

Rolling Shutter Relative Pose Estimation Made Practical

Daniel Barath^{1,2,3}

¹ ETH Zürich, Computer Vision and Geometry Group, dbarath@ethz.ch

² Google Zurich

³ HUN-REN SZTAKI, Budapest, Hungary

Abstract. Rolling shutter (RS) cameras equip virtually all consumer devices, yet RS-aware relative pose estimation has remained impractical: the state-of-the-art solver requires a minimum of 20 point correspondences, making RANSAC-based robust estimation prohibitively expensive due to the exponential dependence of the iteration count on the sample size. We make RS relative pose estimation practical by introducing affine correspondences (ACs) into the RS two-view geometry. We derive novel *RS-corrected affine constraints* that account for the coupling between point perturbations and the row-dependent essential matrix, providing two equations per correspondence beyond the standard epipolar constraint. Building on these constraints, we develop a linearized algebraic solver that estimates pose and RS motion from only 7 ACs. The solver exploits the physical smallness of RS parameters to linearize the constraints, eliminates the 12 RS unknowns via null-space projection, and solves the remaining degree-20 system via action matrices in 1.2ms. On the TUM RS benchmark, our method achieves the best pose and RS parameter accuracy among all tested methods and, uniquely among RS solvers, provides accurate translational velocity estimates – which are poorly conditioned from point correspondences alone due to a $\mathbf{v}\text{-}\mathbf{t}$ coupling. On the global-shutter EuRoC MAV dataset, the solver achieves comparable accuracy to the standard 5-point algorithm, demonstrating that it generalizes well to the GS setting. Code is at https://github.com/danini/rolling_shutter_made_practical.

1 Introduction

Rolling shutter (RS) cameras equip virtually all consumer devices – smartphones, tablets, action cameras, and drones. Unlike global shutter (GS) sensors that expose all pixels simultaneously, RS sensors read out the image row by row. When the camera moves during the readout interval – typically 15–30 ms for consumer CMOS sensors – each row captures the scene from a slightly different viewpoint. This sequential exposure invalidates the standard pinhole model that underpins most multi-view geometry methods [25]: the epipolar constraint, which normally defines a constant essential matrix across the image, becomes row-dependent.

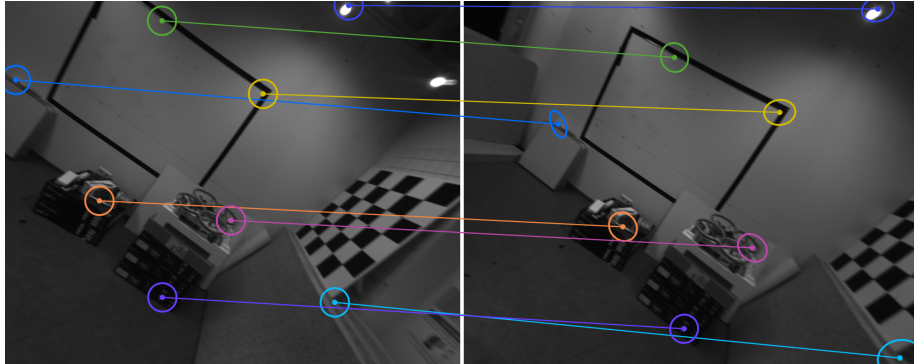


Fig. 1: Affine correspondences on a rolling shutter image pair from the TUM-RS dataset [33]. Circles in the left image and ellipses in the right image represent the local affine frames of 8 matched features. The camera rotates $\sim 8.5^\circ$ during each frame’s readout, producing visible RS distortion. On this pair, a standard GS 5-point solver [29] yields rotation/translation errors of $0.64^\circ/4.33^\circ$; our RS-aware 7-AC solver reduces these to $0.20^\circ/2.54^\circ$ while additionally recovering the RS motion parameters.

