution is in the following points essential to estimation in the spaces of bilinear constraints:

- A complete characterization of the parabolic catadioptric fundamental matrices is given. It is shown that, similar to essential matrices for perspective cameras, all such bilinear forms have two equal non-zero *Lorentzian singular values*, even though they are constraints for uncalibrated cameras.
- The set of all of these fundamental matrices, and in fact the bilinear constraints for perspective cameras as well, can be identified with a quotient of Lie groups, which is known as a homogeneous space. In particular, the catadioptric fundamental matrices form a nine-dimensional manifold in the space of 4×4 matrices.
- The exponential map yields a nowhere-singular and surjective parameterization of these matrices. We present a correct estimation process of the epipolar geometry which at each step is restricted to the manifold.
- Demonstration with real and simulated experiments, the latter of which demonstrates that the proposed estimation algorithm is robust to modest deviations of the assumptions, such as misalignment of the camera and mirror.

Some of these results necessitate elementary concepts from differential geometry, algebra, and their intersection, namely the theory of Lie groups. We attempt to explain all of the concepts, albeit very briefly, and the reader is referred to two highly readable sources [2, 5].

With regards to related work, Baker and Nayar [13] described all possible configurations of central catadioptric cameras. Svoboda and Pajdla were the first to study the epipolar geometry of central catadioptric devices [22]. Kang devised a non-linear self-calibration algorithm for two views [15]. We introduced in [9] a bilinear epipolar constraint for parabolic cameras. Based on this work Sturm [21] derived linear multiview relationships mixing perspective and parabolic cameras. Simultaneously in [10] we introduced multiview linear constraints and derived conditions that an arbitrary matrix be a catadioptric fundamental matrix. A non-linear motion estimation algorithm has been proposed for both hyperbolic and parabolic systems in [7]. For the calibrated case, Gluckman and Nayar [12] estimate motion from the optical flow estimated in catadioptric images. Other calibrated algorithms making assumptions about the environment and produced for visualization can be found in [23, 1]. Calibrated motion estimation algorithms have been also introduced for non-central cameras [17, 18]. A valuable resource for omnidirectional vision literature is the book by Benosman and Kang [4] and the proceedings of the the four Workshops on Omnidirectional Vision. A useful reference on structure-from-motion can be found in [14] and [16].

2. Parabolic catadioptric projections

Suppose that a parabolic mirror is placed above an orthographically projecting camera so that the optical axes of the two are parallel. Such an optical system has a single effective viewpoint at the focus of the paraboloid. If the calibration parameters are known and previously taken into account then the projection induced by the combination of the mirror and lens can be expressed by the following function which we call the *canonical parabolic catadioptric projection* of the point $X = (x, y, z, w)$:

$$q(X) = \begin{cases} \frac{(x,y)}{-z+\operatorname{sign} w\sqrt{x^2+y^2+z^2}} \\ p_\infty \text{ if } x = y = 0 \text{ and } z/w > 0 \\ \text{undefined if } w = 0 \text{ or } X = (0,0,0,1) \end{cases} \tag{1}$$

In contrast, the *canonical perspective projection* is the map

$$p(X) = (x, y, z) \text{ undefined for } X = (0,0,0,1)\,,$$

where the range is $\mathbb{P}^2$, the *projective plane*. This can be expressed as a linear map from $\mathbb{P}^3$ to $\mathbb{P}^2$ using the 3×4 *canonical perspective camera matrix* $P = (I, 0)$; for then $p(X) = PX$. Both canonical projections p and q have the same viewpoint, namely the origin, expressed as $O = (0, 0, 0, 1)$.

Two exceptions are present in the definition of q but not that of p, let us try to justify them. First, notice that the denominator of the first case in (1) is zero only when the numerator is also zero. In particular, for all points $X = (0, 0, z, 1)$ with $z > 0$. Hence, the *projective plane is not a natural representation for parabolic catadioptric images*. Instead we add a single point, p_∞, to $\mathbb{R}^2$, and call this $\mathbb{R}^{2^*}$ or the *extended plane*. Second, observe that points on opposite sides of the viewpoint have different images, in particular $q(x, y, z, 1) \neq q(-x, -y, -z, 1)$. As a consequence, projections of points on the plane at infinity in $\mathbb{P}^3$ are ambiguous since their "sided-ness" with respect to the viewpoint is undefined, i.e., in $\mathbb{P}^3$ is defined such that $(x, y, z, 0) = (-x, -y, -z, 0)$.

In general, the parabolic camera will be uncalibrated and the actual induced projection may differ from the canonical projection by some transformation k which we assume consists of a scale and translation. *We will assume throughout this paper that the skew is zero and the aspect ratio is one.* If (c_x, c_y) is the image center and f is the focal length, then k is of the form:

$$k(u, v) = (c_x + fu, c_y + fv) \text{ and } k(p_\infty) = p_\infty\,. \tag{2}$$

The latter holds because the point at infinity is unaffected by either scale or translation. A general parabolic catadioptric projection can be modeled by a function, $g = k \circ q \circ e$, where $e \in SE(3)$ accounts for the world coordinate system.

