Structure from Motion with Known Camera Positions

Rodrigo Carceroni   Ankita Kumar   Kostas Daniilidis

Dept. of Computer & Information Science — University of Pennsylvania
3330 Walnut Street, Philadelphia, PA 19104
{carceron,ankitak,kostas}@grasp.cis.upenn.edu

Abstract

The wide availability of GPS sensors is changing the landscape in the applications of structure from motion techniques for localization. In this paper, we study the problem of estimating camera orientations from multiple views, given the positions of the viewpoints in a world coordinate system and a set of point correspondences across the views. Given three or more views, the above problem has a finite number of solutions for three or more point correspondences. Given six or more views, the problem has a finite number of solutions for just two or more points. In the three-view case, we show the necessary and sufficient conditions for the three essential matrices to be consistent with a set of known baselines. We also introduce a method to recover the absolute orientations of three views in world coordinates from their essential matrices. To refine these estimates we perform a least-squares minimization on the group cross product $SO(3) \times SO(3) \times SO(3)$. We report experiments on synthetic data and on data from the ICCV2005 Computer Vision Contest.

1. Introduction

Monocular image sequences have been widely used for localization with respect to a reference frame that usually coincides with one of the recorded frames. A subsequent dense matching can yield a 3D model of the scene. The state of the art is represented among others by the 3D modeling approaches by Pollefeys [15] and Nistér [13] and the recursive algorithm by Soatto [2]. When the application scenario is outdoors and we avoid urban canyons, we can equip a monocular camera with a rigidly attached GPS sensor. The sensor's measurement error bounds can constrain the search for the camera positions in a general structure from motion scheme, or be incorporated as a prior probability in a multiple frames algorithm.

To increase our understanding of the problem, we decided to study the exact problem when the positions of the cameras are known with respect to an Earth coordinate system and the missing information are the unknown orientations of the camera coordinate systems with respect to the GPS coordinate system. In particular, we ask the following questions: (1) is the problem more constrained than general structure from motion (if yes, through which constraints and for which number of frames)? (2) is there a generalization of the essential parameters and what are the necessary and sufficient conditions for their decomposability into the unknown rotations? (3) how can we exploit the fact that the unknowns are elements of a compact group, the cross product $SO(3) \times \ldots \times SO(3)$?

In this paper we focus on the geometry of the problem and do not deal with the important issues of matching and image retrieval, namely, finding the feature tuples with mutual correspondences on which to apply our methods.

There is no particular literature yet for the problem we address, but the setting is the same as in the recent ICCV Contest where, given a database of images with associated GPS positions, the problem was to find the GPS position of novel input images. There are several browsers on the Web that relate geographic information with images— *e.g.*, those by Microsoft Research (wwmx.org), Amazon (maps.a9.com), and several add-ons to Google Maps (Flickrmap and Mappr). We can envision camera systems in the future that will automatically annotate image headers with GPS information (see the Ricoh camera model Pro G3). Most of the literature deals with the retrieval aspect of the problem [17, 22]. Johansson and Cipolla [6] estimate pose based on metrically known building facades.

This paper's main contributions are: (1) a proof that cyclic epipolar constraints are necessary and sufficient (in the general case) for a tuple of image points matched across N views to back-project to a unique point in 3D space; (2) a proven enumeration of the necessary and sufficient conditions for three essential matrices to be consistent with three known viewpoints; (3) algorithms that extract initial values for orientations of three views from their essential matrices and iterate on $SO(3) \times SO(3) \times SO(3)$, so as to minimize a

sum of squared epipolar constraints.

In the second section we present the derivation of the algebraic conditions and the minimal number of frames and point correspondences necessary to solve the N-view problem. In the third section we study the decomposability of essential matrix triples. In the fourth section we show how to recover orientations from essential matrices and in the fifth section we present the minimization over SO(3) × SO(3) × SO(3). Finally, we present experiments and conclusions.

2. General Algebraic Constraints

In this section we examine the general algebraic constraints imposed on the absolute orientations of N views by P point correspondences across these views, in order to determine how many such correspondences are necessary to solve the absolute orientation recovery problem when viewing positions are known. More formally, the problem may be posed as how to recover $\mathbf{R}_i \in$ SO(3), $i = 1, \ldots, N$, the *unknown* orthonormal matrices describing the N viewing orientations on an arbitrarily-chosen global Euclidean reference system, from the following *known* quantities: $\mathbf{t}_i \in \Re^3$, $i = 1, \ldots, N$, the viewing positions in the same global reference system; and $\mathbf{u}_{i,k} \in \mathrm{S}^2$, $i = 1, \ldots, N$, $k = 1, \ldots, P$, the unit vectors describing the directions of P *unknown* scene points $\mathbf{p}_k \in \Re^3$ in local Euclidean reference systems aligned with the N views.

The fundamental relationship between these quantities is the incidence of each 3D point $\mathbf{p}_k$ on each viewing ray defined by a viewing position $\mathbf{t}_i$ and a direction $\mathbf{R}_i \mathbf{u}_{i,k}$: for $i = 1, \ldots, N$, $k = 1, \ldots, P$, $\exists! \lambda_{i,k} \in \Re$ such that

$$\mathbf{p}_k = \mathbf{t}_i + \lambda_{i,k} \mathbf{R}_i \mathbf{u}_{i,k}, \qquad (1)$$

where each scalar $\lambda_{i,k}$ that satisfies Eq. (1) is the (Euclidean) depth of point $\mathbf{p}_k$ in view i. This relationship can be expressed in a way that does not involve the coordinates of any 3D scene point $\mathbf{p}_k$: for $k = 1, \ldots, P$, $\exists \lambda_{i,k} \in \Re$, $i = 1, \ldots, N$, such that for $i, j = 1, \ldots, N$,

$$\lambda_{i,k} \mathbf{R}_i \mathbf{u}_{i,k} - \lambda_{j,k} \mathbf{R}_j \mathbf{u}_{j,k} = \mathbf{b}_{i,j}, \qquad (2)$$

where $\mathbf{b}_{i,j} \triangleq \mathbf{t}_j - \mathbf{t}_i$. The depths $\lambda_{i,k}$ that satisfy Eq. (2) are unique if, and only if, the viewpoints $\mathbf{t}_i$ are not all collinear.

