Algebraic Characterization of Essential Matrices and Their Averaging in Multiview Settings

Yoni Kasten* Amnon Geifman* Meirav Galun Ronen Basri
Weizmann Institute of Science
{yoni.kasten,amnon.geifman,meirav.galun,ronen.basri}@weizmann.ac.il

Abstract

Essential matrix averaging, i.e., the task of recovering camera locations and orientations in calibrated, multiview settings, is a first step in global approaches to Euclidean structure from motion. A common approach to essential matrix averaging is to separately solve for camera orientations and subsequently for camera positions. This paper presents a novel approach that solves simultaneously for both camera orientations and positions. We offer a complete characterization of the algebraic conditions that enable a unique Euclidean reconstruction of n cameras from a collection of $\binom{n}{2}$ essential matrices. We next use these conditions to formulate essential matrix averaging as a constrained optimization problem, allowing us to recover a consistent set of essential matrices given a (possibly partial) set of measured essential matrices computed independently for pairs of images. We finally use the recovered essential matrices to determine the global positions and orientations of the n cameras. We test our method on common SfM datasets, demonstrating high accuracy while maintaining efficiency and robustness, compared to existing methods.

1. Introduction

What algebraic conditions make a collection of $\binom{n}{2}$ essential matrices consistent, in the sense that there exist n Euclidean camera matrices that produce them? This fundamental question has not yet been answered in the literature. It is well known that $\binom{3}{2} = 3$ fundamental matrices are consistent if, and only if, the epipole of the third view is transferred correctly between each pair of views, i.e., for every $1 \leq i, j, k \leq 3$, $\mathbf{e}_{ik}^T F_{ij} \mathbf{e}_{jk} = 0$. Recent work [15] presented a set of sufficient and necessary algebraic conditions that make $\binom{n}{2}$ fundamental matrices in general position consistent. One could expect that essential matrices that fulfill those same conditions would be consistent with respect to Euclidean camera matrices. However, these conditions are not sufficient and can be contradicted by a counter example, see one such construction in the supplementary material.

Establishing consistency constraints for essential matrices is an important step toward producing essential matrix averaging algorithms. Given n images $I_1, .., I_n$, a common approach for global Structure from Motion (SfM) begins by robustly estimating essential matrices between pairs of views, $\{E_{ij}\}$, from which an estimate of the relative pairwise rotations $\{R_{ij}\}$ and translations $\{\mathbf{t}_{ij}\}$ are extracted. Motion averaging then is performed typically in two steps: first the absolute camera orientations $\{R_i\}$ are solved by averaging the relative rotations. Then, using the relative translations and the recovered absolute orientations, the absolute camera positions $\{\mathbf{t}_i\}$ are recovered. Finally, the obtained solution is refined by bundle adjustment.

Our goal in this paper is to establish a complete set of necessary and sufficient conditions for the consistency of essential matrices and to use these conditions to formulate a one-step algorithm for averaging essential matrices. To achieve this goal we investigate an object called the n-views essential matrix, which is obtained by stacking the $\binom{n}{2}$ essential matrices into a $3n \times 3n$ matrix whose i, j'th 3×3 block is the essential matrix E_{ij} relating the i'th and the j'th frames. We prove that, in addition to projective consistency (introduced in [15]), this matrix must have three pairs of eigenvalues each of the same magnitude but opposite signs, and its eigenvectors directly encode camera parameters.

We use these results to introduce the first (to the best of our knowledge) essential matrix averaging algorithm. Given a noisy estimate of a subset of $\binom{n}{2}$ essential matrices, our algorithm seeks to find the nearest consistent set of essential matrices. We formulate this problem as constrained optimization and solve it using ADMM. We then incorporate this algorithm in a global SfM pipeline and evaluate our pipeline on the datasets of [25], showing superior accuracies relative to state of the art methods on almost all image collections while also maintaining efficiency.

*Equal contributors

2. Related work

Approaches for Euclidean motion averaging can be divided into two main categories: Incremental methods [16, 20, 1, 14, 26] begin with a small subset of frames and produce an initial reconstruction. The rest of cameras are then used sequentially for reconstruction. These methods are very successful and quite robust. However, they have to apply bundle adjustment refinement at every step to prevent camera drift. Consequently, these method are computationally demanding when applied to large data sets.

Global methods [3, 25, 8, 13, 18], in contrast, recover the motion parameters simultaneously for all the frames. Typical global SfM pipelines proceed by applying a camera orientation solver, followed by a location solver.

Global orientation solvers [3, 17, 24, 9, 5] solve for the absolute orientations of the cameras given relative rotation measurements between pairs of images. [3, 17] derive closed form solutions that minimize a least squares objective constructed from the pairwise relative orientations. These methods are very efficient but due to the relaxed orthonormality requirement, the result is usually suboptimal. Other methods [24, 9, 5] utilize the Lie algebra structure of the rotation group to perform rotation averaging in $SO(3)$. These methods, however, often converge to local minima. Recently, [7] has used the strong duality principle to find the global minimum under certain conditions.

Global location solvers [3, 25, 18] assume known camera orientations and solve for the absolute positions of the cameras by using the noisy relative translations. [3] uses point correspondences to find a least squares solution for the absolute positions. [25] formulates a highly non convex objective and solves for the absolute translations utilizing the Levenberg-Marquet algorithm with random initialization. [18] uses a similar objective while formulating a convex relaxation. [6] uses linear global method that minimizes geometric error in triplet of views while considering feature tracks. All the aforementioned methods are highly dependent on accurate estimation of the absolute rotations of the cameras, which result from a rotation averaging method.

Integrative Methods: [22] formulates the problem of fundamental and essential averaging as global optimization through minimizing the epipolar transfer error. While their method manages to improve the consistency of viewing graphs, it is unable to generate a consistent reconstruction, and so it requires postprocessing steps of both rotation and translation averaging. [19] has introduced the concept of a multi-view essential/fundamental matrix. Their work, however, established only a partial list of constraints. Moreover, their use of a complicated non-convex objective allowed them to only refine a complete reconstruction. [15] has introduced a complete set of consistency conditions for fundamental matrices. They formulate a robust optimization objective and demonstrate state of the art projective reconstructions. Their method, however, is limited to projective settings and is inapplicable to calibrated settings, i.e., for Euclidean reconstruction. [24] suggested a method that optimizes first for camera positions and then for their orientations, and as a post processing simultaneously optimizes for both. However, this method is sensitive to outliers. Recent work explored the properties of the manifold of essential matrices [23]. Their characterization, however, is suitable only for a single essential matrix and not for general multiview settings. Finally, [2, 11] explore general algebraic properties of multi-view settings.

