On the Instability of Relative Pose Estimation and RANSAC's Role

Hongyi Fan
School of Engineering
Brown University
hongyi_fan@brown.edu

Joe Kileel
Department of Mathematics
University of Texas at Austin
jkileel@math.utexas.edu

Benjamin Kimia
School of Engineering
Brown University
benjamin_kimia@brown.edu

Abstract

Relative pose estimation using the 5-point or 7-point Random Sample Consensus (RANSAC) algorithms can fail even when no outliers are present and there are enough inliers to support a hypothesis. These cases arise due to numerical instability of the 5- and 7-point minimal problems. This paper characterizes these instabilities, both in terms of minimal world scene configurations that lead to infinite condition number in epipolar estimation, and also in terms of the related minimal image feature pair correspondence configurations. The instability is studied in the context of a novel framework for analyzing the conditioning of minimal problems in multiview geometry, based on Riemannian manifolds. Experiments with synthetic and real-world data reveal that RANSAC does not only serve to filter out outliers, but RANSAC also selects for well-conditioned image data, sufficiently separated from the ill-posed locus that our theory predicts. These findings suggest that, in future work, one could try to accelerate and increase the success of RANSAC by testing only well-conditioned image data.

1. Introduction

The past two decades have seen an explosive growth of multiview geometry applications such as the reconstruction of 3D object models for use in video games [1], film [19], archaeology [28], architecture [22], and urban modeling (*e.g.*, Google Street View); match-moving in augmented reality and cinematography for mixing virtual content and real video [11]; the organization of a collection of photographs with respect to a scene known as Structure-from-Motion [27] (*e.g.*, as pioneered in photo tourism [2]); robotic manipulation [16]; and meteorology from cameras in automobile manufacture and autonomous driving [22]. One key building block of a multiview system is the relative pose estimation of two cameras [15, 35]. A methodology that is dominant in applications is RANSAC [29]. This forms hypotheses from a few randomly selected correspondences in two views, *e.g.*, 5 in calibrated camera pose estimation [26] and 7 in uncalibrated camera pose estimation [31, 33], and validates these hypotheses using the remaining putative correspondences. The chief stated reason for using RANSAC is robustness against outliers, see [3, 5, 24]. The pose of multiple cameras can then be recovered in either a locally incremental [30] or globally averaging manner [18]. This approach has been quite successful in many applications.

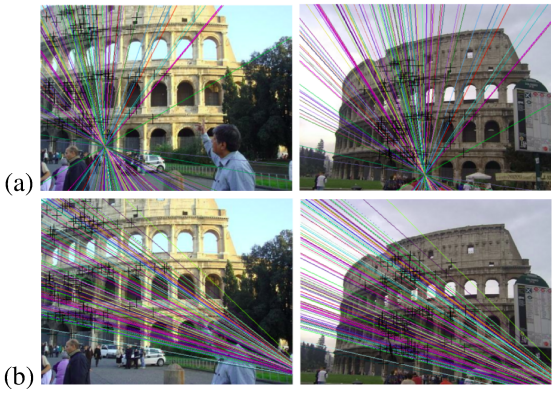

Figure 1. The typical relative pose estimation can fail catastrophically, even with a large number of correspondences (100 correspondences shown in the figure) all of which are inliers. (a) Ground truth epipolar geometry. (b) Erroneous estimated epipolar geometry from the 7-point algorithm and LO-RANSAC [8]. The prime cause of such failure is numerical instability as shown in this paper.

There are, however, a non-negligible number of scenarios where this RANSAC-based approach fails, *e.g.*, in producing the relative pose between two cameras. As an example, when the number of candidate correspondences drops to say 50 to 100 correspondences, as is the case for images of homogeneous and low textured surfaces, the pose estimation process fails. Similarly, when there is repeated texture in the scene, there is a large number of outlier candidate correspondences and again the process can fail. It is curious why the estimation should fail, even if only a few correspondences are available: after all RANSAC can select 5 from 50 in $\binom{50}{5} \approx 2.1$ million combinations, so there are plenty

of veridical correspondences available if the ratio of outliers is low. In an experiment with *no outliers*, either with synthetic data (see Section 5) or real data (see Figure 1), the process can still frequently fail! This is mysterious, unless the role of RANSAC goes beyond weeding out the outliers. Indeed, we will argue that a main role for RANSAC is to *stabilize* the estimation process, without denying its role in dealing with outliers. We will show that the process of estimating pose from a minimal problem is typically unstable to the noise, with a gradation of instability depending on the specific choice of 5 points or 7 points. The role of RANSAC is to integrate non-selected correspondences to improve the stability of the estimation process outcome: if a large number of non-selected candidate correspondences agree, then the hypothesis is both free of outliers but – perhaps more importantly – it is an estimate stable to typical image noise.

This paper inspects the general issue of numerical stability in minimal problems in multiview geometry. We build a framework that connects well-conditioned minimal point configurations with the condition number of the inverse Jacobian of a forward projection map. Using this framework, we compute condition number formulas for the 5-point and 7-point minimal problems. Further, we investigate the issue of ill-posedness, *i.e*. when the condition number is infinite. We obtain characterizations for a world scene to be ill-posed, and requirements for a minimal image point correspondences configuration to be ill-posed.

Much analysis for the degeneracy of two-view geometry has already appeared, *e.g*., [14,17,20,23]. However, this literature has studied when there are multiple solutions to the 3D reconstruction problem. By contrast, we focus on when there exists an unboundedly unstable solution. This analysis is different from previous literature in its focus on minimal problems, where one typically has multiple real solutions. Thus, our theory applies to multiview geometry as it is used in practice: minimal problems solved during RANSAC.

Along with this theoretical analysis, we propose a way to measure the stability of a given minimal image point correspondence set, namely, by measuring the distance from one point on one image to a "degenerate curve" on this image computed using the other point correspondences. This distance gives a means of evaluating the stability of the minimal hypothesis. Our proposal to gauge the stability of a given hypothesis suggests a way to increase the speed and the robustness of RANSAC: by only testing hypotheses which come from sufficiently well-conditioned image data.

The rest of the paper is organized as follows. Section 2 reviews two classic problems for estimating the relative pose, namely the 5-point problem for calibrated cameras and 7-point problem for uncalibrated cameras. Section 3 introduces a novel theoretical framework for analyzing the conditioning of an arbitrary minimal problems. Section 4 presents the specific results of our analysis for relative pose estimation, characterizing ill-posed world scenes and image data as well as proposing a potential way for testing for well-conditioned image data. Finally, Section 5 shows experimental results on synthetic and real data, as a proof-of-concept for our theory and its connection to RANSAC.