Any algebraic constraints among the geometric parameters of multiple views can be derived directly from Eq. (2) [10]. Moreover, when the baselines $\mathbf{b}_{i,j}$ are all known *a priori*, Eq. (2) yields $2N - 3$ independent polynomial constraints [16] on the $3N$ free parameters of matrices $\mathbf{R}_i$. These two facts lead to the following:

Theorem 1 (Number of Correspondences Necessary)
From the knowledge of N absolute viewing positions and projections on these views of P scene points that are in general 3D positions with respect to them, it is only

possible to recover the N views' absolute orientations if

$$N \geq 3, \quad P \geq \left\lceil \frac{3N}{2N - 3} \right\rceil. \qquad (3)$$

Proof. $N \geq 3$ views are needed because with two views the problem is only solvable up to a simultaneous rotation of both views about the baseline. The number of correspondences needed when $N \geq 3$ follows from Lemma 1.

Lemma 1 (Only $2N - 3$ Independent Constraints) *Let* $\mathcal{V} \triangleq \{\mathbf{t}_i, i = 1, \ldots, N\}$ *be any set of $N \geq 3$ viewing positions in $\Re^3$. For every k such that $\mathbf{p}_k$ is not coplanar with any three points in $\mathcal{V}$, Eq. (2) holds if, and only if, the following constraints are satisfied:*

$$\mathbf{u}_1^T \mathbf{E}_{1,2} \mathbf{u}_2 = 0, \qquad (4)$$
$$\mathbf{u}_2^T \mathbf{E}_{2,i} \mathbf{u}_i = 0, \quad i = 3, \ldots, N, \qquad (5)$$
$$\mathbf{u}_i^T \mathbf{E}_{i,1} \mathbf{u}_1 = 0, \quad i = 3, \ldots, N, \qquad (6)$$

where $\mathbf{E}_{i,j} \triangleq \mathbf{R}_i^T \widehat{\mathbf{b}_{i,j}} \mathbf{R}_j$ is an essential matrix, and $\widehat{\mathbf{v}}$ denotes the skew-symmetric matrix associated to $\mathbf{v} \in \Re^3$.

Long Quan [16] has proven Lemma 1 by showing that with points in general position, the *ideal* [3] of the bilinear constraints above is equal to the ideal of all trilinear constraints. In Appendix A we provide a simpler direct proof that Eqs. (4)-(6) imply Eq. (2) in the general case.

Corolary 1 (Irrelevance of baseline lengths) *Because Eqs. (4)-(6) hold even if their l.h.s. is scaled by an arbitrary factor, a direct consequence of Lemma 1 is that only the orientations of the known baselines provide constraints on the unknown viewing orientations.*

For this reason, we will assume from this point on that all baselines have unit norm and, accordingly, that all essential matrices have Frobenius norm equal to $\sqrt{2}$.

3. Stratification of Essential Matrix Triples

It is of course known that the algebraic constraints in Eqs. (4)-(6) can be used to solve the general structure from motion problem, *i.e.*, the problem of simultaneously recovering *relative* viewing positions and orientations from point correspondences. In particular, five point correspondences across a pair of views are generally necessary and sufficient to recover their essential matrix [11]. Straightforward algebraic manipulation of such matrix then yields view-to-view translation (up to a scalar factor) and rotation.

Thus, a fundamental question is: does the knowledge of viewing positions provide *additional* constraints on essential matrices that may be used to recover these matrices with *less* point correspondences than in general structure from motion? In this section we provide an affirmative answer

to the question above, by enumerating minimal sets of constraints that three arbitrary essential matrices have to satisfy in order to be mutually consistent and in order to be consistent with three known baselines:

Definition 1 (Mutual Consistency of Essential Matrices)
Three distinct essential matrices $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ are mutually consistent if, and only if, $\exists\ \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3 \in$ SO(3), $\mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C \in S^2$ such that

$$\mathbf{E}_A = \mathbf{R}_1^T \widehat{\mathbf{b}_A} \mathbf{R}_2, \ \ \mathbf{E}_B = \mathbf{R}_2^T \widehat{\mathbf{b}_B} \mathbf{R}_3, \ \ \mathbf{E}_C = \mathbf{R}_3^T \widehat{\mathbf{b}_C} \mathbf{R}_1 \quad (7)$$

and that $\mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C$ are linearly dependent.

Definition 2 (Consistency with Known Baselines)
Three essential matrices $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ are consistent with a given triple of distinct but linearly dependent baselines, $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1} \in S^2$ if, and only if, $\exists\ \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3 \in$ SO(3) such that Eq. (7) is satisfied with $\mathbf{b}_A = \mathbf{b}_{1,2}, \mathbf{b}_B = \mathbf{b}_{2,3}, \mathbf{b}_C = \mathbf{b}_{3,1}$.

Definitions 1 and 2 induce a stratification of the set of all triples of essential matrices. A real non-zero 3×3 matrix is an essential matrix if, and only if, it has rank two and its two non-zero singular values are identical [11]. The set of matrices that satisfy these constraints and have Frobenius norm equal to $\sqrt{2}$ is five-dimensional. Hence the set of all essential matrix triples has fifteen degrees of freedom. We will demonstrate (in Theorem 2, below) that Definition 1 imposes four independent constraints on such fifteen-dimensional set and hence the subset of essential matrix triples that satisfy it is eleven-dimensional. We will also demonstrate (in Theorem 3, below) that Definition 2 restricts another two degrees of freedom on the set of triples that are already mutually consistent. Hence, it is satisfied only by a nine-dimensional set of essential matrix triples, which agrees with our counting of d.o.f. in Section 2.