3. Linear projection model

Recalling that p is the canonical perspective projection, consider the following question:

> *If $g = k \circ q$, does there exist a function h, independent of k, and a matrix A_k such that for all X,*
>
> $$p(X) = A_k\, h(\, g(X)\,)\,? \tag{3}$$

In other words, can we embed the *uncalibrated*, two-dimensional image plane in a higher dimensional space via h, such that the resulting image points are linearly related to the equivalent perspective projection? Though A_k may be unknown and dependent on k, the function h ought to be independent of k and X. If this were true then in any expression where we would use $p(X)$, e.g., the epipolar constraint, we could substitute the right hand side of (3) and A_k *would be absorbed into any linear constraint*.

We will demonstrate that for parabolic cameras there does exist a family of A_k's and an h. First, let h be defined by:

$$h(u, v) = (2u, 2v, u^2 + v^2 - 1, u^2 + v^2 + 1)^T, \tag{4}$$

with the exception that $h(p_\infty) = (0, 0, 1, 1) = N$. The range of h is the unit sphere S^2 in $\mathbb{P}^3$, equal to the set of points $p \in \mathbb{P}^3$ satisfying the quadratic form: $p^T Q p = 0$ where

$Q = \mathrm{diag}(1,1,1,-1)$. In addition, $h^{-1}|_{S^2}$, the inverse restricted to h's range, is *stereographic projection*, i.e., central projection from N, the north pole. It is not hard to show that

$$(h \circ k)(u,v) \;=\; \big(2(c_x + fu),\, 2(c_y + fv),\, \ldots\,\big)^T,$$

is linearly dependent on $h(u,v)$. Thus, there is a 4×4 matrix K, defined in equation (12) of the appendix, such that

$$(h \circ k)(u,v) \;=\; K\,h(u,\,v)\,.$$

In other words, the similarity transformation k in $\mathbb{R}^{2^*}$ induces a projective transformation K which preserves S^2. This commutative relationship is shown in Figure 1.

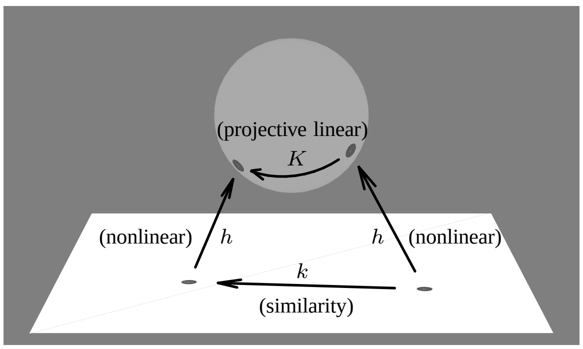

Figure 1: A similarity or change of intrinsics commutes linearly with a projective linear transformation which preserves the sphere.

To denote the application of K we write $k'(X) = K\,X$. Now, suppose a catadioptric projection $g = k \circ q$, so $e = id$; we can rewrite the composition:

$$h \circ g \;=\; h \circ k \circ q \;=\; k' \circ h \circ q\,, \tag{5}$$

Observe the following algebraic derivation:

$$\begin{aligned}(p \circ h \circ q)(x,y,z,1) &= (p \circ h)\,[(x,y)/(-z+r)] \\ &= p\,[2(x,y,z,r)/(r-z)] \\ &= p(x,y,z,1)\,,\end{aligned}$$

where $r = \sqrt{x^2+y^2+z^2}$. This proves that $p \circ h \circ q = p$. Putting this equation and equation (5) together yields:

$$\begin{aligned}p(X) &= (p \circ h \circ q)(X) \\ &= (p \circ k'^{-1} \circ k' \circ h \circ q)(X) \\ &= (p \circ k'^{-1} \circ h \circ g)(X) \\ &= \underbrace{PK^{-1}}_{A_k} h(g(X))\,.\end{aligned}$$

Thus we have verified that equation (3) can be satisfied when $A_k = PK^{-1}$ and h is defined in (4). In subsequent sections we will use the following shorthand notation:

$$\widetilde{x} \;=\; h(x)\,. \tag{6}$$

4. The epipolar constraint

It is well known that two perspective projections $f_1 = p \circ e_1$ and $f_2 = p \circ e_2$ satisfy the following equation known as the epipolar constraint:

$$f_2(X)^T E\, f_1(X) \;=\; 0\,, \tag{7}$$

for some *essential matrix* E depending on the inter-frame motion $e_2 \circ e_1^{-1}$. If $e_2 \circ e_1^{-1} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}$, then $E = \widehat{t}R$.

According to the last section, if $g_1 = k_1 \circ q \circ e_1$ and $g_2 = k_2 \circ q \circ e_2$, then there exist matrices K_1 and K_2, respectively depending on k_1 and k_2 such that

$$f_1(X) \;=\; PK_1^{-1}\widetilde{g_1(X)} \text{ and } f_2(X) \;=\; PK_2^{-1}\widetilde{g_2(X)}$$

where the operator $\widetilde{\cdot}$ is defined in equation (6). Now substitute the right hand sides into equation (7):

$$\widetilde{g_2(X)}^T \underbrace{K_2^{-T} P^T E\, P\, K_1^{-1}}_{F}\, \widetilde{g_1(X)} \;=\; 0\,. \tag{8}$$

The matrix F is called the *parabolic catadioptric fundamental matrix*, first introduced in [9]. Thus, in spite of the nonlinearity of q, there is a bilinear constraint on the liftings of corresponding image points in two parabolic catadioptric views.