2. Minimal Problems for Relative Pose Estimation

This section reviews the relevant setup for the 5-point and 7-point minimal problems. The terms in bold will be used again in the general framework of the next section.

Essential Matrices and the 5-Point Problem: Let $\mathcal{W}$ denote **world scene space**, consisting of the relative pose between two calibrated pinhole cameras together with five world points:

$$\mathcal{W} = \mathrm{SO}(3) \times \mathbb{S}^2 \times (\mathbb{R}^3)^{\times 5} = \{(\mathbf{R}, \hat{\mathbf{T}}, \Gamma_1, \dots, \Gamma_5)\}. \quad (1)$$

Here $\mathrm{SO}(3) = \{\mathbf{R} \in \mathbb{R}^{3\times 3} : \mathbf{R}\mathbf{R}^\top = \mathbf{R}^\top\mathbf{R} = I\}$ is the group of rotation matrices, $\mathbb{S}^2 = \{\hat{\mathbf{T}} \in \mathbb{R}^3 : \|\hat{\mathbf{T}}\|_2 = 1\}$ is the unit sphere (representing the direction of the translation in the relative pose) and Γ_i (for $i \in \{1, .., 5\}$) are the 3D points. Meanwhile, let $\mathcal{X}$ denote the **image data space**, consisting of five image point correspondences:

$$\mathcal{X} = \left(\mathbb{R}^2 \times \mathbb{R}^2\right)^{\times 5} = \{(\gamma_1, \bar{\gamma}_1), \dots, (\gamma_5, \bar{\gamma}_5)\}, \quad (2)$$

where $\gamma_i \in \mathbb{R}^2$ and $\bar{\gamma}_i \in \mathbb{R}^2$ represent corresponding points on the two image planes. Next, let Φ denote the **forward map**, projecting the given world points via the calibrated cameras $[I\ 0] \in \mathbb{R}^{3\times 4}$ and $[\mathbf{R}\ \hat{\mathbf{T}}] \in \mathbb{R}^{3\times 4}$, *i.e*.,

$$\Phi\left(\mathbf{R}, \hat{\mathbf{T}}, \Gamma_1, \dots, \Gamma_5\right) = ((\gamma_1, \bar{\gamma}_1), \dots, (\gamma_5, \bar{\gamma}_5)), \quad (3)$$

where $\gamma_i = \pi(\Gamma_i)$ and $\bar{\gamma}_i = \pi(\mathbf{R}\Gamma_i + \hat{\mathbf{T}})$ where π is projection of the 3D points onto 2D images. The relation between corresponding points on the two images is captured via essential matrix E as $\bar{\gamma}^T E \gamma = 0$, where $E \in \mathbb{R}^{3\times 3}$. Thus, we define the **epipolar space** as the manifold of real essential matrices, which are characterized by ten cubic equations vanishing [9] or in terms singular values as follows:

$$\begin{aligned}\mathcal{Y} &= \{E \in \mathbb{P}(\mathbb{R}^{3\times 3}) : 2EE^\top E - \mathrm{tr}(EE^\top)E = 0, \det(E) = 0\} \\ &= \{E \in \mathbb{P}(\mathbb{R}^{3\times 3}) : \sigma_1(E) = \sigma_2(E) > \sigma_3(E) = 0\}. \end{aligned} \quad (4)$$

Last, the **epipolar map** Ψ is defined as computing an essential matrix from a world scene using the relative pose:

$$\Psi\left(\mathbf{R}, \hat{\mathbf{T}}, \Gamma_1, \dots, \Gamma_5\right) = E = [\hat{\mathbf{T}}]_\times \mathbf{R} \in \mathbb{P}(\mathbb{R}^{3\times 3}). \quad (5)$$

Here $[\hat{\mathbf{T}}]_\times \in \mathbb{R}^{3\times 3}$ is the usual skew matrix representation of cross product with the vector $\hat{\mathbf{T}} \in \mathbb{R}^3$, as in [15, Sec. 9.6].

Then, 5-point problem is the task of determining the possible essential matrices E given the five image point pairs, *i.e.*, computing $\Psi(\Phi^{-1}(\cdot))$. In [26], David Nister developed a solver for this problem. It boils down to computing the real roots of a degree 10 univariate polynomial, giving ≤ 10 real solutions for the essential matrix.

Fundamental Matrices and the 7-Point Problem: For the case of uncalibrated pinhole cameras, the **world scene space** $\mathcal{W}$ is the space of the relative poses together with seven world points:

$$\mathcal{W} = \mathbb{R}^{3\times4} \times \mathbb{R}^{3\times4} \times (\mathbb{R}^3)^{\times 7} = \{\mathcal{P}, \bar{\mathcal{P}}, \Gamma_1, \ldots, \Gamma_7)\}. \quad (6)$$

Here Γ_i (for $i \in \{1, .., 7\}$) are the 3D points and $\mathcal{P}$ and $\bar{\mathcal{P}}$ are the 3×4 projection matrices representing the two cameras, which are defined to be $\mathcal{P} = K[\mathbf{R}|\hat{\mathbf{T}}]$ where K is the intrinsic matrix of the camera. Then, the **image data space** $\mathcal{X}$ is the space of seven image point correspondences, *i.e.*, $\mathcal{X} = \{(\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1), \ldots, (\boldsymbol{\gamma}_7, \bar{\boldsymbol{\gamma}}_7)\}$. The **forward map** Φ is defined to project the world points via the uncalibrated cameras via

$$\Phi(\mathcal{P}, \bar{\mathcal{P}}, \Gamma_1, .., \Gamma_7) = ((\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1), \ldots, (\boldsymbol{\gamma}_7, \bar{\boldsymbol{\gamma}}_7)), \quad (7)$$

where $\boldsymbol{\gamma}_i = \pi(\mathcal{P}\Gamma_i)$ and $\bar{\boldsymbol{\gamma}}_i = \pi(\bar{\mathcal{P}}\Gamma_i)$ with π the projection from 3D to 2D image. The relation between image correspondences is described by a 3×3 rank-2 fundamental matrix F via $\bar{\boldsymbol{\gamma}}_i^T F \boldsymbol{\gamma}_i = 0$. Here the **epipolar space** consists of the manifold of real fundamental matrices *i.e.*,

$$\mathcal{Y} = \{F \in \mathbb{P}(\mathbb{R}^{3\times3}) : \operatorname{rank}(F) = 2\}. \quad (8)$$

The **epipolar map** Ψ sends a world scene to the fundamental matrix associated to the projection matrices [15, Eq. 9.1].