Theorem 2 (Mutual Consistency Constraints) *Essential matrices $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ satisfy Definition 1 if, and only if, $\exists\ \mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C \in$ SO(3), $\mathbf{v}_A, \mathbf{v}_B, \mathbf{v}_C \in S^2$ such that*

$$\mathbf{E}_A = \mathbf{R}_A \widehat{\mathbf{v}_A}, \ \ \mathbf{E}_B = \mathbf{R}_B \widehat{\mathbf{v}_B}, \ \ \mathbf{E}_C = \mathbf{R}_C \widehat{\mathbf{v}_C}, \qquad (8)$$

$$\mathbf{R}_A \mathbf{R}_B \mathbf{R}_C = \mathbf{I}, \qquad (9)$$

$$\mathbf{v}_A^T \left(\mathbf{R}_B \mathbf{v}_B \times \mathbf{R}_A^T \mathbf{v}_C \right) = 0. \qquad (10)$$

Proof. If arbitrary essential matrices $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ can be decomposed as in Eq. (7) then they can be decomposed as in Eq. (8) with

$$\mathbf{R}_A = \mathbf{R}_1^T \mathbf{R}_2, \quad \mathbf{R}_B = \mathbf{R}_2^T \mathbf{R}_3, \quad \mathbf{R}_C = \mathbf{R}_3^T \mathbf{R}_1, \quad (11)$$

$$\mathbf{v}_A = \mathbf{R}_2^T \mathbf{b}_A, \quad \mathbf{v}_B = \mathbf{R}_3^T \mathbf{b}_B, \quad \mathbf{v}_C = \mathbf{R}_1^T \mathbf{b}_C. \quad (12)$$

By substituting Eqs. (11)-(12) on the l.h.s. of Eqs. (9)-(10), it is straightforward to verify that if Definition 1 is satisfied, so are Eqs. (8)-(10).

Conversely, if Eqs. (8)-(9) hold then $\exists\ \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3 \in$ SO(3) such that $\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C$ can decomposed as in Eq. (11). For every such $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \exists\ \mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C \in S^2$ that satisfy Eq. (12). Substituting Eqs. (11)-(12) into Eq. (8) yields Eq. (7). Substituting Eqs. (11)-(12) into Eq. (10) yields the fact that $\mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C$ are linearly dependent. $\quad \square$

Since any three essential matrices can be decomposed as in Eq. (8), what Theorem 2 means is that in order for a triple of essential matrices to be mutually consistent, it is necessary and sufficient that they satisfy four independent constraints: the three constraints in Eq. (9) and the single constraint in Eq. (10). Thus, the set of all mutually consistent essential matrix triples is indeed eleven-dimensional.

Theorem 3 (Known Baseline Constraints) *Let $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ be any mutually consistent triple of essential matrices with Frobenius norms $\sqrt{2}$. Let $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1} \in S^2$ be any triple of distinct but linearly dependent baselines. If $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ are consistent with $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1}$ (according to Definition 2) then the first two singular values of the product matrix $\mathbf{P}_{ABC} \triangleq \mathbf{E}_A \mathbf{E}_B \mathbf{E}_C$ are $|\cos(\mathbf{b}_{1,2}, \mathbf{b}_{2,3})|$ and $|\cos(\mathbf{b}_{2,3}, \mathbf{b}_{3,1})|$. Moreover, if $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1}$ are edges of an acute or right triangle, this constraint on $\mathbf{P}_{ABC}$ is also sufficient to guarantee that Definition 2 is satisfied. If $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1}$ are edges of an obtuse triangle then the additional constraint that the cosine between the left and right null-spaces of $\mathbf{P}_{ABC}$ must be equal in absolute value to $|\cos(\mathbf{b}_{3,1}, \mathbf{b}_{1,2})|$ is needed to guarantee consistency of $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ with respect to $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1}$.*

Proof. Since exchanging the signs of the baselines does not affect Eqs. (4)-(6) we assume without loss of generality that $\mathbf{b}_{3,1}$ is obtained by scaling $\mathbf{b}_{1,2}$ and $\mathbf{b}_{2,3}$ with *negative* factors and adding the results.

Now, because essential matrices $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ satisfy Definition 1 by hypothesis,

$$\mathbf{P}_{ABC} = \mathbf{R}_1^T \mathbf{S}_{ABC} \mathbf{R}_1, \quad \text{where:} \qquad (13)$$

$$\mathbf{S}_{ABC} \triangleq \widehat{\mathbf{b}_A} \widehat{\mathbf{b}_B} \widehat{\mathbf{b}_C}, \qquad (14)$$

and $\mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C$ have unit norm and are linearly dependent. These properties of $\mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C$ imply that

$$\mathbf{S}_{ABC} = \mathbf{U}_S \Sigma \mathbf{V}_S^T, \quad \text{where:} \qquad (15)$$

$$\Sigma \triangleq \text{diagonal}\left(\mathbf{b}_A^T \mathbf{b}_B, \ \mathbf{b}_B^T \mathbf{b}_C, \ 0 \right), \qquad (16)$$

$$\mathbf{U}_S \triangleq [\mathbf{w}_{AC} \quad \mathbf{b}_A \times \mathbf{w}_{AC} \quad \mathbf{b}_A], \qquad (17)$$

$$\mathbf{V}_S \triangleq [\mathbf{b}_C \times \mathbf{w}_{AC} \quad -\mathbf{w}_{AC} \quad \mathbf{b}_C], \qquad (18)$$

$$\mathbf{w}_{AC} \triangleq (\mathbf{b}_A \times \mathbf{b}_C) \ / \ \|\mathbf{b}_A \times \mathbf{b}_C\|. \qquad (19)$$

We omit the tedious algebraic manipulations involved in the derivation of the decomposition above — its validity can be checked with symbolic linear algebra software.

Note that when the elements in the diagonal of Σ are all distinct, the decomposition in Eq. (15) is the singular value decomposition of $\mathbf{S}_{ABC}$, up to permutation of the two non-null singular values (with corresponding permutations of the columns of both $\mathbf{U}_S$ and $\mathbf{V}_S$) and to any sign inversions of the singular values (with corresponding sign inversions on the columns of either $\mathbf{U}_S$ or $\mathbf{V}_S$). In the case where any two singular values of $\mathbf{S}_{ABC}$ are identical, its singular value decomposition is ambiguous, but nonetheless some of the infinite forms in which it may be written are equivalent to Eq. (15), up to the transformations mentioned above.

Eq. (13) means that matrices $\mathbf{P}_{ABC}$ and $\mathbf{S}_{ABC}$ are *orthogonally similar*, which implies that they have the same singular values. Hence, $\exists\, \mathbf{U}_P, \mathbf{V}_P \in \mathrm{SO}(3)$ such that

$$\mathbf{P}_{ABC} \;=\; \mathbf{U}_P\, \Sigma\, \mathbf{V}_P^T. \qquad (20)$$

This shows that the condition that the two first singular values of $\mathbf{P}_{ABC}$ must be $|\cos(\mathbf{b}_{1,2}, \mathbf{b}_{2,3})|$, $|\cos(\mathbf{b}_{2,3}, \mathbf{b}_{3,1})|$ is *necessary* for Definition 2 to be satisfied.