Our paper extends the work of [15, 19] by introducing a complete set of necessary and sufficient conditions for consistency of multiview essential matrices and by proposing an efficient and robust optimization algorithm for essential matrix averaging that incorporates these conditions.

3. Theory

Let $I_1, ..., I_n$ denote a collection of n images of a static scene captured respectively by cameras $P_1, ..., P_n$. Each camera P_i is represented by a 3×4 matrix $P_i = K_i R_i^T [I, -t_i]$ where K_i is a 3×3 calibration matrix, $\mathbf{t_i} \in \mathbb{R}^\mathbf{3}$ and $R_i \in SO(3)$ denote the position and orientation of P_i, respectively, in some global coordinate system. We further denote $V_i = K_i^{-T} R_i^T$, so $P_i = V_i^{-T}[I, -\mathbf{t}_i]$. Consequently, let $\mathbf{X} = (X, Y, Z)^T$ be a scene point in the global coordinate system. Its projection onto I_i is given by $\mathbf{x}_i = \mathbf{X}_i / Z_i$, where $\mathbf{X}_i = (X_i, Y_i, Z_i)^T = K_i R_i^T (\mathbf{X} - \mathbf{t}_i)$.

We denote the fundamental matrix and the essential matrix between images I_i and I_j by F_{ij} and E_{ij} respectively. It was shown in [3] that E_{ij} and F_{ij} can be written as

$$E_{ij} = R_i^T (T_i - T_j) R_j$$

$$F_{ij} = K_i^{-T} E_{ij} K_j^{-1} = V_i (T_i - T_j) V_j^T$$

where $T_i = [\mathbf{t}_i]_\times$.

Throughout this paper we assume that all calibration matrices are known, so our work deals with solving the problem of Euclidean SfM.

The derivations in this paper adopt the definitions "n-view fundamental matrix" and "consistent n-view fundamental matrix" from [15]. We first repeat these definitions, for the sake of clarity, and then define analogous definitions for the calibrated case. In the definitions below we denote the space of all the $3n \times 3n$ symmetric matrices by $\mathbb{S}^{3n}$.

Definition 1. *A matrix $F \in \mathbb{S}^{3n}$, whose 3×3 blocks are denoted by F_{ij}, is called an* **n-view fundamental matrix** *if $\forall i \neq j \in \{1, ..., n\}$, $rank(F_{ij}) = 2$ and $\forall i \; F_{ii} = 0$.*

Definition 2. *An n-view fundamental matrix F is called* **consistent** *if there exist camera matrices $P_1, ..., P_n$ of the form $P_i = V_i^{-T}[I, \mathbf{t}_i]$ such that $F_{ij} = V_i([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times) V_j^T$.*

Definition 3. *A matrix $E \in \mathbb{S}^{3n}$, whose 3×3 blocks are denoted by E_{ij}, is called an* **n-view essential matrix** *if $\forall i \neq j \in \{1 \dots n\}$, $rank(E_{ij}) = 2$, the two singular values of E_{ij} are equal, and $\forall i$ $E_{ii} = 0$.*

Definition 4. *An n-view essential matrix E is called* **consistent** *if there exist n rotation matrices $\{R_i\}_{i=1}^n$ and n vectors $\{\mathbf{t}_i\}_{i=1}^n$ such that $E_{ij} = R_i^T([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times)R_j$.*

Note that any (consistent) n-view essential matrix is also a (consistent) n-view fundamental matrix. In [15] necessary and sufficient conditions for the consistency of the n-view fundamental matrix were proved. The main theoretical contribution of [15] is summarized in Theorem 1. For the consistency of n-view essential matrix, a partial set of necessary conditions were derived in [19]. Those are summarized below in Theorem 2.

Theorem 1. *An n-view fundamental matrix F is consistent with a set of n cameras whose centers are not all collinear if, and only if, the following conditions hold:*

1. *$Rank(F) = 6$ and F has exactly 3 positive and 3 negative eigenvalues.*
2. *$Rank(F_i) = 3$ for all $i = 1, ..., n$, where F_i is the $3 \times 3n$ i^{th} block row of F.*

Theorem 2. *Let E be a consistent n-view essential matrix, associated with rotation matrices $\{R_i\}_{i=1}^n$ and camera centers $\{\mathbf{t}_i\}_{i=1}^n$. E satisfies the following conditions*

1. *E can be formulated as $E = A + A^T$ where $A = UV^T$ and $U, V \in \mathbb{R}^{3n \times 3}$*

$$V = \begin{bmatrix} R_1^T \\ \vdots \\ R_n^T \end{bmatrix} \quad U = \begin{bmatrix} R_1^T T_1 \\ \vdots \\ R_n^T T_n \end{bmatrix} \tag{1}$$

   with $T_i = [\mathbf{t}_i]_\times$ and $\sum_{i=1}^n \mathbf{t}_i = 0$.

2. *Each column of U is orthogonal to each column of V, i.e., $V^T U = 0_{3\times 3}$*
3. *rank(V)=3*
4. *If not all $\{\mathbf{t}_i\}_{i=1}^n$ are collinear, then rank(U) and rank(A) = 3. Moreover, if the (thin) SVD of A is $A = \hat{U}\Sigma\hat{V}^T$, with $\hat{U}, \hat{V} \in \mathbb{R}^{3n\times 3}$ and $\Sigma \in \mathbb{R}^{3\times 3}$ then the (thin) SVD of E is*

$$E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$$

   implying rank(E) = 6.

3.1. Main theoretical results

In this section we derive and prove necessary and sufficient conditions for the consistency of n-view essential matrices in terms of their spectral decomposition. These conditions, in turn, will be used later to formulate a constrained optimization problem and to extract the motion parameters from a consistent n-view essential matrix E .

Theorem 3. *Let $E \in \mathbb{S}^{3n}$ be a consistent n-view fundamental matrix with a set of n cameras whose centers are not all collinear. We denote by $\Sigma_+, \Sigma_- \in \mathbb{R}^{3\times 3}$ the diagonal matrices with the 3 positive and 3 negative eigenvalues of E, respectively. Then, the following conditions are equivalent:*

1. *E is a consistent n-view essential matrix*
2. *The (thin) SVD of E can be written in the form*

$$E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma_+ & \\ & \Sigma_+ \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$$

   with $\hat{U}, \hat{V} \in \mathbb{R}^{3n\times 3}$ such that each 3×3 block of $\hat{V}$, $\hat{V}_i$, $i = 1, ..., n$, is an $\sqrt{n}$-scaled rotation matrix, i.e., $\hat{V}_i = \frac{1}{\sqrt{n}}\hat{R}_i$, where $\hat{R}_i \in SO(3)$. We say that $\hat{V}$ is a block rotation matrix.