Then, the 7-point problem is the task of determining the possible fundamental matrices F given the seven image point pairs, *i.e.*, computing $\Psi(\Phi^{-1}(\cdot))$. The solutions are obtained by computing the real roots of a cubic univariate polynomial, see [15, Sec. 11.1.2].

3. Theoretical Framework

In this section, we present a novel theoretical framework for analyzing the numerical stability of minimal problems in multiview geometry, which generalizes the notation defined in Section 2. The relevant mathematical structure are Riemannian manifolds, which we use to describe the totality of world scenes, image data, and epipolar quantities to be estimated. Riemannian geometry helps, because it allows us to discuss intrinsic distances. Our approach uses tangent spaces, differentials, and the inverse function theorem.

We build on the theory of condition number and ill-posed inputs initiated by Demmel [10], and then extended by Burgisser [7]. We tailored the theory to the setting of minimal problems, where there exist world scenes "in-between" the input image data and the output epipolar quantities.

3.1. Spaces and Maps

Let $\mathcal{W}, \mathcal{X}, \mathcal{Y}$ be Riemannian manifolds, with geodesic distances $d_{\mathcal{W}}(\cdot,\cdot)$, $d_{\mathcal{X}}(\cdot,\cdot)$, $d_{\mathcal{Y}}(\cdot,\cdot)$, tangent spaces denoted by $T(\mathcal{W}, w)$, $T(\mathcal{X}, x)$, $T(\mathcal{Y}, y)$ for points $w \in \mathcal{W}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and inner products on said tangent spaces denoted by $\langle\cdot,\cdot\rangle_{\mathcal{W},w}$, $\langle\cdot,\cdot\rangle_{\mathcal{X},x}$, $\langle\cdot,\cdot\rangle_{\mathcal{Y},y}$. In applications to multiview geometry, we refer to these as

- $\mathcal{W}$ the **world scene space**;
- $\mathcal{X}$ the **image data space**;
- $\mathcal{Y}$ the **epipolar space**.

We restrict to the case $\dim(\mathcal{W}) = \dim(\mathcal{X})$ to model minimal problems in multiview geometry. See Remark 1 below.

Next, assume we are given a differentiable map Φ from world scenes to image data whose domain is an open dense subset of $\mathcal{W}$. We indicate the situation using a dashed right arrow:

$$\Phi : \mathcal{W} \dashrightarrow \mathcal{X}. \quad (9)$$

Assume that the image $\Phi(\mathrm{Dom}(\Phi))$ contains an open dense subset of the codomain $\mathcal{X}$; we summarize this property by calling Φ dominant. We call Φ the **forward map**.

Furthermore, assume that we are provided with a differentiable map from world scenes to epipolar matrices, again defined only on an open dense subset of $\mathcal{W}$:

$$\Psi : \mathcal{W} \dashrightarrow \mathcal{Y}. \quad (10)$$

Again, assume Ψ is dominant. We call Ψ the **epipolar map**.

Now given image data $x \in \mathcal{X}$, we call a function $\Theta : \mathcal{X} \supseteq \mathrm{Dom}(\Theta) \to \mathcal{W}$ a **3D reconstruction map** locally defined around x if $\mathrm{Dom}(\Theta)$ is an open neighborhood of x in $\mathcal{X}$ and Θ is a section of the forward map, that is

$$\Phi \circ \Theta = \mathrm{id}_{\mathrm{Dom}(\Theta)}. \quad (11)$$

In this case, composing Θ with the epipolar map gives a (locally defined) map from image data to the epipolar space:

$$\mathbf{S} := \Psi \circ \Theta : \mathcal{X} \supseteq \mathrm{Dom}(\Theta) \to \mathcal{Y}. \quad (12)$$

We call $\mathbf{S}$ a **solution map** (locally defined around x). The name makes sense because, in minimal problems in vision, the quantity we want to compute is typically an epipolar matrix/tensor while the input is typically image data.

Remark 1 *Assume the above setup. Then minimal problems in multiview geometry are modeled as follows: given image data $x \in \mathcal{X}$, we want to compute all compatible real epipolar matrices/tensors, that is*

$$\Psi(\Phi^{-1}(x)) = \{\Psi(w) : w \in \mathcal{W}, \Phi(w) = x\} \subseteq \mathcal{Y}. \quad (13)$$

These solutions become hypotheses in RANSAC. When we call a problem "minimal", what we mean is the following.

For x in an open dense subset of $\mathcal{X}$, the output $\Psi(\Phi^{-1}(x))$ is a finite set (and not always empty). Often minimality is a consequence of additional structure, which is not required for much of this paper. Typically $\mathcal{W}, \mathcal{X}, \mathcal{Y}$ can be viewed as quasi-projective algebraic varieties and Φ, Ψ are algebraic functions [13]. Then due to the assumption $\dim(\mathcal{W}) = \dim(\mathcal{X})$ and the dominance of Φ, general facts in algebraic geometry imply that generic fibers of Φ are finite sets, so the problem is minimal. See e.g. *[12, Def. 2].*

We want to analyze how sensitive the output of the solution map $\mathbf{S}(x)$ is to realistic levels of noise in the input x. We want to develop quantitative condition number formulas and describe the locus of ill-posed inputs, where a solution map may not even exist locally or has infinite condition number.

3.2. Ill-Posed Locus

Given image data $x \in \mathcal{X}$ and a prescribed world scene $w \in \mathcal{W}$ such that $\Phi(w) = x$, the next lemma shows there exists a unique continuous 3D reconstruction map Θ with $\Theta(x) = w$. Further, Θ is continuously differentiable (C^1).

Lemma 1 *Assume that the forward map Φ is C^1, and that at the world scene $w \in \mathcal{W}$ the forward map differentiates to an isomorphism on tangent spaces. That is, the differential*

$$D\Phi(w) : T(\mathcal{W}, w) \to T(\mathcal{X}, \Phi(w)) \tag{14}$$

is a linear isomorphism. Then there exist open neighborhoods $\mathcal{U}$ of w in $\mathcal{W}$ and $\mathcal{V}$ of $\Phi(w)$ in $\mathcal{X}$ such that $\Phi : \mathcal{U} \to \mathcal{V}$ is bijection, the inverse function is C^1, and

$$D\left((\Phi|_{\mathcal{U}})^{-1}\right)(\Phi(w)) = (D\Phi(w))^{-1}. \tag{15}$$

The lemma follows from the inverse function theorem for manifolds [21]. In words: if the forward Jacobian $D\Phi(w)$ is invertible, then the forward map Φ is locally invertible and its local inverse is differentiable with Jacobian $(D\Phi(w))^{-1}$.

