A geometric solver for calibrated stereo egomotion

Enrique Dunn[1], Brian Clipp[2], Jan-Michael Frahm[1]

[1]University of North Carolina at Chapel Hill
{dunn, jmf}@cs.unc.edu

[2]Applied Research Associates
bclipp@ara.com

Abstract

This paper introduces a novel geometrical solution for the pose estimation of a stereo camera system as commonly used in robotics, where the camera system balances between coverage and overlap. The proposed approach considers a set of features observed, respectively, in four, three and two views. In contrast to most algebraic solutions our constraints are geometrically meaningful. Initially, we use a four view feature to restrict our translation vector to lie on the surface of a sphere while setting orientation as a function of translation up to a single rotational degree of freedom. Next, we use a three view feature to restrict the translation vector to lie on a circle on the sphere, while completely defining orientation as a function of translation. Finally, we use a two view feature to determine the translation vector lying on the intersection of the circle and one of the generator lines of a doubly ruled quadric. We show how for this final step, the problem can be reduced to the intersection of two coplanar circles. We also analyze the degenerate configurations of the proposed solver and perform an experimental evaluation.

1. Introduction

In recent years cameras have established themselves as ubiquitous sensors in robotic systems. Within these systems, stereo cameras can provide single frame scale estimation for visual odometry [16]. However, to achieve cost effectiveness the stereo system has to balance between the overlap among the cameras and the overall scene coverage. As demonstrated in Clipp et al. [3], favoring coverage instead of overlap leads to a degrading performance in the camera pose estimation of the traditional three point pose estimation [5, 13, 6]. We propose a novel approach that requires a significantly smaller overlap region than traditional methods by incorporating features that are only seen in a subset of the frames. Our method uses a novel three feature solver for calibrated stereo cameras. The solver uses a minimal set of geometric constraints, which determine all six degrees of freedom (DOF) of the relative pose. Hence, our method meets the requirements given by robotic camera systems and can be used within typical robust estimation frameworks like RANSAC [1].

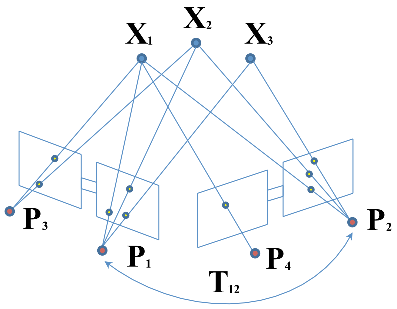

Figure 1. Geometric description of our stereo egomotion instance.

We consider a set of features observed, respectively, in four, three and two views to account for the different overlap characteristics of typical robotic stereo cameras. The sequential analysis of each feature leads to the estimation of different geometric entities, which are determined through straightforward algebraic formulations. This reduces the relative pose estimation to finding the intersection of a pair of coplanar circles.

Our approach differs from the standard implementation of P3P on stereo cameras, where a trio of features is triangulated in one stereo pair and their 3D estimates are used to solve P3P given the corresponding image measurements in the second stereo pair [16]. Instead, we leverage the Euclidean structure provided by having a common feature triangulated in both stereo pairs. In fact, our approach builds exclusively on Euclidean geometry concepts in order to sequentially eliminate the stereo pose parameters. In doing so, we circumvent the use of algebraic geometry methods, which have recently garnered considerable interest in the solution of minimal problems in computer vision [10].

The paper first discusses the related work and defines the notation. Then our geometric solver is introduced and we present the experimental evaluation.

978-1-4577-1102-2/11/$26.00 ©2011 IEEE

2. Related Work

The method described in this paper extends the state of the art in 6DOF motion for rigid, overlapping stereo pairs by introducing a new solution method based on a novel combination of feature correspondences between the four images.We only discuss methods that relate to the scaled, 6DOF motion of a multi-camera system. Nister et al. [16] demonstrate the use of the standard perspective three point (P3P) method [5] to calculate the motion of a stereo camera. Corners in the stereo pair at time zero are matched, triangulated and then the pose of the left (or right) camera of the stereo system are determined using the P3P method. Correspondences are then verified in both the left and right images using RANSAC. Nister also developed a generalized version of the P3P method [15]. This generalized P3P could use three projections of three 3D features in any combination of the one to three images of a trinocular camera.

Ni and Dellaert [14] propose a 6DOF motion estimation for a stereo camera based on decomposing the motion into rotation and translation. They first solve for the rotation of the system using points at infinity and then non-linearly solve for translation. Their method requires an initial solution close to the true motion to start the minimization. Additionally, the requirement for points at infinity poses strong constraints on the applicability in indoor scenes. Our method does not require any particular point distribution enabling indoor scenes.

The generalized camera model [4, 17] models imaging devices where each ray may have its own associated camera center. Pless [17] later used the generalized motion model to formulate constraints for the motion of a generalized camera. An algorithm based on these constraints was presented by Li et al. [11]. Stewenius et al. [19] developed a minimal solution to obtain up to sixty-four motion solutions.Stereo cameras are one of the degenerate camera configurations of the approach of Stewenius et al. [19].

Kim and Chung[9] studied motion estimation for non-overlapping, rigid, two-camera systems and proposed an extended Kalman filter based 6DOF tracking method based on their observations. A minimal solution method for the 6DOF motion of a non-overlapping, rigid, two-camera system was introduced by Clipp et al. [2] and a non-minimal second order cone programming solution was proposed by Kim et al. [8]. Both solution are degenerate in the case of pure translational motion.