This matrix defines the epipolar geometry for the pair of parabolic projections. For example, the locus of all points y satisfying $\widetilde{y}^T F \widetilde{x} = 0$ for any constant x is an *epipolar circle*, and vice versa for all x and any constant y. The equations $F\widetilde{x} = 0$ and $F^T\widetilde{y} = 0$ each have two solutions, and each solution pair defines an *epipolar point pair* $(e_{i,1}, e_{i,2})$, equivalent to one epipole in a perspective image.

5. Self-calibration

In general, the transformations k_1 and k_2 are unknown, therefore we ask:

> *What information about k_1 and k_2 can be inferred from F?*

We shall show that if $k_1 = k_2 = k$ then under most conditions we can infer k from F.

We begin by examining properties of the projective linear transformation specified by the matrix K. Up until now we have only considered its effect on points $\widetilde{x}$ lying on the sphere. Clearly, though, it acts on all of $\mathbb{P}^3$. Let us consider its effect on the origin $O = (0,0,0,1)$. Using the definition of K found in equation (12), we can directly calculate its transformation, $O' = KO$:

$$(2c_x, 2c_y, c_x^2 + c_y^2 + 4f^2 - 1, c_x^2 + c_y^2 + 4f^2 + 1)^T. \tag{9}$$

If c_x, c_y and f are unknown we can calculate them from the point O'. Thus, if we are somehow able to determine O', we can find K.

We now show that O' can be obtained directly from the fundamental matrix. To begin with, since O is the viewpoint of the canonical perspective projection, it satisfies $PO = 0$. Therefore, if $O_1 = K_1 O$, then

$$F O_1 \;=\; K_2^{-T} P^T E \left(P K_1^{-1} K_1 O\right) \;=\; 0 \,.$$

Similarly, $F^T O_2 = 0$ where $O_2 = K_2 O$. Consequently $O_1 \in \mathcal{N}(F)$ and $O_2 \in \mathcal{N}(F^T)$, where $\mathcal{N}(F)$ denotes the nullspace of F. The fundamental matrix, however, is rank two and therefore its nullspaces are two-dimensional. Thus neither O_1 nor O_2 are uniquely defined by these conditions. Yet, if it is known *a priori* that $K_1 = K_2$, then

$$O_1 \;=\; O_2 \;\in\; \mathcal{N}(F) \cap \mathcal{N}(F^T) \,.$$

Provided $\mathcal{N}(E) \neq \mathcal{N}(E^T)$, the intersection is a single point in $\mathbb{P}^3$ and we can uniquely recover c_x, c_y and f. It can be shown that the nullspaces coincide only when $R\,t = t$, i.e., the axis of rotation is about the direction of translation.

6. The Lorentz group

The matrix K, defined in equation (12) of the appendix, can be shown to satisfy $K^T Q K = Q$. If $p' = Kp$ for any point such that $p^T Q p = 0$, then apparently ${p'}^T Q p' = 0$. In other words K preserves point-wise the unit sphere. There are, however, many more matrices A which are not in the form given in equation (12) but which do satisfy $A^T Q A = Q$. In fact like the set of matrices satisfying $A^T A = I$, i.e., the orthogonal group, the set of all matrices A satisfying $A^T Q A = Q$ is a Lie group, i.e., a group whose elements form a manifold. This group is called the *Lorentz group* and is denoted by $O(3,1)$. The $(3,1)$ denotes the fact that the diagonal of Q consists of three $+1$'s and one -1, and there is a study of more general groups $O(p,q)$.

The structure of a matrix Lie group is determined by the possible derivatives of curves through the identity matrix. Suppose that $A(t)$ is a curve restricted to $O(3,1)$ such that $A(0) = I$. Then its derivative at $t = 0$ satisfies:

$$A'(0)^T Q + Q A'(0) \;=\; 0 \,.$$

The derivative is therefore of the form

$$A'(0) \;=\; \begin{pmatrix} \hat{x} & y \\ y^T & 0 \end{pmatrix} \;=\; \mathcal{L}(x,y) \,, \tag{10}$$

for some $x, y \in \mathbb{R}^3$, and therefore the Lorentz group is six-dimensional. We let $\mathfrak{so}(3,1)$ be the set of all such matrices. It can be shown that if $X \in \mathfrak{so}(3,1)$, then $e^X \in O(3,1)$, where $e^X = I + X + X^2/2 + \cdots$. A closed-form formula, analogous to Rodrigues' formula for skew-symmetric matrices, is given in equation of the appendix. It is proven in [2] that e^X is surjective on the connected component of the identity in $O(3,1)$.

Looking back at K, notice that the set of all K as c_x, c_y and f vary over all real numbers, with $f > 0$, is a three-dimensional Lie subgroup of $O(3,1)$. We shall denote by $\mathcal{K}$ the set of all such matrices and call it the *calibration subgroup* (for parabolic projections). It can be shown that when $K \in O(3,1)$, $K \in \mathcal{K}$ if and only if $K p_\infty = p_\infty$.