Relative pose estimation is fundamental to structure from motion (SfM) [21], visual odometry, SLAM, and augmented reality. These systems typically treat the essential matrix $\mathbf{E}$ as constant across the image. On RS images, this introduces a systematic bias that grows with camera velocity and readout time, and in severe cases causes bundle adjustment to diverge or converge to incorrect reconstructions. As RS cameras continue to dominate consumer and industrial markets, developing RS-aware geometric methods is essential.

Under the first-order RS model of Meingast et al. [25], the RS relative pose problem has 17 degrees of freedom: 5 for the relative pose ($\mathbf{R}, \mathbf{t}$) plus 12 for the per-camera angular and translational velocities $(\omega_k, \mathbf{v}_k)$, $k = 1, 2$. Dai et al. [17] showed that a minimum of **20 PCs** is required – far more than the 5 PCs of GS estimation [29]. This gap has practical consequences. The expected RANSAC [19] iteration count scales as η^{-k} (inlier ratio η , sample size k): at 50% inliers, $k = 5$ needs 32 iterations but $k = 20$ over a million – making the 20-PC solver [17] prohibitive and motivating smaller samples.

Affine correspondences (ACs) provide a natural path toward this goal. An AC augments a point match with the local 2×2 affine transformation between image patches [7, 12], yielding 3 equations per correspondence (1 epipolar + 2 affine). In the GS setting, this tripling of constraints enables essential matrix recovery from just 2 ACs [7] instead of 5 PCs, and homographies from 2 ACs [6] instead of 4 PCs. ACs are routinely available from affine-covariant detectors [26] and learned methods [28, 34] at negligible additional cost.

We make RS relative pose estimation practical by bringing ACs to the problem for the first time. We derive *RS-corrected affine constraints* that account for the coupling between point perturbations and the row-dependent essential matrix – a coupling absent in the GS formulation – and build a linearized algebraic

solver that estimates pose and RS parameters from 7 ACs, reducing RANSAC samples from 20 to 7 and iterations from millions to hundreds. On real-world data, we achieve the best pose and RS parameter estimation accuracy among all tested methods, demonstrating that RS-aware relative pose estimation is now practical with affine correspondences. Our key contributions are:

- We derive *RS-corrected affine correspondence constraints* that account for the row-dependent essential matrix – a coupling absent in the GS formulation. Each AC yields 3 equations (1 epipolar + 2 affine), so 7 ACs provide 21 equations for the 17 unknowns of the joint pose + RS problem.
- We develop a 7-AC algebraic solver with algebraic degree 20, based on action matrix methods, that makes RS-aware RANSAC practical: the sample size reduction from 20 to 7 decreases the expected RANSAC iterations by orders of magnitude (e.g., from 10^6 to ~ 600 at 50% inliers).
- We show that the proposed method achieves the best pose and RS estimation accuracy on the TUM-RS benchmark among RS and GS baselines, and substantially improves translational velocity estimation – which is poorly conditioned from PCs alone due to a $\mathbf{v}\text{-}\mathbf{t}$ coupling [17]. On the global-shutter EuRoC MAV dataset [13], the solver generalizes without degradation.

2 Related Work

RS camera modeling and pose estimation. Meingast et al. [25] introduced the first-order RS model in which each scanline’s pose is linearly interpolated between a reference rotation and translation, adding six motion parameters (angular and translational velocity) per frame. This model has been validated across consumer cameras [2, 3, 17] and remains the standard choice when the readout-time rotation is moderate ($\lesssim 10^\circ$). For absolute pose, Albl et al. [3, 4] developed Gröbner-basis minimal solvers under the linearized RS model (6-point, or 5-point with known gravity); Saurer et al. [32] proposed an alternative formulation; and Albl et al. [2] later unified these results. For relative pose, Dai et al. [17] derived the RS epipolar geometry and showed that the linearized two-view problem requires 20 PCs (44 for the quadratic model). The $\mathbf{v}\text{-}\mathbf{t}$ coupling between the translational velocity and the baseline direction had been identified earlier by Ait-Aider and Berry [1]. Zhuang et al. [35] used differential homographies for RS-aware SfM and rectification, Lee et al. [24] a gyroscope to reduce the unknowns, and Hedborg et al. [21] extended bundle adjustment to RS. Hahn et al. [20] recently gave an algebraic analysis of *order-one* RS cameras, under a model in which each world point is imaged at most once – distinct from the first-order model used here, where a point may be observed across several rows. Concurrently, Hruby and Pollefeys [22] reduce the correspondence count via a single-scanline formulation using line correspondences – a different input modality. A common theme is that RS solvers need far more correspondences than their GS counterparts, making RANSAC expensive; our solver reduces the RS sample to 7 ACs via affine-correspondence constraints.