We now come to a central concept in our framework:

Definition 1 *We say that a world scene $w \in \mathcal{W}$ is* **ill-posed** *if the differential $D\Phi(w)$ is not invertible. We say that image data $x \in \mathcal{X}$ is* **ill-posed** *if there exists a world scene $w \in \Phi^{-1}(x)$ such that w is ill-posed.*

Ill-posed world scenes are those failing the condition in the above lemma; therefore, a priori we do not know if the forward map is locally invertible around ill-posed world scenes. Meanwhile ill-posed image data are those such that there is at least one compatible world scene that is ill-posed; hence, there could be problematic behavior around an ill-posed world scene (We emphasize that other world scenes in $\Phi^{-1}(x)$ need not be ill-posed). In a moment, we will see that all of the numerical instabilities in minimal problems must occur at (or near) the ill-posed scenes and image data.

3.3. Condition Number

Our other central theoretical concept is the condition number. We first explain this quite generally (and intuitively), following [7, Ch. 14]. To this end, let $G : \mathcal{X} \supseteq \mathrm{Dom}(G) \to \mathcal{Y}$ be any map defined on an open neighborhood of x in $\mathcal{X}$.

Definition 2 *The* **condition number** *of G at x is defined by*

$$\mathrm{cond}(G, x) := \lim_{\delta \to 0^+} \sup_{\substack{\widetilde{x} \in \mathcal{X} \\ d_{\mathcal{X}}(\widetilde{x}, x) < \delta}} \frac{d_{\mathcal{Y}}\left(G(\widetilde{x}), G(x)\right)}{d_{\mathcal{X}}\left(\widetilde{x}, x\right)}. \tag{16}$$

In a slogan: the condition number captures the limiting *worst-case amplification* of input error in x that the function G can produce in its output $G(x)$, when distances are measured according to the intrinsic metrics on $\mathcal{X}$ and $\mathcal{Y}$.

If G is differentiable, we have a more explicit formula.

Lemma 2 *If G is differentiable then the condition number of G at x equals the operator norm of the differential $DG(x) : T(\mathcal{X}, x) \to T(\mathcal{Y}, y)$,* i.e.

$$\mathrm{cond}(G, x) = \max_{\substack{\dot{x} \in T(\mathcal{X}, x) \\ \|\dot{x}\| = 1}} \|DG(x)(\dot{x})\| =: \|DG(x)\|, \tag{17}$$

where the two norms in the middle quantity are induced by the Riemannian inner products $\langle \cdot, \cdot \rangle_{\mathcal{X}, x}$ and $\langle \cdot, \cdot \rangle_{\mathcal{Y}, G(x)}$.

This is [7, Prop. 14.1], and proven using Taylor's theorem. The lemma reduces computing the condition number of a differentiable map to computing the leading singular value of its Jacobian matrix written with respect to orthonormal bases on the tangent spaces $T(\mathcal{X}, x)$ and $T(\mathcal{Y}, G(x))$.

Here we are most interested in the condition number of solution maps for minimal problems as in Eq. (12). Putting the previous two lemmas together with the chain rule gives:

Lemma 3 *Let $\mathbf{S} = \Psi \circ \Theta : \mathcal{X} \supseteq \mathrm{Dom}(\Theta) \to \mathcal{Y}$ be a solution map as in (12) defined around the image data $x \in \mathcal{X}$. Let $w = \Theta(x) \in \mathcal{W}$ be the corresponding world scene. Assume that w is not ill-posed,* i.e. *$D\Phi(w)$ is invertible. Then, the condition number of $\mathbf{S}$ at x is finite and given by*

$$\mathrm{cond}(\mathbf{S}, x) = \|D\Psi(w) \circ D\Phi(w)^{-1}\|. \tag{18}$$

In particular, $\mathrm{cond}(\mathbf{S}, x)$ can be infinite only if x is ill-posed.

3.4. Relation Between Ill-Posed Loci and Condition Number

As shown in Lemma 3, the condition number at $x \in \mathcal{X}$ of a varying epipolar matrix/tensor can be infinite only if x is ill-posed as in Definition 1. If x is ill-posed, the corresponding world scene $w = \Theta(x) \in \mathcal{W}$ such that $D\Phi(w)$

is rank-deficient might suffer unboundedly large relative changes as x changes. Further, Lemma 1 implies that the number of real 3D reconstructions is locally constant for inputs $x \in \mathcal{X}$ which are not ill-posed. In other words, there can only be a change in the number of real epipolar matrices/tensors when the image data x crosses over the ill-posed locus. Thus, the ill-posed locus captures the "danger zone" where at least one of the solutions to the minimal problem can be unboundedly unstable, and also where real solutions can disappear into (or reappear from) the complex numbers.

In [10], Demmel proved that in some cases, the reciprocal of the distance to the ill-posed locus equals the condition number. For example, this was shown for the problem of matrix inversion. Here, we do not prove a quantitative relationship between the distance to the ill-posed locus and the condition number for solving minimal problems in computer vision as such. But we do numerically demonstrate a close relationship in the case of essential and fundamental matrix estimation in the experiments in Section 5.

4. Main Results

We now present our main theoretical results regarding the instabilities of relative pose estimation, by applying the framework in Section 3 to the minimal problems in Section 2. Due to space limitations, the proofs (and certain explicit formulas) will appear in the supplementary materials.

4.1. Condition Number Formulas

Here we apply the formula (18) based on singular values of the Jacobian matrix to the 5-point and 7-point problems. This yields condition number formulas for essential and fundamental estimation. The expressions are valid if the solution maps passes through non-ill-posed world scenes; in fact they only depend on said world scene. We display the explicit Jacobian matrices in the supplementary materials.