More recently, Clipp et al. [3] have developed a method for estimating the motion of a slightly overlapping camera pair. This configuration increases scene coverage by sacrificing overlap to a minimum while still allowing 6DOF motion estimation regardless of motion. Their method uses one feature visible in all four views of the two poses of a stereo camera to fix the second camera's distance with respect to the four-view feature, which is triangulated in the first view. Three temporal feature correspondences visible in only two-views each are then used to restrict the remaining three degrees of freedom of the second camera. They demonstrated that when the overlap between the two rigidly mounted cameras on a stereo head is reduced the P3P method's accuracy is degraded while their method had constant accuracy.

Instance $\langle k, l, m\rangle$	Ref.	Approach
$\langle$ 0,3,0$\rangle$	[15]	Intersection of a quartic and circle
$\langle$ 0,0,6$\rangle$	[2]	Gröbner Basis + Linear
$\langle$ 1,0,3$\rangle$	[3]	Gröbner Basis
$\langle$ 1,1,1$\rangle$	[this]	Intersection of two circles

Table 1. Minimal solvers for stereo motion estimation.

In contrast to Clipp et al.[3], we target camera systems that still have significant overlap of the cameras as these are very common in robotics. Moreover, our method models the camera motion directly, without the need for a first order approximation of the rotation through a Rodrigues representation. Accordingly, it is better equipped to handle large camera motions. Additionally, we provide a direct geometric solution and avoid the numerical instabilities and implementation difficulties common to Gröbner basis solvers.

3. Problem description and notation

We seek to estimate the 3D rigid transformation between two pairs of cameras, where the intra-pair geometry between each camera pair and the intrinsic parameters are known. We consider the projection of a set of 3D points $\{\mathbf{X}_i \in \mathbb{R}^3 : i = 1\ldots 3\}$, observed by a set of cameras $\{\mathbf{P}_j \in \mathbb{R}^{3\times 4} : j = 1\ldots 4\}$ with corresponding cameras across stereo pairs having consecutive indices (see Figure 1). Estimating the relative pose for calibrated stereo assumes knowledge of each camera's intrinsic calibration matrix $\mathbf{K}_j \in \mathbb{R}^{3\times 3}$ and the pair of rigid motion transformation matrices $\{\mathbf{T}_{k,k+2} = [\mathbf{R}_{k,k+2}|\mathbf{t}_{k,k+2}] : k = 1\ldots 2\}$ defining the geometry for each stereo pair formed by cameras $\mathbf{P}_k$ and $\mathbf{P}_{k+2}$. We take as input a set of homogeneous image measurements $\mathbf{x}_{ij} \in \mathbb{R}^3$, where $i = 1\ldots 3$ denotes the feature index and $j = 1\ldots 5 - i$ is the camera index. This formulation allows us to handle general stereo camera networks, where the estimation of stereo egomotion is the special case where $\mathbf{T}_{1,3} = \mathbf{T}_{2,4}$, $\mathbf{K}_1 = \mathbf{K}_2$ and $\mathbf{K}_3 = \mathbf{K}_4$. In order to enforce incidence constraints among viewing rays, we further define direction vectors of unit length of the form $\hat{\mathbf{x}}_{ij} = \mathsf{N}(\mathbf{K}_j^{-1}\mathbf{x}_{ij})$ where $\mathsf{N}(\cdot) = (\cdot)/\|\cdot\|_2$ normalizes a vector to unit length.

4. Feature scope across stereo pairs

After introducing the notation we shall discuss the different types of input features involved in stereo motion esti-

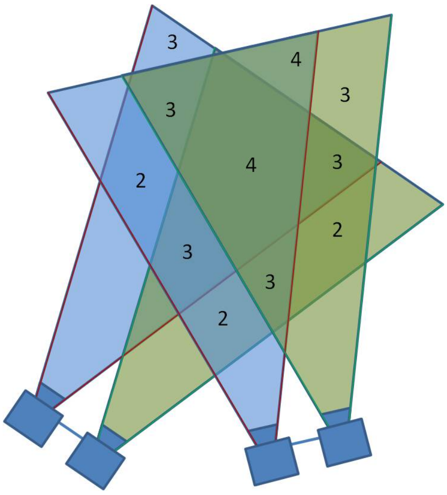

Figure 2. Feature scope in stereo motion estimation as a function of viewing frustums. The distribution of feature regions with common scope within the volume of interest depends on stereo configuration, camera motion and scene geometry.

mation. The input for stereo egomotion estimation consists of $j = 4$ images and a set of feature measurements $\mathbf{x}_{ij}$ corresponding to a 3D point $\mathbf{X}_i$. We denote a feature's *scope* to be the number of images in which a feature is observed. We restrict our interest to features that are observed in both image pairs, since features observed only by one camera pair provide no direct geometrical information on the inter-pair pose. In this way, we can define a general stereo egomotion instance by a 3-tuple $\langle k, l, m \rangle$, where k is the number of features with scope four, l is the number of features of scope three and m is the number of features of scope two. Please see Figure 2 for an illustration of the different scope cardinalities.

For the purpose of this work, we will consider a feature with scope N to provide a total of $N - 1$ geometric constraints for our pose estimation problem. Accordingly, features of different scope can be combined to formulate a fully constrained instance of the relative pose problem for calibrated stereo pairs. In fact, the stereo egomotion problem may be directly formulated as an instance of the well known single camera three point pose problem (P3P) by using a trio of features of scope three (i.e. signature $\langle 0, 3, 0 \rangle$).

Features with a four view scope are found in the intersection of the viewing frustums of the four cameras (please see Figure 2). Given knowledge of the relative orientation and motion among cameras in each stereo pair (i.e. $\mathbf{T}_{1,3}, \mathbf{T}_{2,4}$), these features equate to having a pair of 3D measurements (expressed in each stereo camera coordinate system) through optical triangulation. Moreover, these features constrain the relative pose among camera pairs by three DOF: one DOF in the translation component and two DOF's in the rotational component. In geometric terms, considering a four view feature exclusively, the reference frame of the second camera pair lies on the surface of a sphere with an orientation defined up to a rotation on a tangent plane to the sphere. The sphere center is determined by the position of the triangulation estimate of the four view feature in the first stereo pair. The sphere radius is equal to the magnitude of the triangulation estimate of the four view feature in the second stereo pair. The formulation presented in this work considers an initial four view feature and presents geometric arguments for the fulfillment of the three remaining DOF using additional features of lesser scope.