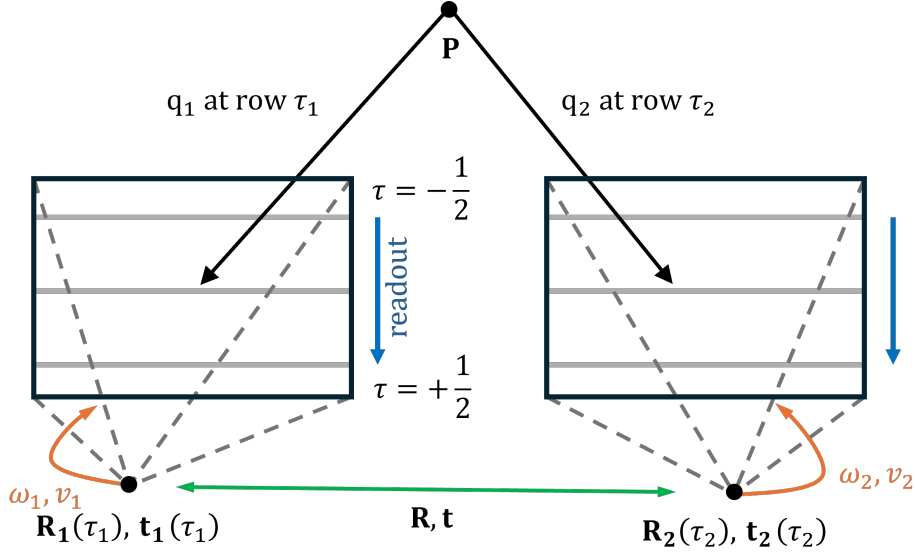


Fig. 2: Rolling shutter two-view geometry. A 3D point P is observed at rows τ_1 and τ_2 in two cameras whose poses vary during readout. Each camera is parameterized by a reference pose $(\mathbf{R}, \mathbf{t})$ and per-frame angular and translational velocities $(\omega_k, \mathbf{v}_k)$, yielding 17 degrees of freedom in total.

Affine correspondences in geometry estimation. An AC augments a point match with the local 2×2 affine transformation between patches, giving richer per-correspondence constraints. Differentiating the epipolar constraint yields two extra equations per AC [12], which Perdoch et al. [30] and Barath and Hajder [7] used to recover the essential matrix from just 2 ACs; homographies follow from 2 ACs [6], and later work consolidated the theory across model types [10, 31]. ACs are widely available from affine-covariant detectors [26] and learned methods [28, 34] at negligible cost, but have not been applied to rolling-shutter geometry – the gap we bridge.

Minimal solvers and RANSAC. Nistér’s 5-point algorithm [29] is the standard GS essential-matrix solver, and automatic Gröbner-basis generators [23] extended such techniques to RS settings [3, 4]. A problem’s algebraic complexity (its number of solutions) governs practicality – our joint 7-AC solver has only 20 – and the η^{-k} cost of RANSAC [19] and its variants [8, 9, 14] in sample size k makes reducing k (here from 20 to 7) decisive.