Proposition 1 (Condition number for E) *Consider the 5-point problem in Section 2. Let $x \in (\mathbb{R}^2 \times \mathbb{R}^2)^{\times 5}$ be given image data, and $w \in \mathrm{SO}(3) \times \mathbb{S}^2 \times (\mathbb{R}^3)^{\times 5}$ a compatible world scene which is not ill-posed. Then there exists a unique continuous 3D reconstruction map Θ locally defined around x such that $\Theta(x) = w$, and an associated uniquely defined solution map $\mathbf{S} = \Psi \circ \Theta$ from image data to essential matrices. The condition number of $\mathbf{S}$ can be computed as the largest singular value of an explicit 5×20 matrix whose entries are functions of w. This matrix naturally factors as a 5×20 matrix multiplied by a 20×20 matrix.*

Proposition 2 (Condition number for F) *Consider the 7-point problem in Section 2. Let $x \in (\mathbb{R}^2 \times \mathbb{R}^2)^{\times 7}$ be given image data, and $w \in \mathbb{R}^{3\times 4} \times \mathbb{R}^{3\times 4} \times (\mathbb{R}^3)^{\times 7}$ a compatible world scene which is not ill-posed. Then there exists a unique continuous 3D reconstruction map Θ locally defined around x such that $\Theta(x) = w$, and an associated uniquely defined solution map $\mathbf{S} = \Psi \circ \Theta$ from image data to essential matrices. The condition number of $\mathbf{S}$ can be computed as the largest singular value of an explicit 7×28 matrix whose entries are functions of w. This matrix naturally factors as a 7×28 matrix multiplied by the inverse of a 28×28 matrix.*

4.2. Ill-Posed World Scenes

Here we derive geometric conditions for a world scene to be ill-posed for the 5-point or 7-point problem. Our characterizations are in terms of the existence of a particular type of quadric surface in $\mathbb{R}^3$, are which should satisfy certain properties related to the given world scene.

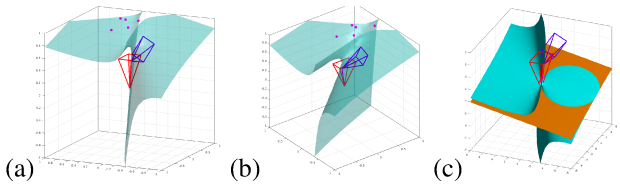

Figure 2. An illustrative example of an ill-posed world scene in the calibrated case. Red and blue pyramid represents two cameras. Magenta points represent the given world points. The green surface is the quadric surface satisfying the three conditions in Theorem 1. (a) and (b) shows two different view angles. Last, a zoomed-out view (c) shows an orange plane perpendicular to the baseline whose intersection with the quadric surface is a circle.

Theorem 1 (Ill-posed world scenes for E) *Consider the 5-point problem in Section 2. Let $w = (\mathbf{R}, \hat{\mathbf{T}}, \Gamma_1, \ldots, \Gamma_5) \in \mathrm{SO}(3) \times \mathbb{S}^2 \times (\mathbb{R}^3)^{\times 5}$ be a world scene such that $\Phi(w)$ exists where Φ is as in Eq. (3). Then w is ill-posed,* i.e. *$D\Phi(w)$ is rank-deficient, if and only if there exists a quadric surface $\mathcal{Q} \subseteq \mathbb{R}^3$ such that:*

- *$\mathcal{Q}$ passes through the given world points $\Gamma_1, \ldots, \Gamma_5$;*
- *$\mathcal{Q}$ contains the baseline of the given relative pose;*
- *and intersecting $\mathcal{Q}$ with any normal affine plane to ℓ produces a circle.*

Here the baseline $\ell \subseteq \mathbb{R}^3$ is the world line passing through the two camera centers, i.e. *$\ell = \mathrm{Span}(-R^\top \hat{\mathbf{T}})$.*

The second requirement implies that $\mathcal{Q}$ is a ruled quadric surface (*i.e.*, covered by an infinity family of lines). Meanwhile, the third item is a non-standard condition implying that $\mathcal{Q}$ must be special within the set of ruled quadric surfaces, namely it must be a so-called "rectangular quadric" [23]. See Figure 2 for visualizations of Theorem 1.

Theorem 2 (Ill-posed world scenes for F) *Consider the 7-point problem in Section 2. Let $w =$*

$(\mathcal{P}, \bar{\mathcal{P}}, \Gamma_1, \ldots, \Gamma_7) \in \mathbb{R}^{3\times 4} \times \mathbb{R}^{3\times 4} \times (\mathbb{R}^3)^{\times 7}$ *be a world scene such that* $\Phi(w)$ *exists where* Φ *is as in Eq. (7). Then* w *is ill-posed,* i.e. $D\Phi(w)$ *is rank-deficient, if and only if there exists a quadric surface* $\mathcal{Q} \subseteq \mathbb{R}^3$ *such that:*

- $\mathcal{Q}$ *passes through the given world points* $\Gamma_1, \ldots, \Gamma_7$*;*
- *and* $\mathcal{Q}$ *contains the baseline of the given relative pose.*

Here the baseline ℓ *is the world line passing through the two camera centers.*

Now the conditions on the quadric surface are the same as in Theorem 1, except the third condition (stemming from calibration) is absent. Our proofs for Theorems 1 and 2 both proceed by unwinding the requirement that there exists a nonzero kernel vector for the forward Jacobian matrix.

4.3. Ill-Posed Image Data

Here we describe the locus of ill-posed image data for the 5-point and 7-point problems. These results rely heavily on the polynomial structure present in both minimal problems (as mentioned in Remark 1). Specifically the proofs use known facts from algebraic geometry due to Sturmfels [34].

Compared to [34], the main contribution of this subsection is that we obtain viable computational schemes for actually visualizing the loci of ill-posed image data. For both the cases of fundamental and essential matrices, we give methods based on numerical homotopy continuation [32] to solve polynomial equations. Implemented in the Julia package HomotopyContinuation.jl [6], these terminate on a desktop computer in ≈ 10 and ≈ 30 seconds, respectively. Details are given in the supplementary materials.