Features with a three view scope are found in the intersection of the shared viewing frustum of one camera pair and the viewing frustum of a single camera in the other camera pair (please see Figure 2). Such features equate to acquiring an absolute 3D measurement in the first stereo pair and constraining the feature's viewing ray in the second camera pair to pass through that 3D position. Accordingly, features of scope three constrain the relative pose by two DOF, which depending on the context, can be any combination of two rotational and/or translational motions. In our approach, we use one such feature to provide one rotational constraint and one translation constraint.

A feature with a two view scope is found in the shared viewing frustum of two cameras belonging to different stereo pairs (see Figure 2). These features resolve a single DOF in the relative stereo pose problem by constraining a viewing ray of unknown geometry to intersect with a known viewing ray. Accordingly, they define incidence relations when analyzed in conjunction with additional features.

5. Our minimal solution

Our approach estimates the relative pose between two stereo cameras by enforcing the distance and incidence constraints of the 3D viewing rays associated with our image measurements. We define our world reference coordinate system to coincide with that of the first camera in the first stereo pair (i.e. $\mathbf{P}_1$) and strive to determine the rigid transformation $\mathbf{T}_{1,2}$ associated with the corresponding camera $\mathbf{P}_2$ in the second stereo pair. Accordingly, for our reference stereo pair we have

$$\mathbf{P}_1 = \mathbf{K}_1 [\mathbf{I}|0], \tag{1}$$

$$\mathbf{P}_3 = \mathbf{K}_3 [\mathbf{I}|0] \, \mathbf{T}_{1,3} \tag{2}$$

Our approach can be summarized as follows. Initially, we use a four view feature to restrict our translation vector to be on a sphere while setting the orientation as a function of translation up to a single rotational DOF. Next, we use a three view feature to restrict the translation vector to lie on a 3D circle on the sphere, while completely defining orientation as a function of translation. Finally, we use a two

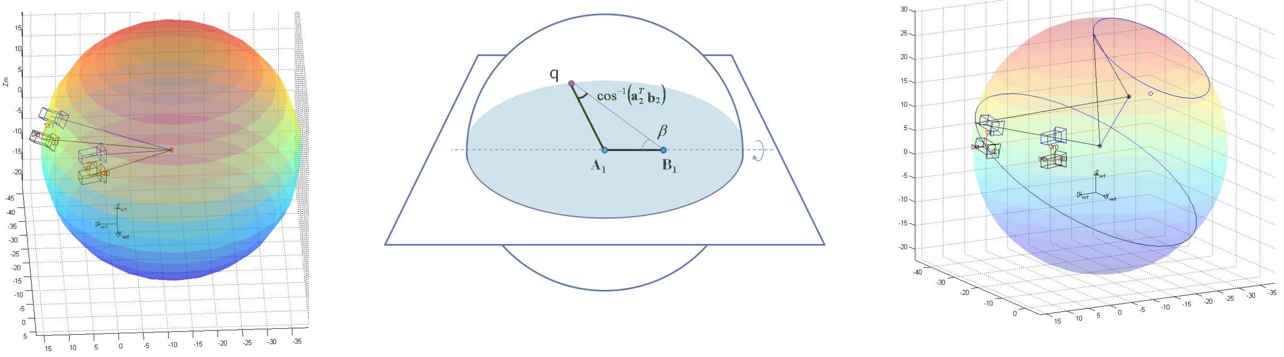

Figure 3. Geometrical constraints from features of scope four and three. At left, a 3D point $\mathbf{X}_1$ is observed by two calibrated stereo pairs enabling triangulation in both pairs. Although the relative motion among pairs is unknown, by enforcing the constant length of the viewing rays, the second camera pair is constrained to lie on a sphere S centered on the triangulation estimate $\mathbf{A}_1$ of the first camera. At middle, given the 3D position of $\mathbf{A}_1$ and $\mathbf{B}_1$, there is a camera position q, located in a great circle of S, which satisfies the incidence constraints among viewing rays $\mathbf{a}_2$ and $\mathbf{b}_2$. At right, there are up to two points q which satisfy incidence constraints for any plane belonging to the pencil defined by the line segment $\overline{\mathbf{A}_1\mathbf{B}_1}$. The generalization to the entire pencil of planes traces a pair of circles orthogonal to $\overline{\mathbf{A}_1\mathbf{B}_1}$.

view feature to determine the translation vector lying on the intersection of the 3D circle and one of the generator lines of a doubly ruled quadric. We will show how for this final step the problem can be reduced to the intersection of two coplanar circles.

5.1. Constraints from a feature of scope four

For the feature $\mathbf{X}_1$ of scope four, we denote its viewing rays across cameras $\mathbf{P}_{1\ldots4}$ as $\mathbf{a}_j = \hat{\mathbf{x}}_{1j}$ for $j = 1\ldots4$ and $\mathbf{A}_j$ as the stereo triangulation estimate of the 3D position of $\mathbf{X}_1$ with respect to the local coordinate system associated with camera $\mathbf{P}_j$. It is possible to locally triangulate $\mathbf{X}_1$ on each stereo pair given knowledge of their calibration and relative motion (i.e. $\mathbf{T}_{1,3}$ and $\mathbf{T}_{2,4}$). While the estimate for $\mathbf{A}_1$ is defined in the coordinate system of $\mathbf{P}_1$, the local estimate of $\mathbf{A}_2$ is defined globally only up to an unknown rigid transformation $\mathbf{T}_{1,2}$. However, from $\mathbf{A}_2$ we identify two invariants. The first is the length of the viewing ray from $\mathbf{P}_2$ to $\mathbf{X}_1$. The second is the angle of that viewing ray with the optical axis of $\mathbf{P}_2$.