3. *$\Sigma_+ = -\Sigma_-$ and the (thin) spectral decomposition of E is of the form*

$$E = [X, Y] \begin{pmatrix} \Sigma_+ & \\ & \Sigma_- \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}$$

   such that $\sqrt{0.5}(X + Y)$ is a block rotation matrix.

Proof. *(1)⇒(2)* Assume that E is a consistent n-view essential matrix. Then, according to Thm. 2, $E = A + A^T$ with $A = UV^T$ and $U, V \in \mathbb{R}^{3n\times 3}$ which take the formas in (1). Since $A = UV^T$ and $rank(A) = 3$, then $A^TA = VU^TUV$ and $A^TA \succeq 0$ with $rank(A^TA) = 3$ (A and A^TA share the same null space). First, we construct a spectral decomposition to A^TA, relying on the special properties of U and V. We have $rank(U) = 3$, and therefore U^TU, which is a 3×3, symmetric positive semidefinite matrix, is of full rank. Its spectral decomposition is of the form $U^TU = QDQ^T$, where $Q \in SO(3)$. (Spectral decomposition guarantees that $Q \in O(3)$. However, Q can be replaced by $-Q$ if $\det(Q) = -1$.) $D \in \mathbb{R}^{3\times 3}$ is a diagonal matrix consisting of the (positive) eigenvalues of U^TU. This spectral decomposition yields the following decomposition

$$A^TA = VQDQ^TV^T. \tag{2}$$

Now, note that

$$Q^TV^TVQ = Q^T\left[R_1, ..., R_n\right] \begin{bmatrix} R_1^T \\ \vdots \\ R_n^T \end{bmatrix} Q = nI_{3\times 3}.$$

By a simple manipulation (2) becomes a spectral decomposition

$$A^T A = \left(\frac{1}{\sqrt{n}}V\right) Q(nD)Q^T \left(\frac{1}{\sqrt{n}}V^T\right). \quad (3)$$

On the other hand, the (thin) SVD of A is of the form $A = \hat{U}\Sigma\hat{V}^T$, where $\hat{U}, \hat{V} \in \mathbb{R}^{3n\times 3}$, $\Sigma \in \mathbb{R}^{3\times 3}$. This means that

$$A^T A = \hat{V}\Sigma^2\hat{V}^T. \quad (4)$$

Due to the uniqueness of the eigenvector decomposition, (3) and (4) collapse to the same eigenvector decomposition, up to some global rotation, $H \in SO(3)$, that is $\frac{1}{\sqrt{n}}VQ = \hat{V}H$, which means that $\hat{V}_i = \frac{1}{\sqrt{n}}R_i^T QH^T$. Since $R_i^T, Q, H^T \in SO(3)$, then $\hat{R}_i =: R_i^T QH^T \in SO(3)$, showing that $\hat{V}$ is a block rotation matrix. Finally, by Thm. 2, the (thin) SVD of E is of the form

$$E = \begin{bmatrix}\hat{U}, \hat{V}\end{bmatrix} \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix} \quad (5)$$

and according to Lemma 5, the eigenvalues of E are Σ and $-\Sigma$. Since the elements on the diagonal of Σ are positive, and E is symmetric with exactly 3 positive eigenvalues Σ_+ and 3 negative eigenvalues Σ_-, it follows that $\Sigma = \Sigma_+$ and $-\Sigma = \Sigma_-$ concluding the proof.

(2)⇒(1) Let E be a consistent n-view fundamental matrix that satisfies condition (2). We would like to show that E is a consistent n-view essential matrix. By condition (2) E can be written as

$$E = \hat{U}\Sigma_+\hat{V}^T + \hat{V}\Sigma_+\hat{U}^T = \bar{U}\hat{V}^T + \hat{V}\bar{U}^T, \quad (6)$$

where $\bar{U} = \hat{U}\Sigma_+$ with $\hat{V}_i = \frac{1}{\sqrt{n}}\hat{R}_i$, $\hat{R}_i \in SO(3)$. By definition $E_{ii} = 0$, and this implies that $\bar{U}_i\hat{V}_i^T$ is skew symmetric. Using Lemma 4, $\bar{U}_i = \hat{V}_i\hat{T}_i$ for some skew symmetric matrix $\hat{T}_i = [\hat{\mathbf{t}}_i]_\times$. Plugging $\bar{U}_i$ and $\hat{V}_i$ in (6) yields

$$\begin{aligned} E_{ij} &= \bar{U}_i\hat{V}_j^T + \hat{V}_i\bar{U}_j^T \\ &= \frac{1}{n}\hat{R}_i\hat{T}_i\hat{R}_j^T - \frac{1}{n}\hat{R}_i\hat{T}_j\hat{R}_j^T \\ &= {R_i}^T([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times)R_j, \end{aligned}$$

where $R_i = \hat{R}_i^T$ and $\mathbf{t}_i = \frac{1}{n}\hat{\mathbf{t}}_i$, concluding the proof.

(2)⇒(3) Let E be an n-view fundamental matrix which satisfies condition (2). This means that the (thin) SVD of E can be written in the form $E = \begin{bmatrix}\hat{U}, \hat{V}\end{bmatrix} \begin{pmatrix} \Sigma_+ & \\ & \Sigma_+ \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$, where $\hat{V}$ is a block rotation matrix. Then, by Lemma 5, the (thin) spectral decomposition of E is $E = [X, Y] \begin{pmatrix} \Sigma_+ & \\ & -\Sigma_+ \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}$, where X, Y are the eigenvectors of E satisfying $X = \sqrt{0.5}(\hat{U} + \hat{V})$ and $Y = \sqrt{0.5}(\hat{V} - \hat{U})$. Since $\hat{V} = \sqrt{0.5}(X + Y)$, and by condition (2) $\hat{V}$ is a block rotation matrix, the claim is confirmed and also $\Sigma_- = -\Sigma_+$.