This demonstrates that a subgroup of $O(3,1)$ is relevant to calibration, what, though, are the effects of the remaining Lorentz transformations? It can be shown that any $A \in O(3,1)$ has the following unique decomposition:

$$A \;=\; \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix} K \;\text{ where }\; R \in O(3) \;\text{ and }\; K \in \mathcal{K} \,.$$

The first part of the decomposition has the effect of rotating the sphere about the origin, and is therefore equivalent to a pure rotation of the scene. Consequently any element of the Lorentz group is a uniquely factorable composition of an element of $\mathcal{K}$, which accounts for the intrinsic parameters of the camera, and an element of $O(3)$, which accounts for the rotation of the scene.

7. Properties of the fundamental matrix

We are interested in the following question not resolved in [10]:

> *Under what condition(s) is an arbitrary 4×4 matrix a parabolic fundamental matrix?*

Recall that if we let $\mathcal{E}$ be the set of perspective fundamental matrices, then $E \in \mathcal{E}$ if and only if E has equal singular values, i.e., for some $U, V \in O(3)$, $E = U^T \operatorname{diag}(1,1,0) V$. Let us denote by $\mathcal{F}$ the set of all parabolic catadioptric fundamental matrices. The following theorem offers an analogous characterization of any $F \in \mathcal{F}$.

Theorem 1. A matrix F can be decomposed into the form defined in equation (8) *if and only if* there exist matrices $U, V \in O(3,1)$ such that

$$F \;=\; U^T \operatorname{diag}(1,1,0,0)\, V \,. \tag{11}$$

Moreover, a decomposition can be found where $K_1 = K_2$ only if $\det(F + F^T) = 0$.

In other words the *Lorentzian singular value decomposition* of F must have equal Lorentzian singular values. This holds even though F is a bilinear constraint for *uncalibrated* cameras. One mistake that should not be made is that an orthogonal projection to $\mathcal{F}$ *will not be obtained by averaging the Lorentzian singular values*, for the Frobenius distance between the original matrix and the projected one will not be minimized. Unfortunately this does not answer the question of optimal projection.

Because of space constraints, we give only an outline of a proof. The forward is trivial because the K_i's absorb the rotations of the singular value decomposition of E, leaving elements of $O(3,1)$. The reverse relies on the decomposition of the element $U \in O(3,1)$ into $U = RK$ where $K \in \mathcal{K}$ and $R \in O(3)$, stated in the previous section. Note that this theorem can be made more specific and restrict U and V to lie in $O(3,1)_0$, the connected component of the identity.

8. Global characteristics of $\mathcal{F}$

Finally we consider the following questions:

> *What is the structure of the set of all parabolic fundamental matrices? If it is a manifold what is its dimension?*

To begin with we make the following observation which is a corollary of Theorem 1: $F \in \mathcal{F}$ if and only if $A^T F \in \mathcal{F}$ for any $A \in O(3,1)_0$. In other words, $\mathcal{F}$ is closed under left multiplication by elements of $O(3,1)_0$. Similarly $F \in \mathcal{F}$ if and only if $FA \in \mathcal{F}$. Consequently, elements of the group $\mathcal{G} = O(3,1)_0 \times O(3,1)_0$ *act* upon elements of $\mathcal{F}$ via a map φ in the following way:

$$\varphi\big(\,(A_1, A_2), F\big) \;=\; A_1^T F A_2\,.$$

Theorem 1 guarantees that the range $\varphi(\mathcal{G}, F_0) \subset \mathcal{F}$. Now suppose that $g = (G_1, G_2)$, $h = (H_1, H_2)$ and $e = (I, I)$. The map φ satisfies $\varphi(e, F) = F$ and $\varphi(g \cdot h, F) = \varphi(g, \varphi(h, F))$, where $g \cdot h = (G_1 H_1, G_2 H_2)$. Any such function is known as a *group action*.

What might this have to do with the structure of $\mathcal{F}$? For example, it is clearly not a group, but the group $\mathcal{G}$ acts on it. We start by noting that φ, and therefore $\mathcal{G}$, can be used to parameterize $\mathcal{F}$. If $\varphi(\mathcal{G}, F_0) = \mathcal{F}$, for any constant F_0, in other words is surjective in $\mathcal{F}$, then $\varphi(\cdot, F_0)$ is a parameterization of $\mathcal{F}$. This is also a corollary of Theorem 1, for if

$$F_{i=0,1} = U_i^T \Sigma V_i \text{ then } \varphi\big((U_0^{-1} U_1, V_0^{-1} V_1), F_0\big) = F_1,$$

where $\Sigma = \operatorname{diag}(1,1,0,0)$. So, $\varphi(\cdot, F_0)$ is surjective in $\mathcal{F}$.

If it were the case that $\varphi_{F_0} = \varphi(\cdot, F_0)$ provided a one-to-one mapping between $\mathcal{G}$ and $\mathcal{F}$, then it would show that $\mathcal{F}$ is a group since φ_{F_0} would be an isomorphism. However, φ_{F_0} is not one-to-one in this case. What, though, is the redundancy in its parameterization of $\mathcal{F}$? Consider the set of elements for which $\varphi_{F_0} = F_0$, i.e.,

$$\mathcal{H}_{F_0} \;=\; \{g \in \mathcal{G} : \varphi(g, F_0) = F_0\} \;\subset\; \mathcal{G}\,.$$

Since it is closed under composition and inversion, and preserves F_0, it is called the *isotropy subgroup* of F_0. Furthermore, it can be shown that it is a 3-dimensional *Lie* subgroup.