From these invariants, the optical center of $\mathbf{P}_2$ (expressed in $\mathbf{P}_1$'s coordinate system) is restricted to lie on a sphere S centered on $\mathbf{A}_1$ with radius $\|\mathbf{A}_2\|$, see Fig. 3. For each candidate point $\mathsf{s} \in \mathsf{S}$ the viewing ray observing $\mathbf{X}_1$ is completely defined and the orientation of the image plane is parameterized by a rotation angle around that viewing ray. In geometric terms, the viewing axis for $\mathbf{P}_2$ can be any of the directions described by a cone with vertex on s, main axis parallel to $\mathbf{A}_1 - \mathsf{s}$ and cone aperture angle equal to $\arccos\left(\mathsf{N}(\mathbf{K}_2^{-1}\mathbf{p}_2) \cdot \mathbf{a}_2\right)$, where $\mathbf{p}_2$ is homogeneous vector for the principal point of camera $\mathbf{P}_2$.

5.2. Constraints from a feature of scope three

For the feature $\mathbf{X}_2$ of scope three, we denote its viewing rays as $\mathbf{b}_j = \hat{\mathbf{x}}_{2j}$ for $j = 1\ldots3$ and $\mathbf{B}_j$ as the stereo triangulation estimate of the 3D position of $\mathbf{X}_2$ with respect to the coordinate system associated with camera $\mathbf{P}_j$. Given the constraints defined by $\mathbf{X}_1$, we can estimate up to two circles $\odot_n$ on S where the optical center of $\mathbf{P}_2$ will be located, see Fig. 3 . Each $\odot_n$ is the intersection of S with a plane π_n orthogonal the line segment $\overline{\mathbf{A}_1\mathbf{B}_1}$. The position of each π_n can be determined by enforcing the following constraint through trigonometric manipulation:

$$\mathsf{N}(\mathbf{A}_1 - \mathsf{q}) \cdot \mathsf{N}(\mathbf{B}_1 - \mathsf{q}) = \mathbf{a}_2 \cdot \mathbf{b}_2, \quad \forall \mathsf{q} \in \odot_n. \tag{3}$$

By examining any arbitrary plane belonging to the pencil defined by $\overline{\mathbf{A}_1\mathbf{B}_1}$, the problem of finding π_n is reduced to describing the geometry of the triangle $\triangle(\mathbf{A}_1\mathbf{B}_1\mathsf{q})$ contained in one of the great circles of S. This triangle will have $\mathbf{A}_1$ and $\mathbf{B}_1$ as two of their vertices and the third vertex q located on the great circle G. A pair of triangles $\triangle(\mathbf{A}_1\mathbf{B}_1\mathsf{q})$ can be determined through the sine rule by noting that $\|A_1 - B_1\|$ is obtained from our triangulation estimates while recalling

$$\angle(\mathbf{A}_1\mathsf{q}\mathbf{B}_1) = \arccos\left(\mathbf{a}_2 \cdot \mathbf{b}_2\right), \tag{4}$$

$$\|\mathbf{A}_1 - \mathsf{q}\| = \|\mathbf{A}_2\|. \tag{5}$$

A pair of circles $\odot_n$ with center $\odot_n^c$ and radius $\odot_n^r$ can be generated by rotating each estimated q around the line segment $\overline{\mathbf{A}_1\mathbf{B}_1}$, see Fig. 3 . Naturally, each $\odot_n$ is contained in a plane orthogonal to $\overline{\mathbf{A}_1\mathbf{B}_1}$. The explicit expressions for $\odot_n^r$ and $\odot_n^c$ are stated in Appendix A. Henceforth, we omit the circle index n from our notation with the understanding that further operations are identical for both possible circles.

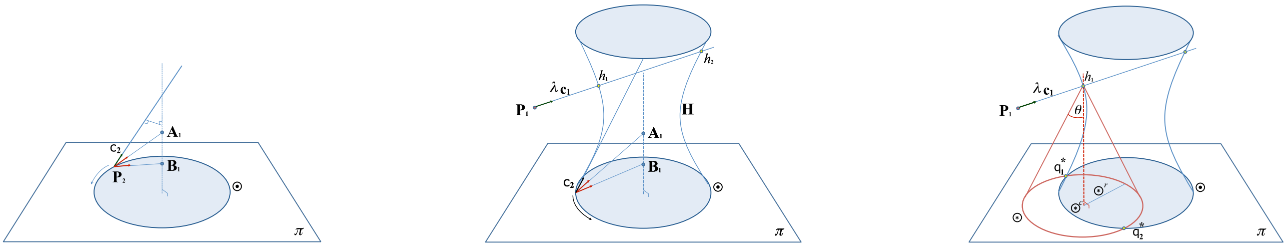

Figure 4. Geometric constraints from a feature of scope two. At left, for every point in $\odot$ the viewing ray c_2 to the unknown position of $\mathbf{X}_3$ is constrained by the known viewing rays to $\mathbf{A}_1$ and $\mathbf{B}_1$. At center, the traversal of c_2 along $\odot$ traces the surface of an hyperboloid of one sheet $\mathbf{H}$. Each intersection h of the viewing ray $\lambda\mathbf{c}_1$ and $\mathbf{H}$ serves as a proxy for $\mathbf{X}_3$ and the hyperboloid generator lines passing through h define valid positions of the optical center of $\mathbf{P}_2$. At right, given that all generator lines must maintain a constant angle θ with the normal to the plane π containing $\odot$, we can define a cone of with vertex at h. Hence, the pair of hyperboloid generator lines passing through h is defined by the intersection of two coplanar circles $\odot$ and $\circledast$.