(3)⇒(2) Let E be a consistent n-view fundamental matrix satisfying condition (3), i.e., its (thin) spectral decomposition is of the form $E = [X, Y] \begin{pmatrix} \Sigma_+ & \\ & \Sigma_- \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}$, where $\sqrt{0.5}(X+Y)$ is a block rotation matrix. Since $\Sigma_+ = -\Sigma_-$, we can use Lemma 5, which implies that the (thin) SVD is of the form $E = \begin{bmatrix}\hat{U}, \hat{V}\end{bmatrix} \begin{pmatrix} \Sigma_+ & \\ & \Sigma_+ \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$, where $\hat{V} = \sqrt{0.5}(X + Y)$, concluding the proof. □

Corollary 1. *Euclidean reconstruction. Let E be a consistent n-view essential matrix with 6 distinct eigenvalues, then it is possible to explicitly determine $R_1 \dots R_n$ and $\mathbf{t}_1, \dots, \mathbf{t}_n$ that are consistent with respect to E.*

Proof. The claim is justified by the following construction, which relies on the spectral characterizations of Thm. 3.

1. Calculate the eigenvectors X, Y of E, and the corresponding three positive eigenvalues, Σ_+, and three negative eigenvalues, Σ_-, respectively.
2. To realize condition (3) of Thm. 3, there are 8 possible choices of $\sqrt{0.5}(X + YI_s)$, where $I_s = \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}$, due to the sign ambiguity of each eigenvector. Then, I_s is chosen such that $\sqrt{0.5}(X + YI_s)$ is block rotation matrix up to a global sign which can be removed.
3. This spectral decomposition determines directly an SVD decomposition in the form of condition (2) of Thm. 3. We would like to emphasize that due to the multiplicity of singular values, a standard SVD method which is performed directly on E, in general, will not produce this special SVD pattern.
4. Having the relation $E_{ij} = \hat{U}_i\Sigma_+\hat{V}_j^T + \hat{V}_i\Sigma_+\hat{U}_j^T$ and since $E_{ii} = 0$ we get that $\hat{U}_i\Sigma_+\hat{V}_i^T$ is skew symmetric. We denote $\bar{U}_i = \hat{U}_i\Sigma_+$ and, by Lemma 4, it holds that $\hat{T}_i = \hat{V}_i^{-1}\bar{U}_i$.
5. Finally, for $i = 1, 2, .., n$, define $R_i =: \sqrt{n}\hat{V}_i^T$ and $\mathbf{t}_i =: \frac{1}{n}\hat{\mathbf{t}}_i$ and it holds that

$$E_{ij} = {R_i}^T([\mathbf{t}_i]_\times - [\mathbf{t}_j]_\times)R_j.$$

This construction yields $\{R_i\}_{i=1}^{i=n}$ and $\{\mathbf{t}_i\}_{i=1}^{i=n}$ which are consistent with respect to E. Moreover, the reconstruction is unique up to a global similarity transformation. Roughly speaking, this can be proven for $n = 3$ by applying an argument from [12]. Next, for $n > 3$, by induction, suppose we obtain two reconstructions, $P_1, ..., P_n$ and $P'_1, ..., P'_n$. By the induction assumption these must include two sets of $n-1$ non-collinear cameras so that each is unique up to a similarity transformation. Such two sets overlap in at least 2 cameras, which in turn imply that the two similarity transformations must be identical. The complete proof is provided in the supplementary material.

3.2. Supporting lemmas

Lemma 4. *[15] Let $A, B \in \mathbb{R}^{3\times3}$ with $rank(A) = 2$, $rank(B) = 3$ and AB^T is skew symmetric, then $T = B^{-1}A$ is skew symmetric.*

Lemma 5. *Let $E \in \mathbb{S}^{3n}$ of rank(6), and $\Sigma \in \mathbb{R}^{3\times3}$, a diagonal matrix, with positive elements on the diagonal. Let $X, Y, U, V \in \mathbb{R}^{3n\times3}$, and we define the mapping $(X, Y) \leftrightarrow (U, V)$: $X = \sqrt{0.5}(\hat{U} + \hat{V})$, $Y = \sqrt{0.5}(\hat{V} - \hat{U})$, $\hat{U} = \sqrt{0.5}(X - Y)$, $\hat{V} = \sqrt{0.5}(X + Y)$.*

Then, the (thin) SVD of E is of the form

$$E = \left[\hat{U}, \hat{V}\right] \begin{pmatrix} \Sigma & \\ & \Sigma \end{pmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{U}^T \end{bmatrix}$$

if and only if the (thin) spectral decomposition of E is of the form

$$E = [X, Y] \begin{pmatrix} \Sigma & \\ & -\Sigma \end{pmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}$$

Proof. The proof is provided in the supplementary material. □

4. Method

Given images $I_1, ..., I_n$, we assume a standard robust method is used to estimate the pairwise essential matrices, which we denote by $\Omega = \{\hat{E}_{ij}\}$. In practice, only a small subset of the pairwise essential matrices are estimated, due to occlusion, large viewpoint and brightness changes as well as objects' motion, and in addition the available estimates are noisy. Our goal therefore is to find a consistent n-view essential matrix $E \in \mathbb{S}^{3n}$ that is as similar to the measured essential matrices as possible.

To make an n-view essential matrix consistent, its blocks of pairwise essential matrices must each be scaled by an unknown factor. [19] proposed an optimization scheme that explicitly seeks for the unknown scale factors, yielding a nonlinear, rank-constrained optimization formulation. The success of this approach critically depends on the quality of its initialization, which in experiments was obtained by applying another, state of the art SfM method.

More recently, [15] proposed an analogous approach for projective SfM. They showed that a consistent 3-view fundamental matrix, which uniquely determines camera matrices (up to a projective ambiguity) from a triplet of images, is invariant to scaling of its pairwise fundamental matrices. This allowed them to formulate an optimization problem that seeks 3-view fundamental matrices that are both consistent and compatible, while avoiding the need to explicitly optimize for the scale factors.

In this paper, we introduce an optimization scheme that is analogous to that of [15], but adapted to calibrated settings. In particular, our scheme uses the algebraic constraints derived in Thm. 3 to enforce the consistency of noisy, and possibly partial essential matrices. Similar to [15], our method simultaneously enforces consistency of camera triplets attached rigidly to each other, allowing us to avoid optimizing explicitly for the unknown scales of the estimated essential matrices. (To that end we further extend Thm. 3 to handle scaled rotations for image triplets, see supplementary material for details.) Our formulation, however, is more involved than in [15] due to the additional spectral constraints required for Euclidean reconstruction.

In the rest of this section we present our constrained optimization formulation and propose an ADMM-based solution scheme. Subsequently, we discuss how to select minimal subsets of triplets to speed up the optimization. Finally, we show how the results of our optimization can be used to reconstruct the absolute orientations and positions of the n cameras.