3 Geometric Background

Rolling Shutter Camera Model. A rolling shutter camera reads out the sensor row by row; if the camera moves during the 15–30 ms readout, each row is captured from a different viewpoint (Fig. 2). We adopt the standard first-order

RS model [2, 17, 25]. Let $\mathbf{R} \in \text{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$, $\|\mathbf{t}\| = 1$, be the relative rotation and translation at the reference rows, and let $\omega_k, \mathbf{v}_k \in \mathbb{R}^3$ ($k = 1, 2$) denote the angular and translational velocities during readout (absorbing the readout time t_{ro}). For a point whose pixel row is y , let $\tau = (y - h/2)/h \in [-\frac{1}{2}, \frac{1}{2}]$ be its scalar normalized row coordinate; the per-scanline pose is

$$\tilde{\mathbf{R}}(\tau_1, \tau_2) = (\mathbf{I} + \tau_2[\omega_2]_{\times}) \mathbf{R} (\mathbf{I} - \tau_1[\omega_1]_{\times}), \quad (1)$$

$$\tilde{\mathbf{t}}(\tau_1, \tau_2) = (\mathbf{t} + \tau_2 \mathbf{v}_2) - \tau_1 \tilde{\mathbf{R}} \mathbf{v}_1, \quad (2)$$

where $[\cdot]_{\times}$ denotes the skew-symmetric matrix, $\tau_1, \tau_2 \in \mathbb{R}$ are the scalar normalized row coordinates of the point in images 1 and 2, respectively, and $\mathbf{I}$ is the identity. The row-dependent essential matrix is

$$\tilde{\mathbf{E}}(\tau_1, \tau_2) = [\tilde{\mathbf{t}}(\tau_1, \tau_2)]_{\times} \tilde{\mathbf{R}}(\tau_1, \tau_2). \quad (3)$$

The unknowns of the RS two-view problem decompose into the *pose parameters* $(\mathbf{R}, \mathbf{t})$ (5 degrees of freedom, as translation is known only up to scale) and the *RS motion parameters* $\theta = (\omega_1, \mathbf{v}_1, \omega_2, \mathbf{v}_2) \in \mathbb{R}^{12}$, giving **17 degrees of freedom** in total for the entire system [17].

Point Correspondences in GS and RS Cameras. In the global shutter (GS) case ($\theta = \mathbf{0}$), $\mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}$ is constant and a point pair $(\mathbf{q}_1, \mathbf{q}_2)$ satisfies

$$\mathbf{q}_2^{\top} \mathbf{E} \mathbf{q}_1 = 0, \quad (4)$$

recoverable from 5 PCs [29]. In the RS case, each pair has its own $\tilde{\mathbf{E}}(\tau_1, \tau_2)$ and the constraint becomes

$$\mathbf{q}_2^{\top} \tilde{\mathbf{E}}(\tau_1, \tau_2) \mathbf{q}_1 = 0. \quad (5)$$

Each PC still yields one equation, but the unknowns now include the 12 RS parameters θ . The known-pose subproblem requires 12 PCs; the joint problem requires **20 PCs** [17] (44 for the quadratic RS model).

Affine Correspondences in GS Cameras. An affine correspondence (AC) augments a point match $(\mathbf{q}_1, \mathbf{q}_2)$ with the local affine map $\mathbf{A}_c = \partial \mathbf{q}_2 / \partial (q_{1,x}, q_{1,y}) \in \mathbb{R}^{3 \times 2}$, available from affine-covariant detectors [26, 28]. Let $\mathbf{a}_u, \mathbf{a}_v \in \mathbb{R}^3$ be the columns of $\mathbf{A}_c$ and $\mathbf{e}_u = (1, 0, 0)^{\top}$, $\mathbf{e}_v = (0, 1, 0)^{\top}$. Differentiating the GS epipolar constraint with respect to $q_{1,x}$ yields [7, 12] equation

$$\mathbf{a}_u^{\top} \mathbf{E} \mathbf{q}_1 + \mathbf{q}_2^{\top} \mathbf{E} \mathbf{e}_u = 0. \quad (6)$$