5.3. Constraints from a feature of scope two

For the feature $\mathbf{X}_3$ of scope two, we denote its viewing rays as $\mathbf{c}_j = \hat{\mathsf{x}}_{3j}$ for $j = 1 \ldots 2$. For any 3D point $\mathsf{q} \in \odot$ the absolute orientation (i.e. expressed in $\mathbf{P}_1$ reference system) of the $\mathbf{c}_2$ is completely defined by the orientation of the vectors $\mathbf{A}_1 - \mathsf{q}$ and $\mathbf{B}_1 - \mathsf{q}$ in conjunction with the known incidence angles among $\mathbf{a}_2, \mathbf{b}_2$ and $\mathbf{c}_2$. That is, given two viewing rays to known 3D positions ($\mathbf{A}_1$ and $\mathbf{B}_1$), the orientation of an additional viewing ray c_2 is resolved by the known inter-ray angle constraints. Moreover, this can be solved in a straightforward manner by the method described in [7].

We note that the viewing ray c_2 can be traversed along all points $\mathsf{q} \in \odot$ and it will maintain a constant relative orientation w.r.t. both the vectors $\odot^c - \mathsf{q}$ and $A_1 - B_1$. Geometrically, this is equivalent to rotating the viewing ray around the axis given by $\overline{\mathbf{A}_1\mathbf{B}_1}$. The obtained set of rays defines an hyperboloid $\mathbf{H}$ of one sheet, which is a doubly ruled quadric surface, see Fig. 4. The hyperbolid can be algebraically represented through a 4×4 quadratic form operating on homogeneous 3D vectors. Moreover, it can be parameterized in terms of the two nearest points and the angle between the 3D lines defined by $L_1(\overline{\mathbf{A}_1\mathbf{B}_1})$ and $L_2(\mathsf{q} + t\mathsf{c}_2)$, where t is a scalar parameter. These properties define a local coordinate system $\mathbf{T}_{\mathbf{H}}$ for the hyperboloid and simplify its parametrization to

$$\mathbf{H} = \mathbf{I}_{4\times 4} + diag(0, 0, -tan^2(\theta), d^2), \qquad (6)$$

where θ is the angle between the line directions and d is the distance between their closest points.

We proceed to determine the 3D intersection of $\mathbf{H}$ with the viewing ray $\lambda\mathbf{c}_1$, where $\lambda \in \mathbb{R}$. The 3D intersection point h is obtained by solving for λ in closed form the following quadratic form

$$\begin{pmatrix} \Omega + \lambda\mathbf{c}_1' \\ 1 \end{pmatrix}^T \mathbf{H} \begin{pmatrix} \Omega + \lambda\mathbf{c}_1' \\ 1 \end{pmatrix} = 0, \qquad (7)$$

where $\Omega = \mathbf{T}_{\mathbf{H}}(0\,0\,0\,1)^T$ and $\mathbf{c}_1' = \mathbf{T}_{\mathbf{H}}\mathbf{c}_1$ are, respectively, the optical center of $\mathbf{P}_1$ and the viewing ray $\mathbf{c}_1$ expressed in the hyperboloid coordinate system. From Equation (7) we can obtain up to two distinct real solutions for λ (see Appendix B for further details). Having estimated λ we can readily determine up to two 3D intersection points $h = \lambda\mathbf{c}_1$ along the viewing direction of $\mathbf{c}_1$.

At this point our geometric constraint analysis has enabled us to estimate a candidate 3D position h for a feature $\mathbf{X}_3$ of scope two. This is notable because the motion $\mathbf{T}_{1,2}$ between the observing cameras is unknown, impeding direct optical triangulation from the image measurements. Moreover, h effectively acts as a proxy for $\mathbf{X}_3$ within our solver and is considered the product of an *indirect* triangulation. A naive approach would be to stop our analysis here and use the estimate of h (along with the estimates for $\mathbf{A}_1$ and $\mathbf{B}_1$) to generate and solve an instance of the well known P3P problem. However, solving such an instance would provide up to four different solutions for each estimate h (discussion on the cardinality of our solution set will be presented in the next section). Instead, we exploit the geometric structure of our problem to obtain a direct geometric estimate with only two solutions for each h.

Having estimated h, the next step is to determine which are the two generator lines of $\mathbf{H}$ that intersect at h, in order to subsequently determine their intersection with $\odot$. Noting that the direction vector of all generator lines has a constant angle θ with the direction vector of the axis defined by $\overline{\mathbf{A}_1\mathbf{B}_1}$, we can abstract our problem as follows. We represent the set of candidates generator lines as a cone with vertex at h, axis parallel to $\overline{\mathbf{A}_1\mathbf{B}_1}$ (i.e. $L_3(h + t\mathsf{N}(\mathbf{A}_1 - \mathbf{B}_1))$ and angle equal to θ. Then, we determine the intersection of this cone with the plane π containing $\odot$. Such intersection yields a second coplanar circle $\circledast$ with center and radius

$$\begin{aligned} \circledast^c &= P(\pi, L_3(h + t\mathsf{N}(\mathbf{A}_1 - \mathbf{B}_1)), \\ \circledast^r &= D(\pi, h)tan(\theta), \end{aligned} \qquad (8)$$

where $P(\cdot, \cdot)$ denotes the intersection between a plane and a line, while $D(\cdot, \cdot)$ denotes Euclidean distance between a plane and a point. The intersection of these circles will yield a pair of points $(\mathsf{q}_1^*, \mathsf{q}_2^*) \in \circledast \bigcap \odot$, corresponding to possible 3D positions of $\mathbf{P}_2$'s optical center. The obtained position of the optical center, along with the absolute orientation of viewing rays for all the features, provides a solution to our relative pose estimation problem.