Now we make the following claim. The *left cosets* of $\mathcal{H}_{F_0}$, i.e., $g\mathcal{H}_{F_0} = \{gh : h \in \mathcal{H}_{F_0}\}$, are in one-to-one correspondence with the elements of $\mathcal{F}$. Thus $\mathcal{F}$ is equivalent to the quotient space $\mathcal{G}/\mathcal{H}_{F_0}$, the quotient space being the partition of $\mathcal{G}$ into the pairwise disjoint left cosets. To prove this claim one must show that $\widetilde{\varphi}_{F_0}(g\mathcal{H}_{F_0}) = \varphi_{F_0}(gh)$ is constant for all $h \in \mathcal{H}_{F_0}$, and therefore $\widetilde{\varphi}_{F_0}(g\mathcal{H}_{F_0})$ is well-defined; then we must show that $\widetilde{\varphi}_{F_0}$ is injective. Both properties follow from the properties of the group action φ or cosets.

A theorem of differential geometry, see [5] for example, states that the quotient of a Lie group $\mathcal{G}$ and some Lie subgroup $\mathcal{H}$ is a manifold of dimension $\dim \mathcal{G} - \dim \mathcal{H}$. This and the arguments above justify the following theorem. Since all of the preceding arguments apply equally well to essential matrices, we can also obtain an analogous result.

Theorem 2. The set of parabolic catadioptric fundamental matrices is equal to a quotient of Lie groups:

$$\mathcal{F} \;\;\cong\;\; O(3,1) \times O(3,1) \,/\, \mathcal{H}_{F_0}\,,$$

where $\mathcal{H}_{F_0}$ is the isotropy subgroup of any $F_0 \in \mathcal{F}$. Consequently $\mathcal{F}$ is a nine-dimensional manifold. Similarly, the set of essential matrices $\mathcal{E}$ is given by:

$$\mathcal{E} \;\;\cong\;\; O(3) \times O(3) \,/\, \mathcal{H}_{E_0}\,,$$

where $\mathcal{H}_{E_0}$ is the isotropy subgroup of any $E_0 \in \mathcal{E}$ in the context of the natural action $\varphi : O(3) \times O(3) \times \mathcal{E} \to \mathcal{E}$.

9. Epipolar geometry estimation

Now that the epipolar constraint is shown to be bilinear in the liftings of the corresponding points, most of the perspective algorithms for motion estimation apply. The most important difference is, so far, the absence of a linear orthogonal projection to $\mathcal{F}$. We are not yet aware of a linear 9-point algorithm. However, supposing $\{(p_i, q_i)\}_{1 \le i \le N}$ is a set of point correspondences, we can minimize, in the usual way, the Sampson approximation to the geometric error [19]:

$$E(F) \;=\; \sum_{i=1}^{n} \frac{\big(\widetilde{p}_i^T F \widetilde{q}_i\big)^2}{\left\| \frac{\partial}{\partial(p_i, q_i)} \widetilde{p}_i^T F \widetilde{q}_i \right\|^2} \;=\; \sum_{i=1}^{n} \frac{f^T b_i b_i^T f}{f^T A_i A_i^T f}$$

where $b_i = \widetilde{q}_i \otimes \widetilde{p}_i$ and

$$A \;=\; \begin{pmatrix} \widetilde{p}_{i\,u} \otimes \widetilde{q}_i & \widetilde{p}_{i\,v} \otimes \widetilde{q}_i & \widetilde{p}_i \otimes \widetilde{q}_{i\,u} & \widetilde{p}_i \otimes \widetilde{q}_{i\,v} \end{pmatrix}.$$

The notations $\widetilde{p}_u$ and $\widetilde{p}_v$ respectively denote partial differential of h with respect to u and v, evaluated at $p = (u_0, v_0)$.

The map φ_{F_0} conveniently provides us with a global and nowhere-singular, though redundant, parameterization of $\mathcal{F}$. We define $\phi_{F_0} : \mathbb{R}^{12} \to \mathcal{F}$ given by:

$$\phi_{F_0}(x_1, y_1, x_2, y_2) \;=\; e^{\mathcal{L}(x_1, y_1)^T} F_0\, e^{\mathcal{L}(x_2, y_2)}\,,$$