Differentiating with respect to $q_{1,y}$ gives the analogous

$$\mathbf{a}_v^{\top} \mathbf{E} \mathbf{q}_1 + \mathbf{q}_2^{\top} \mathbf{E} \mathbf{e}_v = 0. \quad (7)$$

Together with the epipolar constraint (4), each AC provides **3 independent equations** on the essential matrix. Since $\mathbf{E}$ has 5 degrees of freedom, just 2 ACs suffice for essential matrix estimation [7] versus 5 PCs with point-only methods.

4 RS Relative Pose Estimation

Beyond the first-order RS model, we make two explicit approximations, stated here and justified where used: **(i)** constraints are linearized in the motion parameters θ about $\theta = \mathbf{0}$ (valid as $\|\omega\| < 0.15$ rad, Sec. 4.4); and **(ii)** the null-space basis eliminating θ is fixed at the reference rotation $\mathbf{R}_{\text{ref}} = \mathbf{I}$ (Sec. 4.2). The constraints of Sec. 4.1 are exact; the approx. enter only at the solver stage.

4.1 RS-Corrected Affine Constraints

The original GS affine constraints (6)–(7) were derived under the assumption that $\mathbf{E}$ is constant across the image. In the RS case, the essential matrix $\tilde{\mathbf{E}}(\tau_1, \tau_2)$ varies with the row coordinates of both points. When we perturb $\mathbf{q}_1$ to derive the affine constraints, the row coordinates τ_1 and τ_2 change as well, because a shift in the image position of a point shifts the row at which it was read out. This introduces additional terms $\partial\tilde{\mathbf{E}}/\partial\tau_k$ that are absent in the GS formulation.

Derivation. Consider the RS epipolar constraint $F(\mathbf{q}_1, \mathbf{q}_2) = \mathbf{q}_2^\top \tilde{\mathbf{E}}(\tau_1, \tau_2) \mathbf{q}_1 = 0$. When we differentiate with respect to $q_{1,x}$ (horizontal shift), three effects contribute: (i) the explicit dependence of $\mathbf{q}_1$ on its x -coordinate ($\partial\mathbf{q}_1/\partial q_{1,x} = \mathbf{e}_u$); (ii) the induced change in $\mathbf{q}_2$ via the affine map ($\partial\mathbf{q}_2/\partial q_{1,x} = \mathbf{a}_u$); and (iii) the change in the essential matrix caused by the vertical shift of $\mathbf{q}_2$, which alters its readout row τ_2 . Crucially, a horizontal shift of $\mathbf{q}_1$ does not change its own readout row τ_1 (rows are horizontal), but it *may* shift $\mathbf{q}_2$ vertically through $\mathbf{a}_u$, thereby changing τ_2 . Applying the chain rule to all three effects yields

$$\underbrace{\mathbf{a}_u^\top \tilde{\mathbf{E}} \mathbf{q}_1 + \mathbf{q}_2^\top \tilde{\mathbf{E}} \mathbf{e}_u}_{\text{same as GS}} + \underbrace{\mathbf{q}_2^\top \frac{\partial \tilde{\mathbf{E}}}{\partial \tau_2} \mathbf{q}_1 \cdot \frac{f_y}{h} \cdot (\mathbf{a}_u)_y}_{\text{RS correction}} = 0, \quad (8)$$

where f_y is the focal length in pixels, h the image height, $(\mathbf{a}_u)_y$ the y -component of $\mathbf{a}_u$ (which converts the image-space vertical shift to a normalized-row shift via f_y/h), and $\partial\tilde{\mathbf{E}}/\partial\tau_2$ is given below.