6. Solution cardinality and degeneracy

The proposed solver will report up to eight possible solutions. More specifically, there are up to two circles $\odot_n$ from which to estimate $\mathbf{H}_n$. These circles depend on the existence of a valid triangle $\triangle(\mathbf{A}_1\mathbf{B}_1\mathsf{q})$ (see Fig. 3) contained in one of the great circles of S. There will exist at least one triangle that satisfies Equation (3), with the existence of a second triangle contingent on $\mathbf{B}_1$ being located within the volume of S, i.e. $\|\mathbf{A}_1 - \mathbf{B}_1\| \leq \|\mathbf{A}_2\|$. Each of the two $\mathbf{H}_n$ can be intersected by $\lambda\mathbf{c}_1$ at up to two 3D points h_λ. For each of these four possible intersections a cone can be defined, from which a pair of solutions can be estimated.

Due to the geometric nature of our solver, there exist three notable 3D feature configurations that are degenerate. The first configuration arises when all 3D features are collinear. This corresponds geometrically to the pair of coplanar circles to be intersected to become concentric, due to the fact that the quadric surface $\mathbf{H}$ degenerates into a cone with the viewing ray $\lambda\mathbf{c}_1$ intersecting at the cone's vertex (see Fig. 5 for an illustration). The second degenerate 3D feature geometry arises when the feature $\mathbf{X}_3$ of scope two is located in the plane π containing $\odot$. For this scenario the hyperbolic surface effectively "flattens" into a single plane, as the generator lines are orthogonal to the axis defined by the line $L_1(\overline{\mathbf{A}_1\mathbf{B}_1})$ (Fig. 5 shows this case). We can still solve for the motion intersecting the viewing ray $\mathbf{c}_1$ and our degenerate quadric. Hence, we can use this 3D feature estimate to generate a P3P instance. The third degeneracy arises when the third viewing ray is one of the hyperboloid generators. Under such circumstances all the points along the viewing ray intersect the hyperboloid and no point can be selected to define the cone of candidate generator lines (see Fig. 5 for an illustration).

For the sake of notation simplicity, we have chosen to describe our solver using a feature $\mathbf{X}_3$ of scope two being observed by cameras $\mathbf{P}_1$ and $\mathbf{P}_2$. This is unnecessarily restrictive, as such feature may be observed by any inter-camera pair combination. Our flexibility is due to the fact that the intra-pair calibration is known. Hence any viewing rays observed by cameras $\mathbf{P}_3$ and $\mathbf{P}_4$ can be expressed in terms of the displaced and rotated equivalent in $\mathbf{P}_1$ and $\mathbf{P}_2$. Accordingly, the variables forming Eq. 7 should reflect the modified geometry, but the rest of the approach remains unchanged. Moreover, if $\mathbf{X}_3$ is observed by camera $\mathbf{P}_4$ the second degenerate condition is now triggered by finding $\mathbf{X}_3$ on the plane parallel to π at an offset equal to the one for the optical center of $\mathbf{P}_4$ traversing $\odot$ to generate the hyperboloid.

7. Experiments

In this section we evaluate our method on synthetic and real data to evaluate the sensitivity of our solver to image measurement noise. The first experiment utilizes feature points on a plane placed in front of both cameras, while the second experiment uses features in a cube volume. Within each of these sets, the features are distributed according to a uniform distribution. We considered random motions of our stereo cameras with 600 by 800 pixel image resolution (i.e. both camera pairs were allowed to move independently but had to overlap in their fields of view). For each test configuration we projected a triplet of 3D features to all cameras and perturbed the image measurement with different levels of Gaussian noise. We performed three measurements to evaluate the quality of our pose estimation. We first evaluate the orientation inaccuracies by measuring the angle of rotation associated with the rotation matrix given by $f_1 = \angle(\hat{\mathbf{R}}\ddot{\mathbf{R}}^T)$, where $\hat{\mathbf{R}}$ denotes our rotation estimate and $\ddot{\mathbf{R}}$ denotes our ground truth. Next, we measure the angle in radians between the estimated and the ground truth translation vectors, by evaluating $f_2 = \arccos\big(\mathsf{N}(\hat{\mathbf{t}}) \cdot \mathsf{N}(\ddot{\mathbf{t}})\big)$, where similarly, $\hat{\mathbf{t}}$ denotes our translation vector estimate and $\ddot{\mathbf{t}}$ denotes our ground truth. Finally, we measure the ratio of lengths among $\hat{\mathbf{t}}$ and $\ddot{\mathbf{t}}$ by computing $f_3 = \min(\|\hat{\mathbf{t}}\|, \|\ddot{\mathbf{t}}\|)/\max(\|\hat{\mathbf{t}}\|, \|\ddot{\mathbf{t}}\|)$. Our results are depicted in Figure 6 where the reported value for each noise level measure is the mean value of a sample of 1000 test configurations.

Starting with very precise translational estimates in the case of noise free estimates, our results as expected depict a graceful degradation in accuracy as a function of image noise. This is expected for a minimal solver given the lack of redundancy. While the solver directly provides the 3D position of the camera center of $\mathbf{P}_2$, the rotation needs to be computed from the viewing rays connecting the estimated camera center and the (now known) 3D position of the three observed features. Such computation is performed using Horn's method [7] for absolute orientation. Accordingly, rotation estimates are dependent on the accuracy of both the translation estimate as well as the (direct and indirect) 3D triangulation estimates of all features.