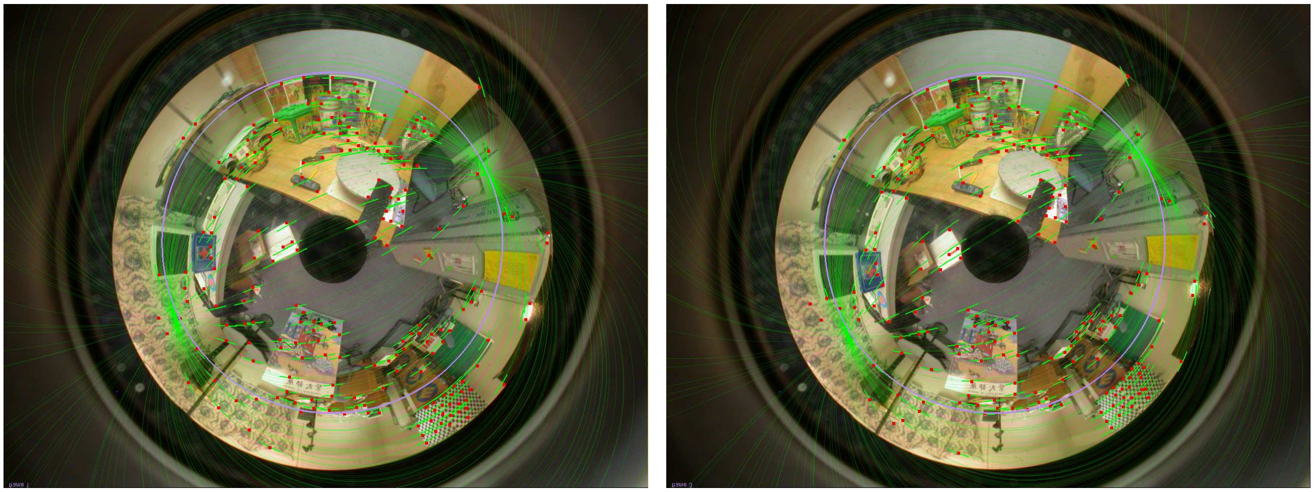

Figure 2: Estimation of the epipolar geometry found by minimizing the Sampson error. Each green line is the epipolar circle determined from the point in the other frame, hence the points do not lie exactly on the epipolar circles. The epipolar circle corresponding to a given point is highlighted in the region of the point. Note that in parabolic catadioptric images there are always two epipoles present in the image. The light purple circle encodes the intrinsic parameters; its center is the image center and its radius is twice the focal length, it has been found by choosing the least singular value of $(F, F^T)^T$.

where $\mathcal{L}$ is defined in (10). Equation (13) gives an explicit formula for the exponential map. If $p = (x_1, y_1, x_2, y_2)$, then we can directly calculate the Jacobian $J = \partial/\partial p E(\phi_{F_0}(p))$ and the 12×12 Hessian $H = \partial^2/\partial p^2 E(\phi_{F_0}(p))$ using the approximation $e^X = I + X$.

Algorithm 1. (Motion estimation)

1. Obtain an initial estimate F_0 by guessing (poor) calibration matrices K_1 and K_2, and estimating an essential matrix E_0. Then $F_0 = K_2^{-T} P^T E_0 P K_1^{-1}$.

2. With the initial estimate, minimize the function $E(F)$ using Levenberg-Marquardt by explicitly calculating J and H . Let $\hat{F}$ be the final local minimum.

3. If it is assumed that $K_1 = K_2$, then the least singular vector of $(\hat{F}, \hat{F}^T)$ is an estimate of the vector O' in equation (9), whereby we can estimate c_x, c_y and f and consequently K.

4. $\hat{E} = PK^T \hat{F} K P^T$ is an estimate of the essential matrix which can be factored to obtain four possible motions. Only one leads to a consistent reconstruction.

We have implemented this algorithm and tested it using features obtained from an automatic point tracker [20]. The estimated epipolar geometry is shown in Figure 2. In Figure 3 we show three images, the left and right of which are rectifications of the images of Figure 2. We lack the space here to describe the rectification in this paper, except to say that it is *conformal* and so locally there is no distortion; it is described in [11]. Using the disparity estimated with a stereo algorithm we show in the middle image a warping of the right image to match the left image.

10. Model deviations

We have, until now, assumed that the projection model is ideal, that the camera is exactly orthographic and that the mirror is aligned and has the shape of a paraboloid. In this section we examine through simulation the effects that deviations from these assumptions have on the estimation of the motion. When these deviations occur, there is no closed form solution for the projection of a point in space. Hence, in the simulations, to calculate the projection of a point we use the principle of least action in which the image of a point in space is determined from the path with the shortest distance from that point to the focus of the camera.

We now describe four simulations in which a single parameter is varied and errors in rotation and translation are shown. In Figure 4 are graphs of the error in translation and rotation as a function of the independent variable. The error in rotation equals $\| \log \hat{R} R_0^{-1} \|$ where $\hat{R}$ and R_0 are the estimated and true rotations respectively. The error in translation is similarly $\cos^{-1} \hat{t}^T t_0$, where $\hat{t}$ and t_0 are the estimated and true translations respectively. From left to right in Figure 4:

- A perspective camera with focal length 20 times the focal length of the mirror. The optical axes remain parallel but are displaced a distance x lying between 0.5 to 3 times the mirror's focal length.
- An orthographic camera and a mirror which is rotated at angle of θ about its focus, where $0^\circ \leq \theta \leq 5^\circ$.
- A perspective camera, with optical axis aligned with mirror, and where the ratio ρ of the distance between the mirror focus and camera focus to the mirror focal length varies between ∞ (orthographic) and 13.

Figure 3: Left: Rectification of left image in Figure 2. Middle: Warping of right rectification to left using estimated disparity. Right: Rectification of right image in Figure 2.