4.1. Optimization

In multi-view settings, it is common to define a viewing graph $G = (V, W)$, with nodes $v_1, \ldots, v_n$, corresponding to the n cameras, and $w_{ij} \in W$ if $\hat{E}_{ij}$ is one of the estimated pairwise essential matrices. Let τ denote a collection of m 3-cliques of cameras where $m \leq \binom{n}{3}$. The collection may be the full set of the 3-cliques in G, or a chosen subset as described in Sec. 4.2. We index the elements of τ by $k = 1, ..., m$, where $\tau(k)$ denote the k^{th} triplet. The collection τ determines a partial selection of measured essential matrices, Ω, that plays a role in the optimization problem, where it holds that if $\hat{E}_{ij} \in \Omega$ then $\hat{E}_{ij}^T = \hat{E}_{ji} \in \Omega$.

We define the measurements matrix $\hat{E} \in \mathbb{S}^{3n}$ to have $\hat{E}_{ij}$ as its $(i, j)^{th}$ block if $\hat{E}_{ij} \in \Omega$ and $0_{3\times3}$ in the rest of its blocks. In our optimization problem we look for E that is as close as possible to $\hat{E}$ under the constraints that its 9×9 blocks, induced by $\{\tau(k)\}_{k=1}^m$ and denoted by $\{E_{\tau(k)}\}_{k=1}^m$, are consistent 3-view essential matrices. In general, such E is inconsistent and incomplete, but as we explain in Sec. 4.3 it is possible to retrieve a set of n absolute rotations and translations that is compatible with its essential matrices up to scale, which in turn implies that the completion of the missing entries is consistent.

We formulate our constrained optimization as follows

$$\begin{aligned} \underset{E}{\text{minimize}} \quad & \sum_{k=1}^{m} ||E_{\tau(k)} - \hat{E}_{\tau(k)}||_F^2 \qquad (7) \\ \text{subject to} \quad & E = E^T \\ & E_{ii} = 0_{3\times 3} \\ & rank(E_{\tau(k)}) = 6 \\ & \Sigma_+(E_{\tau(k)}) = -\Sigma_-(E_{\tau(k)}) \\ & X(E_{\tau(k)}) + Y(E_{\tau(k)}) \text{ is block rotation,} \end{aligned}$$

with $i = 1, ..., n$ and $k = 1, ..., m$, where $\Sigma_+(E_{\tau(k)}), \Sigma_-(E_{\tau(k)}) \in \mathbb{R}^3$ denote the 3 largest (descending order) and 3 smallest (ascending order) eigenvalues of $E_{\tau(k)}$ respectively and $X(E_{\tau(k)}) \in \mathbb{R}^{9\times 3}$ and $Y(E_{\tau(k)}) \in \mathbb{R}^{9\times 3}$ are their corresponding eigenvectors.

Solving (7) is challenging due to its rank and spectral decomposition constraints. We solve this optimization problem using ADMM. To that end, as part of the ADMM method [4] $4m$ auxiliary matrix variables of size 9×9 are added: $2m$ variables duplicating $\{E_{\tau(k)}\}_{k=1}^m$, denoted $B = \{B_k\}_{k=1}^m$ and $D = \{D_k\}_{k=1}^m$, as well as $2m$ Lagrange multipliers, $\Gamma = \{\Gamma_k\}_{k=1}^m$ and $\Phi = \{\Phi_k\}_{k=1}^m$. This yields the following constrained optimization problem

$$\begin{aligned} \max_{\Gamma,\Phi} \min_{E,B,D} \quad & \sum_{k=1}^{m} L(E_{\tau(k)}, B_k, \Gamma_k, D_k, \Phi_k) \qquad (8) \\ \text{subject to} \quad & E = E^T \\ & E_{ii} = 0_{3\times 3} \\ & rank(B_k) = rank(D_k) = 6 \\ & \Sigma_+(B_k) = -\Sigma_-(B_k) \\ & X(D_k) + Y(D_k) \text{ is block rotation} \end{aligned}$$

with $i = 1, ..., n$ and $k = 1, ..., m$, where

$$L(E_{\tau(k)}, B_k, \Gamma_k, D_k, \Phi_k) = \left\| E_{\tau(k)} - \hat{E}_{\tau(k)} \right\|_F^2 + \frac{\alpha_1}{2} \left\| B_k - E_{\tau(k)} + \Gamma_k \right\|_F^2 + \frac{\alpha_2}{2} \left\| D_k - E_{\tau(k)} + \Phi_k \right\|_F^2 .$$

We initialize the auxiliary variables at $t = 0$ with

$$B_k^{(0)} = \hat{E}_{\tau(k)}, D_k^{(0)} = \hat{E}_{\tau(k)}, \Gamma_k^{(0)} = 0, \Phi_k^{(0)} = 0.$$

Then, we solve the optimization problem iteratively by alternating between the following four steps.

(i) Solving for E.

$$\begin{aligned} E^{(t)} = \underset{E}{\text{argmin}} \quad & \sum_{k=1}^{m} \left\| E_{\tau(k)} - \hat{E}_{\tau(k)} \right\|_F^2 \qquad (9) \\ & + \frac{\alpha_1}{2} \left\| B_k^{(t-1)} - E_{\tau(k)} + \Gamma_k^{(t-1)} \right\|_F^2 \\ & + \frac{\alpha_2}{2} \left\| D_k^{(t-1)} - E_{\tau(k)} + \Phi_k^{(t-1)} \right\|_F^2 \\ \text{subject to} \quad & E = E^T \text{ and } E_{ii} = 0_{3\times 3} \end{aligned}$$

This is a convex quadratic problem and can be solved efficiently, using a closed form solution.