For the vertical derivative ($q_{1,y}$), a shift of $\mathbf{q}_1$ now directly changes τ_1 , introducing an additional correction term:

$$\underbrace{\mathbf{a}_v^\top \tilde{\mathbf{E}} \mathbf{q}_1 + \mathbf{q}_2^\top \tilde{\mathbf{E}} \mathbf{e}_v}_{\text{same as GS}} + \underbrace{\mathbf{q}_2^\top \frac{\partial \tilde{\mathbf{E}}}{\partial \tau_1} \mathbf{q}_1 \cdot \frac{f_y}{h}}_{\text{RS correction (from } \tau_1)} + \underbrace{\mathbf{q}_2^\top \frac{\partial \tilde{\mathbf{E}}}{\partial \tau_2} \mathbf{q}_1 \cdot \frac{f_y}{h} \cdot (\mathbf{a}_v)_y}_{\text{RS correction (from } \tau_2)} = 0. \quad (9)$$

Equation (9) has *two* correction terms: a vertical shift of $\mathbf{q}_1$ changes τ_1 directly (the $\partial\tilde{\mathbf{E}}/\partial\tau_1$ term) and may also shift $\mathbf{q}_2$ vertically via $(\mathbf{a}_v)_y$ (the $\partial\tilde{\mathbf{E}}/\partial\tau_2$ term).

Row derivatives of the essential matrix. From (1)–(2), the derivatives of $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{t}}$ with respect to the row coordinates are

$$\frac{\partial \tilde{\mathbf{R}}}{\partial \tau_1} = -(\mathbf{I} + \tau_2[\omega_2]_{\times}) \mathbf{R} [\omega_1]_{\times}, \quad \frac{\partial \tilde{\mathbf{t}}}{\partial \tau_1} = -\tilde{\mathbf{R}} \mathbf{v}_1 - \tau_1 \frac{\partial \tilde{\mathbf{R}}}{\partial \tau_1} \mathbf{v}_1, \quad (10)$$

$$\frac{\partial \tilde{\mathbf{R}}}{\partial \tau_2} = [\omega_2]_{\times} \mathbf{R} (\mathbf{I} - \tau_1[\omega_1]_{\times}), \quad \frac{\partial \tilde{\mathbf{t}}}{\partial \tau_2} = \mathbf{v}_2 - \tau_1 \frac{\partial \tilde{\mathbf{R}}}{\partial \tau_2} \mathbf{v}_1, \quad (11)$$

Combining these with the product rule applied to the essential matrix (3) gives $\partial \tilde{\mathbf{E}} / \partial \tau_k = [\partial \tilde{\mathbf{t}} / \partial \tau_k]_{\times} \tilde{\mathbf{R}} + [\tilde{\mathbf{t}}]_{\times} \partial \tilde{\mathbf{R}} / \partial \tau_k$.

Summary. Each AC yields three equations on the RS two-view geometry: the epipolar constraint (5) plus two new affine constraints (8)–(9). The RS correction terms vanish when $\theta = \mathbf{0}$ (since $\partial \tilde{\mathbf{E}} / \partial \tau_k \rightarrow 0$ when there is no RS motion), recovering the standard GS constraints as a special case. These constraints are *general*: they hold for any RS two-view configuration under the first-order RS model and do not require any linearization. In the remainder of this section, we exploit them to build efficient minimal solvers for RANSAC.

4.2 Minimal Solver from 7 Affine Correspondences

The RS-corrected AC constraints (5), (8)–(9) involve all 17 unknowns (5 pose + 12 RS motion). Solving this full nonlinear system directly is algebraically intractable: even the known-pose subproblem alone (12 RS unknowns) already has 254 isolated solutions (supplementary material), and the full 17-DoF system further couples the pose with the row-dependent (τ_1, τ_2) terms.