- An orthographic camera and a mirror whose eccentricity ε varies from 0.9 to 1.1 (1 is a paraboloid).

We find that the motion estimation algorithm is robust to modest deviations from the ideal assumptions. It appears that the usage of a perspective camera has little effect on the results, though there is a greater effect if the optical axes are not aligned.

11. Conclusion

In conclusion we have presented a framework for structure from motion for parabolic catadioptric systems. The methods presented formulate the structure from motion in a way which linearly decouples the structure and the motion despite a non-linear projection model. We demonstrated local and global properties of the manifold of fundamental matrices and the formulation of the set as a quotient of groups, and which has given us insight into the perspective case. We propose an algorithm for estimation of the epipolar geometry which minimizes on the nine-dimensional manifold of fundamental matrices. Finally, we have shown that these algorithms are robust to small deviations from ideal assumptions.

We hope to address some of the following problems arising out of this work:

Estimation algorithms. Ideally we would like to find linear algorithms for estimation of fundamental matrices. Another issue is that the decreased resolution per viewing angle effectively decreases the magnitude of the baseline, therefore algorithms robust to small baselines should be addressed.

Harmonic analysis. We believe one of the most important contributions of this paper is the characterization of the set of bilinear forms as a quotient space of Lie groups. Because of their symmetry, it is possible to define the Fourier transform of functions on such spaces. Is it possible to combine elements of structure-from-motion with methods in signal processing?

Control on $\mathcal{F}$. Control has been investigated on quotient spaces, e.g. [6], can these ideas be extended to visual servoing directly on the quotient space $\mathcal{F}$?

Acknowledgements. The authors are grateful for support through the following grants: NSF-IIS-0083209, NSF-EIA-0218691, NSF-IIS-0121293 and DARPA DAAH-19-02-1-0383. The authors would also like to thank Rana Molana, Rene Vidal, Jana Košecká and Ameesh Makadia for their helpful comments on the paper.

Appendix

First we give the matrix K depending on the intrinsic parameters c_x, c_y and f:

$$K = \begin{pmatrix} 2f & 0 & 2f c_x & -2f c_x \\ 0 & 2f & 2f c_y & -2f c_y \\ -c_x & -c_y & 1 - c_x^2 - c_y^2 + f^2 & 1 + c_x^2 + c_y^2 - f^2 \\ c_x & -c_y & 1 - c_x^2 - c_y^2 - f^2 & 1 + c_x^2 + c_y^2 + f^2 \end{pmatrix}. \quad (12)$$

Second we give the analog to Rodrigues' formula for matrices $A \in \mathfrak{so}(3,1)$:

$$e^A = \left(\frac{\lambda_2 \sinh \lambda_1 - \lambda_1 \sinh \lambda_2}{\lambda_1 \lambda_2 (\lambda_1^2 - \lambda_2^2)} \right) A^3 + \left(\frac{\cosh \lambda_1 - \cosh \lambda_2}{\lambda_1^2 - \lambda_2^2} \right) A^2 + \left(\frac{-\lambda_2^3 \sinh \lambda_1 + \lambda_1^3 \sinh \lambda_2}{\lambda_1 \lambda_2 (\lambda_1^2 - \lambda_2^2)} \right) A \left(\frac{-\lambda_2^2 \cosh \lambda_1 + \lambda_1^2 \cosh \lambda_2}{\lambda_1^2 - \lambda_2^2} \right) I. \quad (13)$$

where $\{\lambda_1, -\lambda_1, \lambda_2, -\lambda_2\}$ are the eigenvalues of A. If A has two zero eigenvalues, then there is a simpler form resembling Rodrigues' formula which we leave to the reader.

References

[1] D. Aliaga. Accurate catadioptric calibration for real-time pose estimation in room-size environments. In *Proc. Int. Conf. on Computer Vision*, pages 127–134, Vancouver, Canada, Jul. 9-12, 2001.

[2] A. Baker. *Matrix Groups: An Introduction to Lie Group Theory*. Springer Verlag, 2002.

[3] S. Baker and S. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35:175–196, 1999.

[4] R. Benosman and S. Kang. *Panoramic Vision*. Springer-Verlag, 2000.

[5] W. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, London, 1986.