To see that the same condition is also *sufficient* for consistency with known baselines that correspond to acute or right triangles, remember that $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ satisfy Definition 1 by hypothesis. Let $\mathbf{b}_A, \mathbf{b}_B, \mathbf{b}_C \in \mathrm{S}^2$ be any three unit vectors that satisfy Eq. (7). If the condition on the singular values of $\mathbf{P}_{ABC}$ is satisfied, Eq. (13) and Eqs. (15)-(20) imply that $|\cos(\mathbf{b}_A, \mathbf{b}_B)| = |\cos(\mathbf{b}_{1,2}, \mathbf{b}_{2,3})|$, and $|\cos(\mathbf{b}_B, \mathbf{b}_C)| = |\cos(\mathbf{b}_{2,3}, \mathbf{b}_{3,1})|$. Now, knowledge of two absolute cosine values for any acute or right triangle uniquely defines the triangle's geometry up to a rigid transformation and a uniform scaling factor. Hence, $\exists\, \mathbf{R} \in \mathrm{SO}(3)$ such that

$$[\mathbf{b}_A\ \mathbf{b}_B\ \mathbf{b}_C] \;=\; \mathbf{R}\ [\mathbf{b}_{1,2}\ \mathbf{b}_{2,3}\ \mathbf{b}_{3,1}]. \qquad (21)$$

Substituting Eq. (21) back into Eq. (7) yields

$$\mathbf{E}_A = \tilde{\mathbf{R}}_1^T \widehat{\mathbf{b}_{1,2}} \tilde{\mathbf{R}}_2,\ \ \mathbf{E}_B = \tilde{\mathbf{R}}_2^T \widehat{\mathbf{b}_{2,3}} \tilde{\mathbf{R}}_3,\ \ \mathbf{E}_C = \tilde{\mathbf{R}}_3^T \widehat{\mathbf{b}_{3,1}} \tilde{\mathbf{R}}_1,\ \ (22)$$

$$\tilde{\mathbf{R}}_1 \triangleq \mathbf{R}^T \mathbf{R}_1, \qquad \tilde{\mathbf{R}}_2 \triangleq \mathbf{R}^T \mathbf{R}_2, \qquad \tilde{\mathbf{R}}_3 \triangleq \mathbf{R}^T \mathbf{R}_3, \qquad (23)$$

which implies that Definition 2 is satisfied.

For baselines that correspond to an obtuse triangle, however, the knowledge of two absolute cosines leaves an additional two-way ambiguity in the triangle's geometry: it is in general impossible to determine if the obtuse angle is the one that has the larger known absolute cosine or the one whose absolute cosine is unknown (if the two known absolute cosines are identical, then the obtuse angle is necessarily the one with unknown absolute cosine). In this case, an additional constraint must be satisfied by matrices $\mathbf{U}_P$ and $\mathbf{V}_P$ in order to guarantee that Definition 2 holds.

To obtain such constraint, we substitute Eq. (15) and Eq. (20) into Eq. (13), which yields

$$\mathbf{U}_P\, \Sigma\, \mathbf{V}_P^T = \mathbf{R}_1^T\, \mathbf{U}_S\, \Sigma\, \mathbf{V}_S^T\, \mathbf{R}_1. \qquad (24)$$

A necessary and sufficient condition in order to $\exists\, \mathbf{R}_1 \in \mathrm{SO}(3)$ such that Eq. (24) holds is that

$$\mathbf{V}_P^T\, \mathbf{U}_P = \mathbf{V}_S^T\, \mathbf{U}_S. \qquad (25)$$

Substituting Eqs. (17)-(19) on Eq. (25) and simplifying the resulting expression, we get

$$\mathbf{V}_P^T\, \mathbf{U}_P = \begin{bmatrix} 0 & \cos(\mathbf{b}_C, \mathbf{b}_A) & \sin(\mathbf{b}_C, \mathbf{b}_A) \\ -1 & 0 & 0 \\ 0 & -\sin(\mathbf{b}_C, \mathbf{b}_A) & \cos(\mathbf{b}_C, \mathbf{b}_A) \end{bmatrix}. \qquad (26)$$

Eq. (26) shows that given the knowledge of Σ, matrices $\mathbf{U}_P$ and $\mathbf{V}_P$ do indeed contain very little extra information about consistency with respect to known baselines. Nonetheless, Eq. (26) implies that when the baselines correspond to an obtuse triangle, the extra constraint that the cosine between the left and right null spaces of $\mathbf{P}_{ABC}$ must be equal in absolute value to $|\cos(\mathbf{b}_{3,1}, \mathbf{b}_{1,2})|$ is exactly what is needed to eliminate the two-way ambiguity in the triangle's geometry. $\qquad\square$

Since the two constraints on the singular values of $\mathbf{P}_{ABC}$ guarantee consistency with respect to known baselines up to a finite ambiguity in the worst case, the set of essential matrix triples that are consistent with respect to any *particular* triple of know baselines is indeed nine-dimensional.

4. Linear Recovery of Orientations

Importantly, the proof that we presented for Theorem 3 is constructive, in the sense that the decomposition defined in Eqs. (13)-(19) can be used to recover the absolute orientations of three views from their essential matrices and a set of three known baselines, even if each one of these three essential matrices is estimated independently, from correspondences across only one of the three pairs of views.

More specifically, to compute viewing orientations from any triple of independently-estimated essential matrices $\mathbf{E}_A, \mathbf{E}_B, \mathbf{E}_C$ and from known baselines $\mathbf{b}_{1,2}, \mathbf{b}_{2,3}, \mathbf{b}_{3,1}$, we start by: (1) enforcing the constraints in Eqs. (9)-(10), (2) computing matrices $\mathbf{U}_P$ and $\mathbf{V}_P$ through a singular value decomposition of the product matrix $\mathbf{E}_A \mathbf{E}_B \mathbf{E}_C$, and (3) computing matrices $\mathbf{U}_S$ and $\mathbf{V}_S$ according to the expressions in Eqs. (17)-(19). Then, we perform the (possible) permutation and the sign-flips needed in the columns of $\mathbf{U}_P$ and $\mathbf{V}_P$, in order for Eq. (25) to be satisfied.