(ii) Solving for B_k. For all $k = 1, ..., m$

$$\begin{aligned} B_k^{(t)} = \underset{B_k}{\text{argmin}} \quad & ||B_k - E_{\tau(k)}^{(t)} + \Gamma_k^{(t-1)}||_F^2 \qquad (10) \\ \text{subject to} \quad & rank(B_k) = 6 \\ & \Sigma_+(B_k) = -\Sigma_-(B_k) \end{aligned}$$

The minimizer for this sub-problem is obtained in the following way. By construction, $E_{\tau(k)}^{(t)} - \Gamma_k^{(t-1)}$ is a symmetric matrix, and we denote its (full) spectral decomposition by $U\Sigma U^T$, where $U \in \mathbb{R}^{9\times 9}$ and $\Sigma \in \mathbb{R}^{9\times 9}$ is a diagonal matrix in which the eigenvalues are arranged in a descending order. Then, the update is

$$B_k^{(t)} = U\Sigma^* U^T, \qquad (11)$$

where $\Sigma^* \in \mathbb{R}^{9\times 9}$ is a diagonal matrix with the entries

$$\Sigma_{ii}^* = \begin{cases} \frac{1}{2}(\Sigma_{ii} - \Sigma_{10-ii}) & i \neq 4,5,6 \\ 0 & i = 4,5,6 \end{cases} \qquad (12)$$

(iii) Solving for D_k. For all $k = 1, ..., m$

$$\begin{aligned} D_k^{(t)} = \underset{D_k}{\text{argmin}} \quad & ||D_k - E_{\tau(k)}^{(t)} + \Phi_k^{(t-1)}||_F^2 \qquad (13) \\ \text{subject to} \quad & rank(D_k) = 6 \\ & X(D_k) + Y(D_k) \text{ is a block rotation matrix} \end{aligned}$$

We minimize this sub-problem by an iterative process, which we repeat until convergence. We begin with $D_k = E_{\tau(k)}^{(t)} - \Phi_k^{(t-1)}$, which is symmetric by construction. We apply spectral decomposition to D_k, and extract $X(D_k)$, $Y(D_k)$, $\Sigma_+(D_k)$ and $\Sigma_-(D_k)$. Assuming no eigenvalue multiplicities, the eigenvectors are determined uniquely up to a sign (this argument is justified in 5.1). We denote by I_s, a diagonal matrix of size 3×3, such that each diagonal element is either 1 or -1. There are eight configurations for I_s from which we select the best, in the sense that on average each 3×3 block, of the form, $\sqrt{0.5}[X + YI_s]_i$, $i = 1, 2, 3$, is close to scaled rotation, using the following score,

$$I_s^* = \underset{I_s}{\text{argmax}} \sum_{i=1}^{3} \frac{\left\| diag((X_i + Y_i I_s)^T (X_i + Y_i I_s)) \right\|_2}{\|(X_i + Y_i I_s)^T (X_i + Y_i I_s)\|_F}.$$

Next, let $V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$ be the projection of $\sqrt{0.5}(X + YI_s^*)$ so that V_i is the closest scaled rotation to $\sqrt{0.5}[X + YI_s^*]_i$. Projection to scaled $SO(3)$ is obtained through removal of the singular values from the SVD decomposition, setting the scale factor to the average of the singular values, and possibly negating the scale factor to make the determinant positive. Let $U = \sqrt{0.5}(X - YI_s^*)$, and $\tilde{X} = \sqrt{0.5}(U + V)$ and $\tilde{Y} = \sqrt{0.5}(V - U)$. We then update the value of D_k to be the symmetric matrix

$$D_k = [\tilde{X}, \tilde{Y}] \begin{pmatrix} \Sigma_+ & \\ & \Sigma_- \end{pmatrix} \begin{bmatrix} \tilde{X}^T \\ \tilde{Y}^T \end{bmatrix}$$

and repeat these steps until convergence.

(iv) Updating Γ_k, Φ_k. For all $k = 1, \ldots, m$

$$\Gamma_k^{(t)} = \Gamma_k^{(t-1)} + B_k^{(t)} - E_{\tau(k)}^{(t)} \tag{14}$$

$$\Phi_k^{(t)} = \Phi_k^{(t-1)} + D_k^{(t)} - E_{\tau(k)}^{(t)} \tag{15}$$

4.2. Graph construction and outliers removal

As explained above, to apply our optimization algorithm, it is required to determine a collection τ of camera triplets, which is a subset of the given camera triplets. The selection of a subset allows for better efficiency and robustness. Similarly to [16, 15], we consider a weighted viewing graph G whose weights for each edge w_{ij} is assigned to be the number of the inlier matches relating I_i and I_j. We begin by selecting disjoint maximal spanning trees from G, from which we extract an initial subset of triplets. We then, remove near collinear and inconsistent triplets. We next build a triplet graph G_T whose nodes, which represent image triplets, are connected by an edge whenever two triplets share the same two cameras. Finally, we greedily remove nodes from G_T. starting with the least consistent triplet (using the rotation consistency score defined below), a node is removed as long as the connectivity of G_T is preserved and the total number of cameras associated with G_T does not decrease.

To define collinear and consistency scores for each triplet we denote the angles in the triangle formed by three cameras i, j, k by $\theta_i, \theta_j, \theta_k$ respectively. We measure each angle using the known relative translations $\mathbf{t}_{ij}, \mathbf{t}_{ik}, \mathbf{t}_{jk}$, i.e., $\cos\theta_i = (\frac{\mathbf{t}_{ij}^T \mathbf{t}_{ik}}{||\mathbf{t}_{ij}||||\mathbf{t}_{ik}||})$. Then, the collinearity score of cameras $\{i, j, k\}$ is the minimal angle in $\{\theta_i, \theta_j, \theta_k\}$. The consistency score of translations is defined by $|\theta_i + \theta_j + \theta_k - \pi|$ and the consistency score of rotations by $||R_{ij} R_{jk} R_{ki} - I||$.

4.3. Location and orientation retrieval

After solving (7), we extract from E the collection of 3-view essential matrices $\{E_{\tau(k)}\}_{k=1}^m$, which, due to the optimization, are consistent w.r.t scaled rotations. Next, using Corollary 1 with additional block normalizing at step 3, three rotations $\{R_1^{\tau(k)}, R_2^{\tau(k)}, R_3^{\tau(k)}\}$ and three translations $\{t_1^{\tau(k)}, t_2^{\tau(k)}, t_3^{\tau(k)}\}$ are extracted from each $E_{\tau(k)}$, which are uniquely defined up to a similarity transformation. Now, any two triplets in τ that share two cameras a, b agree on E_{ab}. Since the sign of E_{ab} is fixed, it determines the cameras a, b up to 2 unique configurations [10]. Therefore, each one of the two triplets must agree with one of the two configurations. As a result, assuming that both triplets defines the same configuration for a and b, there is a unique similarity transformation between the two triplets. In practice, in our experiments we observe that this is always the case.