Theorem 3 (Ill-posed image data for E) *Consider the 5-point problem in Section 2. Let* $((\gamma_1, \bar{\gamma}_1), \ldots, (\gamma_5, \bar{\gamma}_5)) \in (\mathbb{R}^2 \times \mathbb{R}^2)^{\times 5}$ *be image data. Then* x *is ill-posed,* i.e. *there exists some compatible world scene which is ill-posed, only if a certain polynomial* $\mathbf{P}$ *in the entries of* $\gamma_1, \bar{\gamma}_1, \ldots, \bar{\gamma}_5$ *vanishes. This polynomial has degree* 30 *separately in each of the points* $\gamma_1, \ldots, \bar{\gamma}_5$*. In particular, if we fix numerical values for* $\gamma_1, \bar{\gamma}_1, \ldots, \gamma_5$ *but keep* $\bar{\gamma}_5 \in \mathbb{R}^2$ *as variable, then (generically)* $\mathbf{P}$ *specializes to a degree* 30 *polynomial just in* y_5*, and its vanishing set is a degree* 30 *curve in the second image plane. Moreover given the values for* $\gamma_1, \bar{\gamma}_1, \ldots, \gamma_5$*, we can compute an explicit plot of this curve in* $\mathbb{R}^2$ *by plotting the real roots of the curve intersected with various vertical lines swept across the second image plane.*

We call the curve in Theorem 3 a 4.5**-point curve**, because it is specified by four-and-a-half image point pairs, namely $\gamma_1, \bar{\gamma}_1, \ldots, \gamma_5$. See Figure 4 for sample renderings.

Theorem 4 (Ill-posed image data for F) *Consider the 7-point problem in Section 2. Let* $((\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1), \ldots, (\boldsymbol{\gamma}_7, \bar{\boldsymbol{\gamma}}_7)) \in (\mathbb{R}^2 \times \mathbb{R}^2)^{\times 7}$ *be image data. Then* x *is ill-posed,* i.e. *there exists some compatible world scene which is ill-posed, only if a certain polynomial* $\mathbf{P}$ *in the entries of* $\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1, \ldots, \bar{\boldsymbol{\gamma}}_7$ *vanishes. This polynomial has degree* 6 *separately in each of the points* $\boldsymbol{\gamma}_1, \ldots, \bar{\boldsymbol{\gamma}}_7$*. In particular, if we fix numerical values for* $\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1, \ldots, \boldsymbol{\gamma}_7$ *but keep* $\bar{\boldsymbol{\gamma}}_7 \in \mathbb{R}^2$ *as variable, then (generically)* $\mathbf{P}$ *specializes to a degree* 6 *polynomial just in* $\bar{\boldsymbol{\gamma}}_7$*, and its vanishing set is a degree* 6 *curve in the second image plane. Moreover given the values for* $\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1, \ldots, \boldsymbol{\gamma}_7$*, we can compute an explicit plot of this curve in* $\mathbb{R}^2$ *by plotting the real roots of the curve intersected with various vertical lines swept across the second image plane.*

We call the curve in Theorem 4 a 6.5**-point curve**, because it is specified by six-and-a-half image point pairs, namely $\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1, \ldots, \boldsymbol{\gamma}_7$. See Figure 4 for sample renderings.

5. Experimental Results

Our experimental results are mostly on synthetic data, although at the end of this section some illustration is shown on real data.

Data Generation: We generate random valid configurations consisting of the world scene $(\mathbf{R}, \hat{\mathbf{T}}, \Gamma_1, \ldots, \Gamma_N)$, intrinsic matrix K, and 2D point pairs on the image plane $(\gamma_1, \bar{\gamma}_1, \cdots, \gamma_N, \bar{\gamma}_N)$ which are expressed as $(\boldsymbol{\gamma}_1, \bar{\boldsymbol{\gamma}}_1, \cdots, \boldsymbol{\gamma}_N, \bar{\boldsymbol{\gamma}}_N)$ in pixel units. Here $N = 5$ or $N = 7$, depending on whether the camera is calibrated or uncalibrated. We generate random instances as follows:

- $\mathbf{R}$: The QR decomposition of a random 3×3 matrix with i.i.d. standard normal entries gives an orthonormal matrix sample;
- $\hat{\mathbf{T}}$: A uniformly sampled vector from the unit sphere with radius as 1 meter;
- Γ_i: uniformly sampled points with depth in [1, 20] meters;
- K: chosen so that the image size is 640×480, focal length is set to 32 millimeters, and principle point is the image center;
- $(\gamma_i, \bar{\gamma}_i)$, and $(\boldsymbol{\gamma}_i, \bar{\boldsymbol{\gamma}}_i)$: projections of Γ_i onto two images.

We discard instances where any of the 2D points land outside the image's boundary or cases where 3D points locate at the back of the camera.

Instability Revelation: We first aim to demonstrate that instabilities do empirically occur for both calibrated and uncalibrated relative pose estimation minimal problems. To this end, we generate 3000 synthetic minimal problems each for calibrated/uncalibrated as described above. For each minimal problem instance, we add i.i.d. noise to the image points drawn from the spherical Gaussian $\mathcal{N}(0, \sigma^2 I_2)$ for different noise levels σ. Then, we separately solve the original and perturbed problems, and compare them. We define an estimate to be unstable if either

of the following criteria holds: *(i)* **Large error in the solutions for the perturbed points**: the error in the fundamental or essential matrix after normalization is defined by $e = \text{mean}(\text{abs}(\text{abs}(\bar{M}./M) - \mathbf{1}\mathbf{1}^\top))$. Here "./" denotes element-wise division, M is the ground-truth model, $\bar{M}$ is the nearest estimated model, and $\mathbf{1}\mathbf{1}^\top$ is the 3×3 matrix with each element 1. Then *(i)* holds if e exceeds a threshold τ. *(ii)* **Change in the number of real solutions**: this behavior is troublesome because the true epipolar matrix can disappear into the complex plane if there is a variation in the number of real solutions.

Figure 3 (a) and (b) shows the fraction of the erroneous estimations out of the 3000 instances at various small to moderate noise levels and error thresholds. It is clear that for random perturbations, the ratio of erroneous cases cannot be ignored even when the noise is small. In practice, given a sufficient number of correspondences, unstable instances are weeded out by RANSAC through maximizing the number of inliers. Even when *all* correspondences are inliers, RANSAC is still needed to overcome the instabilities of relative pose estimation.

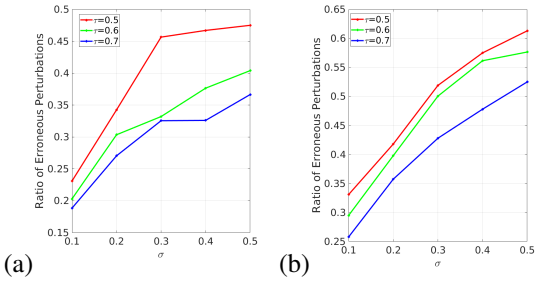

Figure 3. Ratio of erroneous estimations out of 3000 random synthetic minimal problems at different noise levels σ and error thresholds τ for (a) Fundamental matrix and (b) Essential matrix.