After this is done, Eq. (24) must also be satisfied, from which we get the following six constraints on $\mathbf{R}_1$:

$$[\mathbf{U}_S\ \mathbf{V}_S]^T \mathbf{R}_1 = [\mathbf{U}_P\ \mathbf{V}_P]^T. \qquad (27)$$

We compute the orthonormal $\mathbf{R}_1$ that minimizes the least-squares residuals between the l.h.s. and the r.h.s. of Eq. (27) by: (1) performing a singular value decomposition of the 3×3 matrix $[\mathbf{U}_S\ \mathbf{V}_S][\mathbf{U}_P\ \mathbf{V}_P]^T$, (2) replacing the computed matrix of singular values with the identity matrix, (3)

multiplying the computed left and right matrices of singular vectors to obtain $\mathbf{R}_1$.

By applying analogous sequences of steps to the product matrices $\mathbf{E}_B \mathbf{E}_C \mathbf{E}_A$ and $\mathbf{E}_C \mathbf{E}_A \mathbf{E}_B$ we recover the orientations $\mathbf{R}_2$ and $\mathbf{R}_3$, respectively. In practice, when eight or more correspondence across three views are available, essential matrices can be estimated linearly [5]. Hence, in such cases the procedure outlined above allows recovery of viewing orientations with a fixed number of linear steps.

5. Minimization on SO(3) × SO(3) × SO(3)

The method described in Section 4 is sufficient to solve the absolute viewing orientation problem in the ideal, noise-free case, from at least five point correspondences across at least three views. However, under practical conditions in which image measurements are noisy, direct minimization of the general algebraic constraints in Eqs. (4)-(6) usually yields more accurate estimates of viewing orientations — see Section 6 for some empirical evidence of this. In this section we address the following optimization problem

$$\text{find arg } \min_{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3} \mathcal{E} \triangleq \sum_{i=1}^{3} \sum_{k=1}^{P} (\epsilon_{i,j,k})^2, \quad (28)$$

$$\epsilon_{i,j,k} \triangleq \mathbf{u}_{i,k}^T \mathbf{R}_i^T \widehat{\mathbf{b}_{i,j}} \mathbf{R}_j \mathbf{u}_{j,k}, \quad j = i \bmod 3 + 1. \quad (29)$$

The problem above involves optimization on a nine-dimensional manifold (SO(3) × SO(3) × SO(3)) that is not isomorphic to the vector space $\Re^9$ [19]. Such problems are often solved by assigning a unique global parameterization to the manifold and then, from a given initial point on the manifold, computing successive refinements by taking steps along the manifold's tangent planes and then projecting points on these tangent planes back to the manifold. Here we adopt a more stable and efficient approach suggested by Taylor and Kriegman [21]: rather than using a unique global parameterization, we cover the manifold with a set of local parameterizations that are *individually* diffeomorphic to a minimum-dimensional Euclidean space. Such approach allows optimization to be performed *directly* on the manifold, with no need for re-projection. More specifically, we extend the solution that Taylor and Kriegman proposed for minimization on SO(3) [21] to the nine-dimensional manifold SO(3) × SO(3) × SO(3).

In order to do so, we parameterize this manifold in the neighborhood of any fixed point $\langle \mathbf{R}_1^{(0)}, \mathbf{R}_2^{(0)}, \mathbf{R}_3^{(0)} \rangle$ as the following function of parameter vector $\Omega \triangleq \left[\omega_1^T \, \omega_2^T \, \omega_3^T \right]^T, \omega_i \in \Re^3, i = 1, \ldots, 3$:

$$\mathbf{R}_i = \mathbf{R}_i^{(0)} \exp(\widehat{\omega}_i), \quad i = 1, \ldots, 3. \quad (30)$$

where $\exp(\cdot)$ is the matrix exponential operator [4]. Such parameterization is diffeomorphic to $\Re^9$ within the compact

neighborhood defined by $\|\omega_i\| < \pi, \; i = 1, \ldots, 3$. Moreover, the derivatives of the unknown viewing orientations with respect to any free parameters ϕ, φ in Ω are

$$\frac{\partial \mathbf{R}_i}{\partial \phi} = \mathbf{R}_i^{(0)} \widehat{\mathbf{x}_\phi}, \quad \frac{\partial^2 \mathbf{R}_i}{\partial \phi \, \partial \varphi} = \mathbf{R}_i^{(0)} \frac{\widehat{\mathbf{x}_\phi} \, \widehat{\mathbf{x}_\varphi} + \widehat{\mathbf{x}_\varphi} \, \widehat{\mathbf{x}_\phi}}{2}, \quad (31)$$

where $\mathbf{x}_\phi$ is either $[1 \, 0 \, 0]^T$, or $[0 \, 1 \, 0]^T$, or $[0 \, 0 \, 1]^T$, respectively if ϕ is the first, or second, or third element of ω_i.

To optimize the error metric in Eq. (28) we compute its gradient, $\mathbf{g}$, and its Hessian, $\mathbf{H}$, using the chain rule and Eq. (31). More specifically, in order to refine a given set of estimated viewing orientations we use values of $\mathbf{R}_i$, $\mathbf{g}$ and $\mathbf{H}$ computed according to Eqs. (30)-(31) within a standard iterative quadratic optimization algorithm (*i.e.*, an unconstrained optimization method similar to Levenberg-Marquardt's algorithm).

6. Experiments

The methods introduced in Sections 4 and 5 assume that the centers of projection of all three cameras are known exactly. In such ideal scenario these methods (when combined) should yield more accurate estimates of absolute viewing orientations than traditional structure from motion, because they enforce the extra constraints enumerated in Section 3. In practice, however, GPS measurements are noisy. The main goal of the experiments presented in this section is to determine the levels of noise in image and GPS measurements for which one should use the methodology proposed in this work.

More specifically, we compare the accuracy and efficiency of four alternatives for finding absolute viewing orientations from multiple point correspondences across three views and GPS measurements: (1) traditional structure-from-motion (SfM), represented by the eight-point algorithm [5], (2) SfM followed by the method proposed in Section 4 to enforce consistency with *known baselines linearly* (SfM-KB-L), (3) SfM-KB-L followed by the method proposed in Section 5 to enforce consistency with *known baselines non-linearly* (SfM-KB-NL), and (4) a RANSAC-based [12] method (Algorithm 1) that samples the set of possible viewing orientations of one camera, uses the matched features in a preemptive way to filter out bad orientation hypotheses, and then applies the non-linear optimization described in Section 5 to refine the remaining hypotheses (RANSAC-KB-NL).