	Our Method		**Chatterjee et al. [5]**		**Martinec et al. [17]**	
Data Set	R_f	R_d	R_f	R_d	R_f	R_d
Vienna Cathedral	**0.1141**	**4.7328**	0.1514	7.8472	-	-
Piazza del Popolo	**0.0595**	**2.4098**	0.2287	11.6022	0.5901	27.2359
NYC Library	**0.1200**	**4.8751**	0.1226	6.0765	0.5395	28.4385
Alamo	**0.0751**	**3.0489**	0.0879	4.4958	0.1503	7.3180
Metropolis	**0.30**	**15.6613**	0.4612	26.5636	1.1381	58.6143
Yorkminster	**0.1499**	**6.6343**	0.1526	7.8017	1.3434	73.3679
Montreal ND	**0.0608**	**2.4652**	0.1049	5.8742	0.2419	11.928
Tower of London	**0.1250**	**5.0731**	0.1366	5.8872	0.3435	16.6457
Ellis Island	0.0636	2.5784	**0.0499**	**2.2486**	0.0616	2.4955
Notre Dame	**0.0619**	**2.5091**	0.0876	4.7463	0.2109	10.5894

Table 1. Errors in estimated camera orientations, compared to ground truth measurements. R_f denotes the mean Frobenius norm error, averaged over the different cameras, and R_d is the mean angular error in degrees. Empty cells represent missing information.

By the construction process described in Sec. 4.2, the collection of triplets τ form a connected triplet graph. It is therefore possible to traverse the graph and apply a similarity transformation on the three cameras of each new node $\tau(k)$, and as a result bring all the cameras to a common Euclidean frame.

5. Experiments

To evaluate our approach, we implemented the SfM pipeline described next and tested it on ten challenging collections of unordered internet photographs of various sizes from [25]. Each dataset is provided with ground truth camera parameters. We use our method to recover camera parameters and compare them to the parameters obtained with existing methods, before and after bundle adjustment (BA).

5.1. SfM pipeline

The input to our algorithm is a set of independently estimated pairwise essential matrices, along with the number of inlier matches for each pairwise essential matrix. We build a triplet graph G_T as we describe in Sec. 4.2, removing any triplet whose (a) collinearity score is below 0.17 radians, (b) rotation consistency score exceeds 1.1, or (c) translation consistency score exceeds 1 radians. The final connected graph G_T defines the collection τ of triplets of cameras.

Next we apply our optimization algorithm as is described in Sec. 4.1. During optimization we observed that the eigenvalues of B_k and D_k (10) were always distinct for $k = 1, \ldots, m$. This means that the optimization variables, $E_{\tau(k)}$, indeed converge to 3-view consistent essential matrices with distinct eigenvalues. At this stage, we follow Corollary 1 to recover camera positions and orientations for each triplet of cameras and align all the recovered camera matrices by similarity transformations, as is described in

		Our Method					LUD [18]					1DSFM [25]				Cui [6]			
Data set	N_c	$\bar{x}$	$\tilde{x}$	$\bar{x}_{BA}$	$\tilde{x}_{BA}$	N_r	$\bar{x}$	$\tilde{x}$	$\bar{x}_{BA}$	$\tilde{x}_{BA}$	N_r	$\tilde{x}$	$\bar{x}_{BA}$	$\tilde{x}_{BA}$	N_r	$\tilde{x}$	$\bar{x}_{BA}$	$\tilde{x}_{BA}$	N_r
Vienna Cathedral	836	**9.6**	4.2	5.4	1.2	674	10	5.4	10	4.4	750	6.6	2e4	**0.5**	757	**3.5**	**4.0**	2.6	578
Piazza del Popolo	328	7.2	3.5	**2.5**	**0.8**	275	**5**	**1.5**	4	1.0	305	3.1	200	2.6	303	2.6	3.2	2.4	294
NYC Library	332	**3.3**	2.2	**1.1**	0.47	277	6	**2.0**	7	1.4	320	2.5	20	**0.4**	292	1.4	6.9	0.9	288
Alamo	577	2.5	1.2	**0.8**	0.35	482	**2**	**0.4**	2	**0.3**	547	1.1	2e7	**0.3**	521	0.6	3.7	0.5	500
Metropolis	341	15.2	6.9	**2.7**	1.4	168	**4**	**1.6**	4	1.5	288	9.9	70	**1.2**	288	-	-	-	-
Yorkminster	437	5.6	**2.7**	**1.9**	0.8	341	**5**	**2.7**	4	1.3	404	3.4	500	**0.2**	395	3.7	14	3.8	341
Montreal ND	450	1.9	1.0	**0.6**	**0.4**	416	**1**	**0.5**	1	**0.4**	435	2.5	1	0.9	425	0.8	1.1	**0.4**	426
Tower of London	572	**11.6**	5.0	**4**	1.0	414	20	**4.7**	10	3.3	425	11	40	**0.4**	414	4.4	6.2	1.1	393
Ellis Island	227	14.1	6.1	5.3	1.7	211	-	-	-	-	-	3.7	40	**0.4**	213	3.1	**1.8**	0.6	211
Notre Dame	553	1.8	0.8	**0.4**	**0.2**	529	**0.8**	**0.3**	0.7	**0.2**	536	10	7	2.1	500	**0.3**	0.8	**0.2**	539

Table 2. Camera positions error in meters evaluated on the data sets of [25]. N_c is the number of images in each dataset, $\bar{x}, \tilde{x}$ are the mean and median error respectively before bundle adjustment, and $\bar{x}_{BA}, \tilde{x}_{BA}$ are the mean and median errors after bundle adjustment. N_r are the number of reconstructed cameras. Empty cells represent missing information.

	Our Method			LUD [18]			1DSFM [25]			Cui [6]		
Data set	T_{R+T}	T_{BA}	T_{Tot}	T_{R+T}	T_{BA}	T_{Tot}	T_{R+T}	T_{BA}	T_{Tot}	T_{R+T}	T_{BA}	T_{Tot}
Vienna Cathedral	**68**	293	**566**	787	208	1467	323	3611	3934	-	717	959
Piazza del Popolo	**26**	27	**87**	88	31	162	42	213	255	-	93	144
NYC Library	**28**	58	125	102	47	200	47	382	429	-	48	**90**
Alamo	**47**	155	**327**	385	133	750	152	646	798	-	362	621
Metropolis	**16**	70	**93**	67	38	142	47	224	271	-	-	-
Yorkminster	**33**	116	207	103	148	297	71	955	1026	-	63	**108**
Montreal ND	**41**	170	494	271	167	553	93	1043	1136	-	226	**351**
Tower of London	**41**	120	241	88	86	228	61	750	811	-	121	**221**
Ellis Island	**21**	53	140	-	-	-	29	276	305	-	64	**95**
Notre Dame	**52**	277	**720**	707	126	1047	205	2139	2344	-	793	1159

Table 3. Runtime in seconds for results in Table 2. T_{R+T} denotes the time for motion averaging (for other methods rotation and translation estimation). T_{BA} the time for bundle adjustment and T_{Tot} is the total running time of the method, including the additional time for building the triangle cover. Empty cells represent image collections not tested by the authors. In addition, Cui [6] does not report results before BA.