Instability Detection: The methods described in Section 4.3 are now applied to compute the 4.5-point degenerate curves for the uncalibrated case and 6.5-point curves for the calibrated case. The scenario is that 4(6) correspondences are fixed and for the 5th(7th) correspondence the point on one image is fixed and the locus of all unstable points is derived as a curve. Figure 4 shows several sample curves plotted on the second image plane along with the given image points. For the uncalibrated case, the degree of the 6.5-point curve is 6, while for calibrated case the degree of the 4.5-point curve is 30. The curves split the image plane into different connected components, wherein the number of real solutions is locally constant. In the language of [4], the curves are the "real discriminant loci".

In another experiment, we separate the 3000 random synthetic minimal problems into three categories: stable cases, unstable cases, and the borderline cases (given that the condition number is a continuous indication of the stability). Here, an instance is sorted according to the the number of erroneous estimates among $n = 20$ perturbations, denoted by $\hat{n}$. If $\hat{n} \in [0, n/3]$, we count the instance as stable; if $\hat{n} \in [2n/3, n]$, we count the instance as unstable; and if $\hat{n} \in [n/3, 2n/3]$, we count the instance as borderline. In this experiment, we use $\tau = 0.5$ and $\sigma = 0.3$. For the uncalibrated case, the average distance from the 7th point to the 6.5-point curve is 2.35 pixels among unstable cases, while for the stable cases it is 22.12 pixels. For the calibrated case, the average distance from the 5th point to the 4.5-point curve is 0.32 pixels for unstable cases, while for the stable case it is is 14.95 pixels. From these statistical differences (see Figure 5), we observe that the stable and unstable categories can be distinguished by thresholding on the distance between the last point to the X.5-point curve.

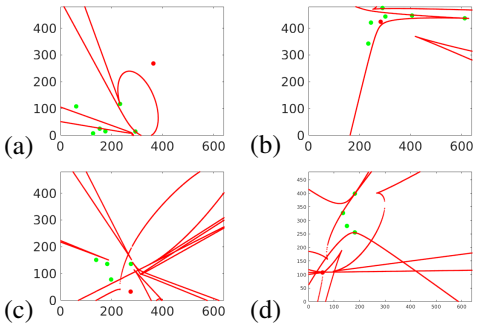

Figure 4. Sample results for the X.5-point degenerate curve in Theorems 3 and 4. Correspondences used in computing the curve are shown as green; the red points are the 5th/7th correspondence for calibrated/uncalibrated relative pose estimation, respectively. The red curve is the X.5-point curve computed using homotopy continuation. Stability is directly correlated to distance of the second point from the curve. (a) A stable configuration for uncalibrated estimation. (b) An unstable configuration for uncalibrated estimation. (c) A stable configuration for calibrated case. (d) An unstable configuration for calibrated case.

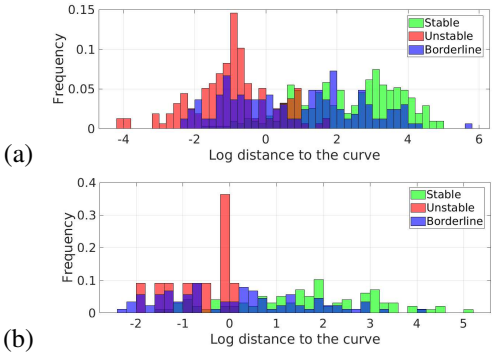

Figure 5. Histogram of distance from the last point to the degenerate curve sorted by: stable cases (green), unstable cases (red) and borderline cases (blue). (a) Uncalibrated estimation. (b) Calibrated estimation. Stable and unstable categories are separated by distance to the curve.

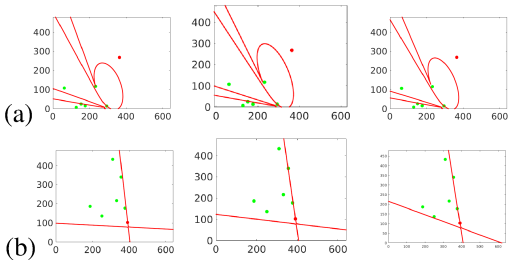

Figure 6. Illustrative result indicating the stability of the degenerate curve under various noise shows remarkable stability of the curve. (a) The degenerate curve of a stable uncalibrated configuration. (b) The degenerate curve of an unstable uncalibrated configuration. Curves for calibrated estimation are shown in the supplementary materials.

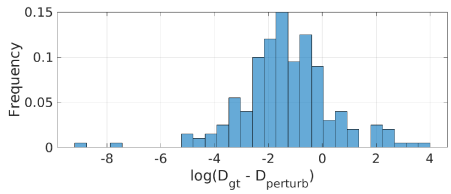

Figure 7. Histogram of the difference between two distances: from the target point to the X.5-point curve without perturbation; and from the target point to the X.5-point curve with perturbation on the other points.

Stability of Instability: Here we show that the degenerate curve is mostly stable to the presence of the noise so that our idea is not only theoretically correct, but can also be used in the practical setting of noisy images. Figure 6 shows some examples of the X.5-curves when adding noise to the corresponding points. The distribution of the log-difference between distance from the point to the curve in the noiseless and noisy case, Figure 7, showing the perturbation does not drastically change the distance.

Illustration with Real Data: Based on the above, the X.5-point curve can be used with real images to detect near-degenerate minimal cases. To show this, we use image pairs given by the RANSAC 2020 dataset [25] where standard point correspondences are available. Figure 8 shows that for a solution with large error compared to the ground-truth, the remaining selected point is close to the degenerate curve. More results are in the supplementary materials.

In another test with real data, we randomly took 1000 inlier minimal samples from each image pair in the dataset that had more than 100 inlier correspondences. We found that only 50% of these minimal configurations had large distances to the degenerate curve, so about half were unstable. However, *after* running RANSAC on all inliers and taking the winning hypothesis, we found about 90% of the winning hypotheses had large distances to the curve. This shows, indeed, that RANSAC selects stable configurations.