We evaluate the methods enumerated above on two datasets: a synthetic dataset composed of 100 instances in which features and viewpoints are generated randomly and then perturbed with controlled amounts of noise, and a real dataset from the *ICCV2005 Computer Vision Contest* [20].

Synthetic dataset. For each synthetic scene generated, viewpoints $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ were independently drawn from uni-

Algorithm 1 RANSAC-based method to compute absolute orientations of three views (RANSAC-KB-NL).

Require: knowledge of 3 viewing positions;
Require: $P > 3$ point correspondences across 3 views;
1: **for** each $\mathbf{R}_1 \in$ uniform-sampling (SO(3)) **do**
2:     choose 3 among P point correspondences;
3:     $\mathbf{R}_2, \mathbf{R}_3 \leftarrow$ roots (4th-degree-polynomials) [14];
4:     use other $P-3$ correspondences preemptively [12] to discard $[\mathbf{R}_1\,\mathbf{R}_2\,\mathbf{R}_3]$ hypotheses with little support;
5: **end for**
6: **for** 32 hypotheses with smallest Sampson errors **do**
7:     refine $[\mathbf{R}_1\,\mathbf{R}_2\,\mathbf{R}_3]$ as in Section 5;
8: **end for**
9: choose $[\mathbf{R}_1\,\mathbf{R}_2\,\mathbf{R}_3]$ with smallest Sampson error.

---

form distributions in the following cuboids:

$$\begin{aligned}
\mathbf{t}_1 &\in [0 < x < 15, \quad 0 < y < 5, \quad 0 < z < 1], \\
\mathbf{t}_2 &\in [0 < x < 5, \quad 10 < y < 15, \quad 0 < z < 1], \\
\mathbf{t}_3 &\in [10 < x < 15, \ 10 < y < 15, \ 0 < z < 1],
\end{aligned}$$

and $P = 30$ features were independently drawn from a uniform distribution in the following cuboid:

$$\mathbf{p}_k \in [0 < x < 15, \ 30 < y < 45, \ 0 < z < 10],$$

$k = 1, \ldots, P$. For each given value of experimental parameter α (synthetic feature noise), the direction of each feature in each view was then perturbed by a α-degree rotation about an axis independently drawn from a uniform distribution on S^2. Moreover, for each given value of experimental parameter δ (synthetic viewpoint uncertainty), the values of $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ fed as input to the methods that require knowledge of viewing positions were perturbed by translations independently drawn from a uniform distribution on the square $[-\delta/2 < x < \delta/2, \ -\delta/2 < y < \delta/2, \ z = 0]$.

Figure 1 shows the distribution of the errors in the orientations[1] estimated by each of the four techniques in study for different values of the synthetic feature noise (α) and zero synthetic viewpoint uncertainty.

In such ideal scenario the *median* error of SfM-KB-NL is significantly smaller than that of traditional SfM (as expected). However, in a few instances the non-linear optimization diverges, which results in SfM-KB-NL errors of more than ten degrees. RANSAC-KB-NL, on the other hand, does not suffer from this shortcoming, while still yielding a median accuracy similar to that of SfM-KB-NL. The trade-off is that RANSAC-KB-NL is significantly slower than the other alternatives. Average execution times

---

[1]We define the error of each estimated triple of orientations as the maximum among the angles (in degrees) of the three rotations that must be applied to transform each estimated orientation into the corresponding actual orientation. To compute this metric for SfM results, which are only relative (not absolute) orientations, we express actual orientations in the coordinate system of one of the three input views.
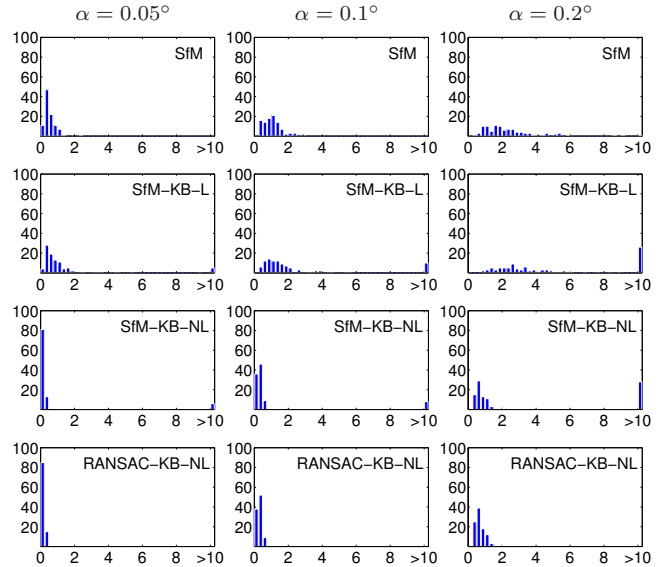

Figure 1. Histograms of errors in orientations[1] estimated by applying SfM, SfM-KB-L, SfM-KB-NL and RANSAC-KB-NL to 100 synthetic scenes, with $\delta = 0$. Each bar (except those with the label ">10") is the percentage of errors within a $0.25°$ interval.

per scene on a Pentium4 3.4GHz 1GB RAM running Matlab 7.0 over Windows XP were: SfM, 0.02s; SfM-KB-L, 0.02s; SfM-KB-NL, 1s; RANSAC-KB-NL, 110s.

More realistic scenarios are those in which there is uncertainty in the coordinates of the cameras' centers of projection. Figures 2 and 3 show the error distributions produced by the three techniques that require knowledge of viewing positions, when there is viewpoint uncertainty. Each one of these figures corresponds to a different value of α. They should thus be compared against the graphs with corresponding α values in the top row of Figure 1.

Through such comparison, it becomes clear that for any given level of noise in image features, there is a level of uncertainty in GPS measurements for which the accuracy of the methods proposed in this paper is roughly equivalent to that of traditional SfM. For instance, by comparing the last row of Figure 2 against the middle graph in the top row of Figure 1, we verify that with feature noise $\alpha = 0.1°$ this "break even" point between RANSAC-KB-NL and traditional SfM is somewhere near $\delta = 0.25$. For levels of GPS uncertainty lower than that, RANSAC-KB-NL will generally yield more accurate results than traditional SfM.