Sec. 4.3. This yields a set of cameras with absolute positions and orientations. The obtained camera parameters are finally refined using BA.

5.2. Results

To evaluate our recovered camera orientations, we compare our results to those obtained with the methods of Chatterjee et al. [5] and Martinec et al. [17]. For a fair comparison we evaluate these methods on same subset of images used in our method. Moreover, since in contrast to our method these solvers do not estimate camera positions we evaluate the results before BA. The results are summarized in Table 1. Our method outperforms these two solvers in nine out of the ten datasets.

To evaluate our recovered camera positions we compare our method to the following position solvers: Cui et al. [6], 1DSFM [25] and LUD [18]. The results are summarized in Table 2. Note that [18, 6] use point matches in their pipelines, while both our method and [25] do not use point correspondences until the final BA. In general, the latter approaches allow for faster optimization, but result in inaccuracies before BA. On the other hand, it allows for greater improvement in the final BA, compared to [18, 6]. Indeed, as can be seen in the table, while our method surpasses all the other methods before bundle adjustment in three out of the ten datasets (according to mean error), it achieves superior performance on eight out of ten after BA.

Table 3 further compares execution times, before and after BA, showing that our method is efficient, compared to existing methods.

5.3. Technical details

For Bundle Adjustment we used the Theia standard SfM library [21]. Camera position results for [18, 6, 25] in Tables 2 and 3 are taken from the papers. We ran our experiments on an Intel(R)-i7 3.60GHz with Windows. Bundle Adjustment was performed on a Linux machine Intel(R) Xeon(R) CPU @ 2.30GHz with 16 cores.

6. Conclusion

We have provided in this paper algebraic conditions for the consistency of essential matrices in multiview settings and an algorithm for their averaging given noisy and partial measurements. In future research we will seek to further incorporate collinear camera triplets in the averaging algorithm, explore numerical properties, and design online consistency enforcement algorithms for SLAM settings.

Acknowledgment This research was supported in part by the Minerva foundation with funding from the Federal German Ministry for Education and Research.

References

[1] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M Seitz, and Richard Szeliski. Building rome in a day. In *2009 IEEE 12th international conference on computer vision*, pages 72–79. IEEE, 2009.

[2] Chris Aholt and Luke Oeding. The ideal of the trifocal variety. *Mathematics of Computation*, 83(289):2553–2574, 2014.

[3] Mica Arie-Nachimson, Shahar Z Kovalsky, Ira Kemelmacher-Shlizerman, Amit Singer, and Ronen Basri. Global motion estimation from point matches. In *2012 Second Int. Conf. on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 81–88. IEEE, 2012.

[4] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, 3(1):1–122, 2011.

[5] Avishek Chatterjee and Venu Madhav Govindu. Robust relative rotation averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):958–972, 2018.

[6] Zhaopeng Cui, Nianjuan Jiang, Chengzhou Tang, and Ping Tan. Linear global translation estimation with feature tracks. *ArXiv preprint, ArXiv:1503.01832*, 2015.

[7] Anders Eriksson, Carl Olsson, Fredrik Kahl, and Tat-Jun Chin. Rotation averaging and strong duality. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 127–135, 2018.

[8] Thomas Goldstein, Paul Hand, Choongbum Lee, Vladislav Voroninski, and Stefano Soatto. Shapefit and shapekick for robust, scalable structure from motion. In *European Conf. on Computer Vision*, pages 289–304. Springer, 2016.

[9] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013.

[10] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[11] Anders Heyden and Kalle Åström. Algebraic properties of multilinear constraints. *Mathematical Methods in the Applied Sciences*, 20(13):1135–1162, 1997.

[12] Robert J. Holt and Arun N. Netravali. Uniqueness of solutions to three perspective views of four points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):303–307, 1995.

[13] Nianjuan Jiang, Zhaopeng Cui, and Ping Tan. A global linear method for camera pose registration. In *IEEE Int. Conf. on Computer Vision*, pages 481–488, 2013.

[14] Yoni Kasten, Meirav Galun, and Ronen Basri. Resultant based incremental recovery of camera pose from pairwise matches. *arXiv preprint arXiv:1901.09364*, 2019.

[15] Yoni Kasten, Amnon Geifman, Meirav Galun, and Ronen Basri. Gpsfm: Global projective sfm using algebraic constraints on multi-view fundamental matrices. *ArXiv preprint, ArXiv:1812.00426*, 2018.

[16] Manfred Klopschitz, Arnold Irschara, Gerhard Reitmayr, and Dieter Schmalstieg. Robust incremental structure from motion. In *Proc. 3DPVT*, volume 2, pages 1–8, 2010.

[17] Daniel Martinec and Tomas Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[18] Onur Ozyesil and Amit Singer. Robust camera location estimation by convex programming. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2674–2683, 2015.

[19] Soumyadip Sengupta, Tal Amir, Meirav Galun, Tom Goldstein, David W Jacobs, Amit Singer, and Ronen Basri. A new rank constraint on multi-view fundamental matrices, and its application to camera location recovery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4798–4806, 2017.

[20] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.

[21] Chris Sweeney. Theia multiview geometry library: Tutorial & reference. http://theia-sfm.org.

[22] Chris Sweeney, Torsten Sattler, Tobias Hollerer, Matthew Turk, and Marc Pollefeys. Optimizing the viewing graph for structure-from-motion. In *IEEE Int. Conf. on Computer Vision*, pages 801–809, 2015.

[23] Roberto Tron and Kostas Daniilidis. The space of essential matrices as a Riemannian quotient manifold. *SIAM Journal on Imaging Sciences*, 10(3):1416–1445, 2017.

[24] Roberto Tron and René Vidal. Distributed image-based 3-d localization of camera sensor networks. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 901–908. IEEE, 2009.

[25] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In *European Conf. on Computer Vision*, pages 61–75. Springer, 2014.

[26] Changchang Wu. Towards linear-time incremental structure from motion. In *2013 International Conference on 3D Vision-3DV 2013*, pages 127–134. IEEE, 2013.