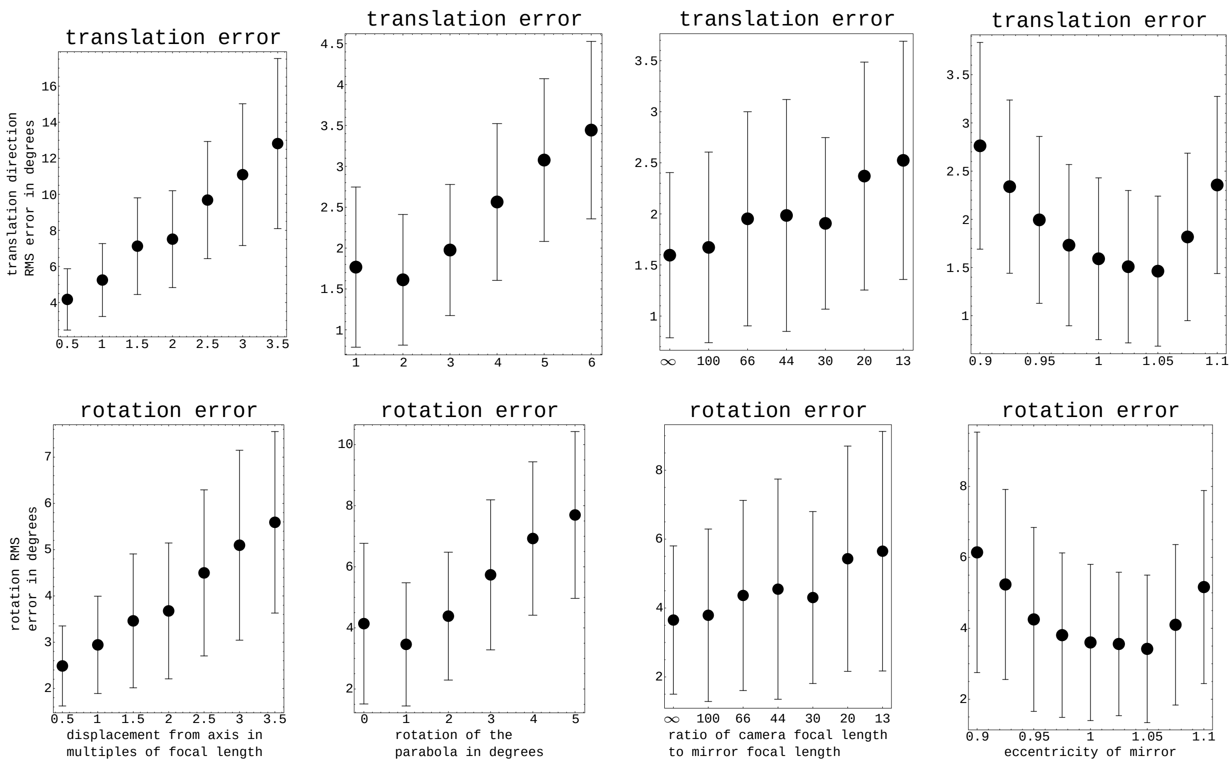

Figure 4: Experiments demonstrating robustness to deviations from the ideal assumptions. See text for details.

[6] R. W. Brockett. System theory on group manifolds and coset spaces. *SIAM J. Control*, 10(2):265–284, 1972.

[7] P. Doubek and T. Svoboda. Reliable 3d construction from a few catadioptric images. In *Workshop on Omnidirectional Vision, Copenhagen, June 22*, pages 71–78, 2002.

[8] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 43:223–243, 2001.

[9] C. Geyer and K. Daniilidis. Structure and motion from uncalibrated catadioptric views. In *IEEE Conf. Computer Vision and Pattern Recognition*, Hawaii, Dec. 11-13, 2001.

[10] C. Geyer and K. Daniilidis. Properties of the catadioptric fundamental matrix. In *Proc. Seventh European Conference on Computer Vision*, pages 140–154, Copenhagen, Denmark, 2002.

[11] C. Geyer and K. Daniilidis. Conformal rectification of omnidirectional stereo pairs. In *Workshop on Omnidirectional Vision and Camera Networks, Madison, WI, June 22*, 2003.

[12] J. Gluckman and S. Nayar. Ego-motion and omnidirectional cameras. In *Proc. Int. Conf. on Computer Vision*, pages 999–1005, Bombay, India, Jan. 3-5, 1998.

[13] J. Gluckman and S. Nayar. Rectifying transformations that minimize resampling effects. In *IEEE Conf. Computer Vision and Pattern Recognition*, Hawaii, Dec. 11-13, 2001.

[14] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge Univ. Press, 2000.

[15] S. Kang. Catadioptric self-calibration. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages I–201–207, Hilton Head Island, SC, June 13-15, 2000.

[16] Y. Ma, S. Soatto, J. Košecká, and S. Sastry. *An Invitiation to 3D Vision: From Images to Geometric Models*. Springer Verlag, New York, 2003.

[17] J. Neumann, C. Fermüller, and Y. Aloimonos. Eyes from eyes: New cameras for structure from motion. In *Workshop on Omnidirectional Vision, Copenhagen, June 22*, pages 19–26, 2002.

[18] R. Pless. Discrete and differential two-view constraints for general imaging. In *Workshop on Omnidirectional Vision, Copenhagen, June 22*, pages 53–59, 2002.

[19] P. Sampson. Fitting conic sections to very scattered data: an iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, 18:97–108, 1982.

[20] J. Shi and C. Tomasi. Good features to track. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 593–600, Seattle, WA, June 21-23, 1994.

[21] P. Sturm. Mixing catadioptric and perspective cameras. In *Workshop on Omnidirectional Vision, Copenhagen, June 22*, pages 37–44, 2002.

[22] T. Svoboda, T. Pajdla, and V. Hlavac. Epipolar geometry for panoramic cameras. In *Proc. 5th European Conference on Computer Vision*, pages 218–231, 1998.

[23] C. J. Taylor. Video plus. In *IEEE Workshop on Omnidirectional Vision, Hilton Head, SC, June 12*, pages 3–10, 2000.