Real dataset. In order to provide some evidence that the methods based on prior knowledge about viewing positions are useful in challenging real-life scenarios, we applied the most accurate of them (RANSAC-KB-NL) to a triple of images from the final dataset of the *ICCV2005 Computer Vision Contest* [20]. To do so, we *manually* selected and established correspondences among $P = 37$ features that are visible across these three images, as shown in Figure 4.
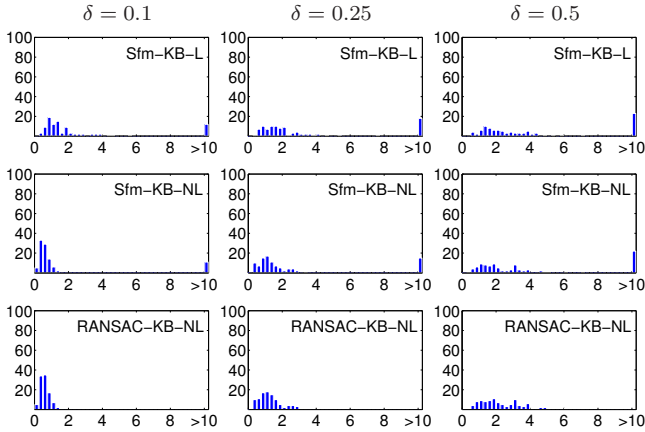
Figure 2. Histograms of errors in orientations[1] estimated by applying SfM, SfM-KB-L, SfM-KB-NL and RANSAC-KB-NL to 100 synthetic scenes, with $\alpha = 0.1°$.
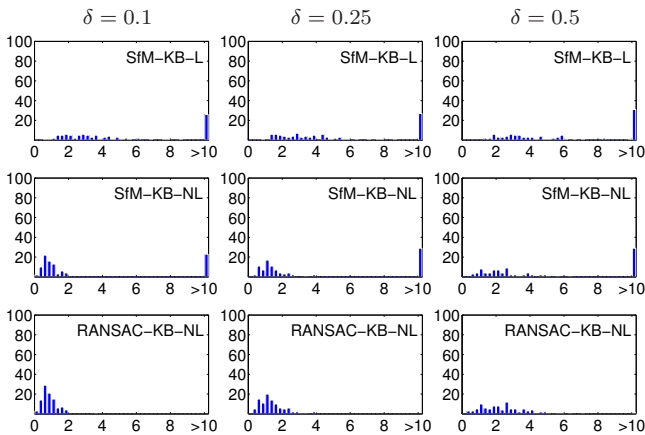

Figure 3. Histograms of errors in orientations[1] estimated by applying SfM, SfM-KB-L, SfM-KB-NL and RANSAC-KB-NL to 100 synthetic scenes, with $\alpha = 0.2°$.

Each of these three input images has 1200×1600 pixels. Their internal calibration matrices were calculated assuming a $49°$ horizontal field-of-view, which yields a focal length of about 1700 pixels. Viewing positions obtained from GPS measurements imply the following baseline lengths: $\|\mathbf{b}_{1,2}\| = 8.5m$, $\|\mathbf{b}_{2,3}\| = 8.3m$, $\|\mathbf{b}_{3,1}\| = 0.26m$, with an uncertainty of about 2m per measurement. Because views 1 and 3 are very close, this is a nearly singular configuration, *i.e.* almost a worst-case scenario for methods based on epipolar constraints.

In order to evaluate the accuracy of the orientations estimated with RANSAC-KB-NL, we used them (in conjunction with GPS measurements and the cameras' intrinsic matrices) to compute optical rays through the selected features, in world coordinates. Then, we recovered a sparse 3D model of the scene by finding the least-squares intersections of each triple of corresponding rays. Finally, we reprojected all estimated 3D points back to each image and compared the resulting image coordinates against those of the manually-selected features.

As shown in Figure 4, none of the Root-Mean-Square (RMS) reprojection errors generated by RANSAC-KB-NL on the three input views were larger than 3.2 pixels, which is less than 0.2% of the focal length. We also used the sparse 3D structure computed with RANSAC-KB-NL to localize four other views that were part of the Contest's final dataset (pictures 0687, 5296, 6673 and 7632), using a pose estimation method [1]. The localization errors obtained for those four viewpoints were, respectively, 9.3m, 1.1m, 8.5m, 6.0m, which are all within the range for which positive scores were attributed in the Contest's final round [20].

7. Limitations and Future Work

In this paper we provided the algebraic framework to work with multiple views that are accompanied by GPS positions. Our next goal is to use the framework above but include uncertainty in the viewpoint positions using a bounded error model that can be incorporated with inequality constraints. To make the system more practicable we will study the case of varying focal length and corresponding conditions for fundamental matrices to be consistent with given viewpoints. In both cases of calibrated and uncalibrated cameras it is very important to be able to identify degenerate cases and in particular planar scenes that can be modeled with collineations.

Moreover, we have shown what is the minimum number of point correspondences needed to recover N viewing orientations when viewing positions are known, but we have not shown that such minimum sets of correspondences are *sufficient* to disambiguate the problem completely. In fact, due to the polynomial nature of the constraints created by point correspondences, finite ambiguities almost certainly *do* exist in minimal cases. Performing reliable orientation recovery with very few point correspondences is thus a major challenge that we intend to tackle by applying recent developments in global optimization of general computer vision problems [7, 8, 18] to the particular problem studied in this paper.