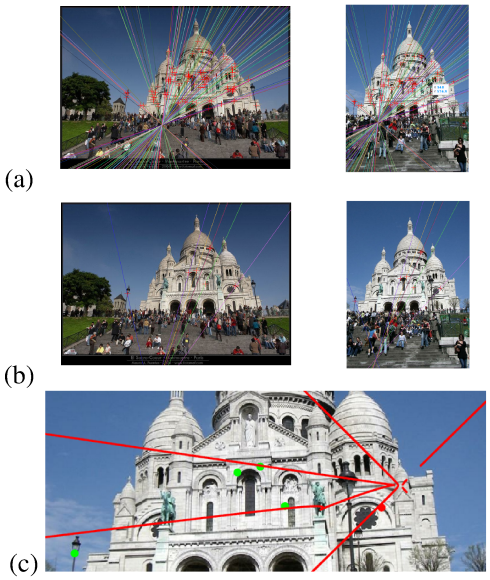

Figure 8. An example with real data to demonstrate an unstable minimal configuration with all-inlier correspondences. (a) The ground-truth epipolar geometry of a pair of images. (b) The closest solution found by the 7-point algorithm give 7 inliers. (c) Zoomed-in image showing that the remaining point is close to the degenerate curve, indicating that this is poorly-conditioned data.

6. Conclusion

In this paper, we developed a general framework for analyzing the numerical instabilities of minimal problems in multiview geometry. We applied this to the problem of relative pose estimation, namely, the popular 5-point and 7-point problems. We derived condition number formulas, and we characterized the ill-posed world and image scenes.

Numerical experiments on synthetic and real data and supported our theoretical findings. We observed numerical instabilities for image data landing close to the 4.5- and 6.5-point degenerate curves, which are used to describe ill-posed problem instances in Theorems 3 and 4.

This paper related the numerical instabilities of minimal problems to the function of RANSAC inside SfM reconstructions. Given *all* inlier data, RANSAC is still needed to overcome the ill-conditioning of relative pose estimation.

In future work, we could apply our theory to other minimal problems, *e.g.*, partially calibrated relative pose estimation or three-view geometry. In addition, we would like to develop a real-time means of recognizing and filtering out poorly-conditioned image data. Such could be applied before solving minimal problems and running RANSAC.

Acknowledgements: The authors are grateful to have participated the Algebraic Vision Research Cluster at ICERM, Brown University in Spring 2019, where they met and the seeds of this project were planted. Kimia and Fan gratefully acknowledge the support of NSF award 1910530.

References

[1] D. Ablan. *Digital Photography for 3D Imaging and Animation*. Wiley, 2007. 1

[2] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building Rome in a day. *Communications of the ACM*, 54(10):105–112, 2011. 1

[3] D. Barath, J. Noskova, M. Ivashechkin, and J. Matas. MAGSAC++, a fast, reliable and accurate robust estimator. In *CVPR*, pages 1304–1312, 2020. 1

[4] E. A. Bernal, J. D. Hauenstein, D. Mehta, M. H. Regan, and T. Tang. Machine learning the real discriminant locus. *arXiv preprint arXiv:2006.14078*, 2020. 7

[5] E. Brachmann and C. Rother. Neural-guided RANSAC: Learning where to sample model hypotheses. In *CVPR*, pages 4322–4331, 2019. 1

[6] P. Breiding and S. Timme. HomotopyContinuation. jl: A package for homotopy continuation in Julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018. 6

[7] P. Bürgisser and F. Cucker. *Condition: The Geometry of Numerical Algorithms*, volume 349. Springer Science & Business Media, 2013. 3, 4

[8] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In *Joint Pattern Recognition Symposium*, pages 236–243. Springer, 2003. 1

[9] M. Demazure. *Sur Deux Problemes De Reconstruction*. PhD thesis, INRIA, 1988. 2

[10] J. W. Demmel. On condition numbers and the distance to the nearest ill-posed problem. *Numerische Mathematik*, 51(3):251–289, 1987. 3, 5

[11] T. Dobbert. *Matchmoving: The Invisible Art of Camera Tracking*. Sybex, 2005. 1

[12] T. Duff, K. Kohn, A. Leykin, and T. Pajdla. PL1P-Point-Line minimal problems under partial visibility in three views. In *ECCV*, pages 175–192. Springer, 2020. 4

[13] J. Harris. *Algebraic Geometry: A First Course*, volume 133. Springer Science & Business Media, 2013. 4

[14] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2

[15] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. CUP, 2nd edition, 2004. 1, 2, 3

[16] B. Horn. *Robot Vision*. MIT press, 1986. 1

[17] F. Kahl and R. Hartley. Critical curves and surfaces for euclidean reconstruction. In *ECCV*, pages 447–462. Springer, 2002. 2

[18] Y. Kasten, A. Geifman, M. Galun, and R. Basri. Algebraic characterization of essential matrices and their averaging in multiview settings. In *CVPR*, pages 5895–5903, 2019. 1

[19] M. Kitagawa and B. Windsor. *MoCap for Artists: Workflow and Techniques for Motion Capture*. Focal Press, 2008. 1

[20] J. Krames. Zur ermittlung eines objektes aus zwei perspektiven.(ein beitrag zur theorie der "gefährlichen örter".). *Monatshefte für Mathematik und Physik*, 49(1):327–354, 1941. 2

[21] J. M. Lee. *Smooth Manifolds*. Springer, 2013. 4

[22] T. Luhmann, S. Robson, S. Kyle, and I. Harley. *Close Range Photogrammetry: Principles, Methods, and Applications*. Wiley, 2007. 1

[23] S. Maybank. *Theory of Reconstruction From Image Motion*, volume 28. Springer Science & Business Media, 2012. 2, 5

[24] D. Mishkin. Benchmarking robust estimation methods. Tutorial: RANSAC in 2020, *CVPR*, 2020. 1

[25] D. Mishkin. RANSAC tutorial 2020 dataset. https://github.com/ducha-aiki/ransac-tutorial-2020-data. 8

[26] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770, 2004. 1, 3

[27] O. Özyeşil, V. Voroninski, R. Basri, and A. Singer. A survey of structure from motion. *Acta Numerica*, 26:305–364, 2017. 1

[28] M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, and J. Tops. Image-based 3D acquisition of archaeological heritage and applications. In *Conference on Virtual Reality, Archaeology, and Cultural Heritage*, pages 255–262. ACM, 2001. 1

[29] R. Raguram, J.-M. Frahm, and M. Pollefeys. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In *ECCV*, pages 500–513. Springer, 2008. 1

[30] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *CVPR*, pages 4104–4113, 2016. 1

[31] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, pages 519–528, 2006. 1

[32] A. J. Sommese and C. W. Wampler. *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. World Scientific, 2005. 6

[33] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999. 1

[34] B. Sturmfels. The Hurwitz form of a projective variety. *Journal of Symbolic Computation*, 79:186–196, 2017. 6

[35] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2010. 1