Experiments on real data were performed by using our solver within a RANSAC framework as part of a SfM pipeline. The stereo capture system consisted of a pair 5MP cameras recording at 15fps. The system was calibrated and lens distortion was corrected. KLT [18] features were tracked at every frame of the video sequence and keyframes stored once sufficient scene flow was detected. SIFT [12]

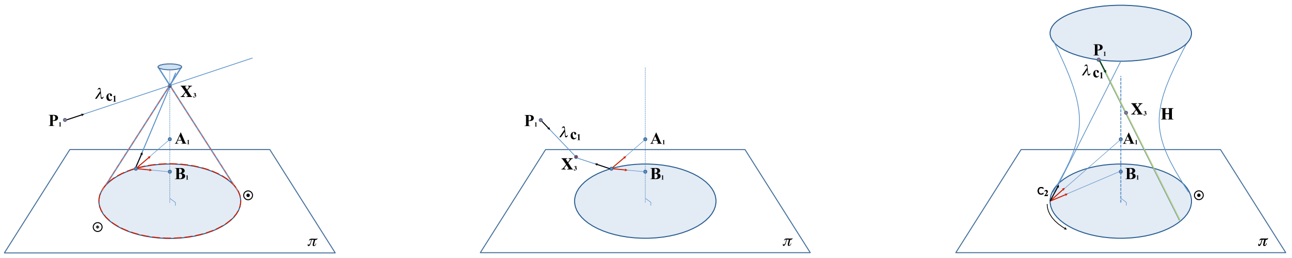

Figure 5. Degeneracy of the solver. At left, three collinear points trace a cone surface and constrain the intersection of this surface with the additional viewing ray to be at the cone's vertex. Accordingly, the intersection for the circles ⊙ and ⊛ is not uniquely defined. At center, when the feature of scope two is coplanar with ⊙, the traced quadric surface degenerates into a plane where the generator lines are not uniquely defined. At right, When the viewing ray of the third feature lies on the the surface of the hyperboloid (e.g. it is a generator line) the intersection of such ray is not uniquely defined.

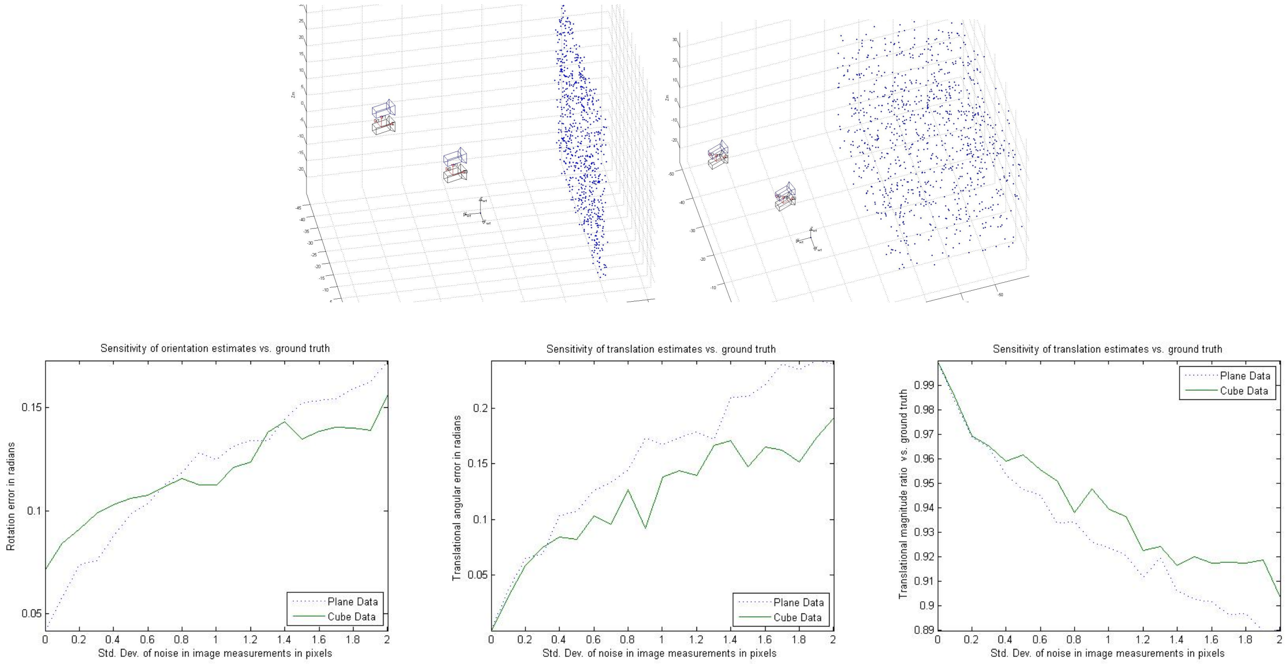

Figure 6. Top: Synthetic test scenarios used in our experimentation. Bottom: Sensitivity plots for applying gaussian perturbations to the set of input image measurements

features were computed for each keyframe and these features were fed into our solver to perform motion estimation among consecutive keyframes. Figure 7 illustrates the scene under observation as well as the obtained sparse 3D structure. Over the 33 keyframes we measure an average reprojection error of 2 pixels in the input 5MP images.

8. Discussion and future work

We introduced a minimal geometric solver for the motion estimation of stereo cameras. The proposed method pushes the state of the art by addressing the set of robotic stereo cameras that have previously been unexplored. The proposed constraints are geometrically meaningful and deliver up to eight solutions for the camera position. Additionally, we identified the degenerate camera motions of our method and performed an experimental evaluation on real and synthetic data. Future work will address the analysis of error propagation within the solver. Also, the inclusion of the solver within a model selective RANSAC framework. In this regard, we believe that robustness against online variations in feature scope is an important issue which still has not been fully addressed within robust estimation.

Acknowledgements. We like to thank SPAWAR, DOE grant DEFG52-08NA28778, NASA grant NNX09AR17G subcontract S100013 and Nvidia for their support.