A. Proof of Lemma 1

The fact that Eq. (2) implies Eqs. (4)-(6) is well known [9]. To see that the converse holds, observe that Eq. (4) implies that the vectors $\mathbf{v}_1$, $\mathbf{b}_{1,2}$ and $\mathbf{v}_2$ are linearly dependent, where $\mathbf{v}_i \triangleq \mathbf{R}_i \mathbf{u}_i$. Moreover, because $\mathbf{p}$ is in general position with respect to $\mathbf{t}_1$ and $\mathbf{t}_2$ by hypothesis, $\mathbf{v}_1$ and $\mathbf{v}_2$ are
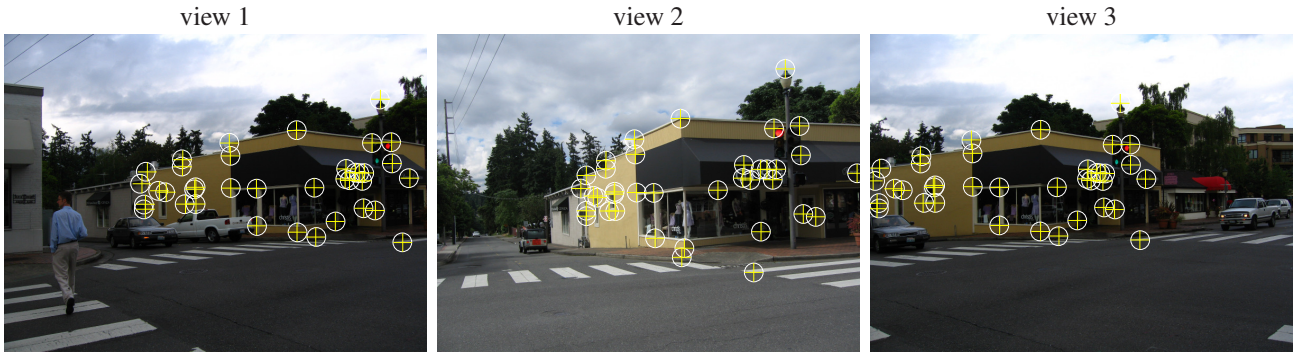
Figure 4. Real dataset from the *ICCV2005 Computer Vision Contest* [20] and results obtained on this dataset by RANSAC-KB-NL. White circles represent manually selected features. Yellow crosses represent reprojections of the 3D points reconstructed with RANSAC-KB-NL. RMS reprojection errors on views 1, 2 and 3 are: 2.9 pixels, 0.39 pixels and 3.2 pixels, respectively.

linearly independent. Hence

$$\exists! \ \lambda_1, \mu_2 \in \Re \quad | \quad \lambda_1 \mathbf{v}_1 - \mu_2 \mathbf{v}_2 = \mathbf{b}_{1,2}. \qquad (32)$$

Analogously, Eqs. (5)-(6) imply that

$$\exists! \ \lambda_2, \mu_i \in \Re \quad | \quad \lambda_2 \mathbf{v}_2 - \mu_i \mathbf{v}_i = \mathbf{b}_{2,i}, \qquad (33)$$

$$\exists! \ \lambda_i, \mu_1 \in \Re \quad | \quad \lambda_i \mathbf{v}_i - \mu_1 \mathbf{v}_1 = \mathbf{b}_{i,1}. \qquad (34)$$

Now because $\mathbf{t}_1$, $\mathbf{t}_2$ and $\mathbf{t}_i$ are points in the Euclidean space,

$$\mathbf{b}_{1,2} + \mathbf{b}_{2,i} + \mathbf{b}_{i,1} = 0. \qquad (35)$$

By adding Eqs. (32)-(34) and substituting Eq. (35) on the r.h.s. of the resulting expression, we obtain

$$\left(\lambda_1 - \mu_1 \right) \mathbf{v}_1 + \left(\lambda_2 - \mu_2 \right) \mathbf{v}_2 + \left(\lambda_i - \mu_i \right) \mathbf{v}_i = 0. \qquad (36)$$

Since $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_i$ are linearly independent by hypothesis, Eq. (36) implies that $\mu_1 = \lambda_1$, $\mu_2 = \lambda_2$, $\mu_i = \lambda_i$. Substituting this on Eqs. (32)-(34) yields Eq. (2). □

References

[1] A. Ansar and K. Daniilidis. Linear pose estimation from points or lines. In *Proc. European Conf. on Comput. Vision*, volume 4, pages 282–296, 2002.

[2] A. Chiuso, P. Favaro, H. Jin, and S. Soatto. Structure from motion causally integrated over time. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):523–535, 2002.

[3] D. A. Cox, J. B. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer-Verlag, 1996.

[4] M. Curtis. *Matrix Groups*. Springer-Verlag, 1979.

[5] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(6):580–593, 1997.

[6] B. Johansson and R. Cipolla. A system for automatic pose-estimation from a single image in a city scene. In *Proc. IASTED Intl. Conf. Signal Proc. Pattern Recog. Appl.*, 2002.

[7] F. Kahl. Multiple view geometry and the L-infinity norm. In *Proc. IEEE Intl. Conf. on Comput. Vision*, volume 2, pages 1002–1009, 2005.

[8] F. Kahl and D. Henrion. Globally optimal estimates for geometric reconstruction problems. In *Proc. IEEE Intl. Conf. on Comput. Vision*, volume 2, pages 978–985, 2005.

[9] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[10] Y. Ma, K. Huang, R. Vidal, J. Kosecka, and S. Sastry. Rank conditions on the multiple-view matrix. *Intl. J. Comput. Vision*, 59(2):115–137, 2004.

[11] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–770, 2004.

[12] D. Nistér. Preemptive RANSAC for live structure and motion estimation. *Mach. Vision Appl.*, 16(5):321–329, 2005.

[13] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, volume 1, pages 560–567, 2004.

[14] D. Nistér and F. Schaffalitzky. Four points in two or three calibrated views: theory and practice. *Intl. J. Comput. Vision*, to appear, 2006.

[15] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *Intl. J. Comput. Vision*, 59(3):207–232, 2004.

[16] L. Quan. Algebraic relations among matching constraints of multiple images. Technical Report RR-3345, INRIA, 1998.

[17] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or how do I organize my holiday snaps? In *Proc. European Conf. on Comput. Vision*, volume 1, pages 414–431, 2002.

[18] H. Stewenius, F. Schaffalitzky, and D. Nistér. How hard is 3-view triangulation really? In *Proc. IEEE Intl. Conf. on Comput. Vision*, volume 1, pages 686–693, 2005.

[19] J. Stuelpnagel. On the parametrization of the three-dimensional rotation group. *SIAM Review*, 6(4):422–430, 1964.

[20] R. Szeliski. ICCV2005 Computer Vision Contest, 2005. http://research.microsoft.com/iccv2005/Contest/.

[21] C. Taylor and D. Kriegman. Minimization on the lie group SO(3) and related manifolds. Technical Report 9405, Dept. Electrical Engineering, Yale University, 1994.

[22] T. Yeh, K. Tollmar, and T. Darrell. Searching the web with mobile images for location recognition. In *Proc. IEEE Conf. on Comput. Vision and Pattern Recognition*, volume 2, pages 76–81, 2004.