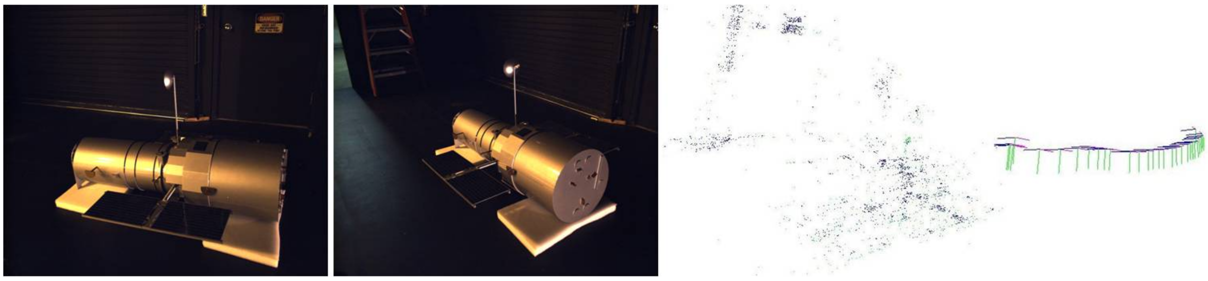

Figure 7. Experiments on captured data. At left, two frames out of a 900 frame video sequence (only left camera shown). At right, reconstructed sparse 3D structure and camera path.

References

[1] R. Bolles and M. Fischler. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981. 1

[2] B. Clipp, J.-M. Frahm, M. Pollefeys, J.-H. Kim, and R. Hartley. Robust 6dof motion estimation for non-overlapping multi-camera systems. In *IEEE Workshop on Applications of Computer Vision*, 2008. 2

[3] B. Clipp, C. Zach, J.-M. Frahm, and M. Pollefeys. A new minimal solution to the relative pose of a calibrated stereo camera with small field of view overlap. In *ICCV*, 2009. 1, 2

[4] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. In *ICCV*, pages 108–115, 2001. 2

[5] R. Haralick, C. Lee, K. Ottenberg, and M. Nolle. Review and analysis of solutions of the three point perspective pose estimation problem. *IJCV*, 13(3):331–356, 1994. 1, 2

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 1

[7] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4(4):629–642, 1987. 5, 6

[8] J. Kim, R. Hartley, J. Frahm, and M. Pollefeys. Visual odometry for non-overlapping views using second-order cone programming. In *ACCV*, pages 353–362, 2007. 2

[9] J.-H. Kim and M. J. Chung. Absolute motion and structure from stereo image sequences without stereo correspondence and analysis of degenerate cases. *Pattern Recognition*, 39(9):1649–1661, 2006. 2

[10] Z. Kukelova, M. Bujnak, and T. Pajdla. Automatic generator of minimal problem solvers. In *ECCV*, 2008. 1

[11] H. Li, R. Hartley, and J. Kim. A linear approach to motion estimation using generalized camera models. In *CVPR*, 2008. 2

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004. 6

[13] C. McGlone, E. Mikhail, and J. Bethel. *Manual of Photogrammetry, 5th Edition*. American Society of Photogrammetry and Remote Sensing, Bethesda, MD, 2004. 1

[14] K. Ni and F. Dellaert. Stereo tracking and three-point/one-point algorithms - a robust approach in visual odometry. In *ICIP*, pages 2777–2780, 2006. 2

[15] D. Nistér. A minimal solution to the generalised 3-point pose problem. In *CVPR*, pages I: 560–567, 2004. 2

[16] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *CVPR*, volume 01, pages 652–659, 2004. 1, 2

[17] R. Pless. Using many cameras as one. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2003. 2

[18] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994. 6

[19] 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. 2

A. Estimation of $\odot_n$

The center $\odot_n^c$ and radius $\odot_n^r$ of each circle is given by

$$\phi = \left(\frac{\|\mathbf{A}_2\| \sin(\angle(\mathbf{A}_1\mathbf{q}\mathbf{B}_1))}{\|\mathbf{A}_1 - \mathbf{B}_1\|} \right), \qquad (9)$$

$$\beta_1 = \pi - \angle(\mathbf{A}_1\mathbf{q}\mathbf{B}_1) - \arcsin(\phi), \qquad (10)$$

$$\beta_2 = \pi - \angle(\mathbf{A}_1\mathbf{q}\mathbf{B}_1) - \left[\frac{\pi}{2} + \arccos(\phi)\right], \qquad (11)$$

$$\odot_n^c = \mathbf{A}_1 + \|\mathbf{A}_2\| \cos(\beta_n) \mathsf{N}(\mathbf{B}_1 - \mathbf{A}_1), \qquad (12)$$

$$\odot_n^r = \|\mathbf{A}_2\| \sin(\beta_n) \qquad (13)$$

B. Intersection of $\Omega + \lambda \mathbf{c}_1'$ and $\mathbf{H}$

The scalar λ from Equation (7) is determined by solving a quadratic equation

$$\alpha\lambda^2 + \gamma\lambda + \xi = 0, \qquad (14)$$

where

$$\alpha = \begin{pmatrix} \mathbf{c}_1' \\ 0 \end{pmatrix}^T \mathbf{H} \begin{pmatrix} \mathbf{c}_1' \\ 0 \end{pmatrix}, \qquad (15)$$

$$\gamma = 2 \begin{pmatrix} \mathbf{c}_1' \\ 0 \end{pmatrix}^T \mathbf{H} \begin{pmatrix} \Omega \\ 1 \end{pmatrix}, \qquad (16)$$

$$\xi = \begin{pmatrix} \Omega \\ 1 \end{pmatrix}^T \mathbf{H} \begin{pmatrix} \Omega \\ 1 \end{pmatrix}. \qquad (17)$$