Key observation: the RS parameters enter approximately linearly. The epipolar constraint (5) is a polynomial of multi-degree $(2, 1, 2, 1)$ in $(\omega_1, \mathbf{v}_1, \omega_2, \mathbf{v}_2)$: it is linear in each translational velocity $\mathbf{v}_k$ and quadratic in each angular velocity ω_k . However, the RS parameters are physically small ($\|\omega_k\| < 0.15$ rad for virtually all consumer cameras; see Sec. 4.4), so the quadratic terms $O(\|\omega\|^2) < 5 \times 10^{-3}$ are negligible compared to measurement noise. We therefore **linearize the RS constraints in θ** around the GS solution $\theta = \mathbf{0}$, keeping only first-order terms.

For a given pose $(\mathbf{R}, \mathbf{t})$, setting $\theta = \mathbf{0}$ reduces the essential matrix to $\mathbf{E}_0 = [\mathbf{t}]_{\times} \mathbf{R}$ and the row derivatives to the compact forms

$$\left. \frac{\partial \tilde{\mathbf{E}}}{\partial \omega_{1,k}} \right|_{\theta=\mathbf{0}} = -\tau_1 [\mathbf{t}]_{\times} \mathbf{R} [\mathbf{e}_k]_{\times}, \quad \left. \frac{\partial \tilde{\mathbf{E}}}{\partial v_{2,k}} \right|_{\theta=\mathbf{0}} = \tau_2 [\mathbf{e}_k]_{\times} \mathbf{R}, \quad (12)$$

with analogous expressions for ω_2 and $\mathbf{v}_1$. Each of the three per-AC constraints becomes a linear function of θ with coefficients that depend on $(\mathbf{R}, \mathbf{t})$ and the data $(\mathbf{q}_1, \mathbf{q}_2, \tau_1, \tau_2, \mathbf{A}_c)$. Stacking all $3N$ linearized constraints for N ACs yields

$$\mathbf{J}(\mathbf{R}, \mathbf{t}) \theta = -\mathbf{r}_0(\mathbf{R}, \mathbf{t}), \quad (13)$$

where $\mathbf{r}_0$ collects the GS residuals (evaluating each constraint at $\theta = \mathbf{0}$) and $\mathbf{J}$ is the $3N \times 12$ Jacobian matrix. The system is linear in θ but nonlinear in $(\mathbf{R}, \mathbf{t})$.

Eliminating the RS parameters via null-space projection. The key idea of our solver is to *eliminate* the 12 RS unknowns θ from the linear system (13) and obtain conditions that constrain only the 5 pose unknowns $(\mathbf{R}, \mathbf{t})$.

With $N = 7$ ACs, the system (13) has $3 \times 7 = 21$ rows and 12 columns. It is overdetermined in θ : there are more equations than unknowns. Such a system has a solution only if the right-hand side $-\mathbf{r}_0$ lies in the column space of $\mathbf{J}$. Equivalently, $\mathbf{r}_0$ must be orthogonal to every vector in the left null space of $\mathbf{J}$. Let $\mathbf{U}_\perp \in \mathbb{R}^{21 \times 9}$ be an orthonormal basis for this $21 - 12 = 9$ -dimensional subspace. Then (13) has a solution if and only if

$$\mathbf{U}_\perp^\top \mathbf{r}_0(\mathbf{R}, \mathbf{t}) = \mathbf{0}. \quad (14)$$

This projects out the 12 RS unknowns entirely, yielding **9 polynomial equations in the 5 pose unknowns $(\mathbf{R}, \mathbf{t})$** alone. Once a pose satisfying (14) is found, the RS parameters are recovered from (13) by least squares fitting.

Fixed null-space basis. Strictly, $\mathbf{U}_\perp$ depends on $(\mathbf{R}, \mathbf{t})$ through $\mathbf{J}$, which would make (14) implicitly nonlinear in the pose. We avoid this complication by computing $\mathbf{U}_\perp$ once at a fixed reference rotation $\mathbf{R}_{\text{ref}} = \mathbf{I}$ (identity) and a generic translation. This is an approximation, justified by two observations. First, $\mathbf{J}$ depends on the pose only through the linearization point of the RS model, and its left null space varies smoothly – so moderate deviations from $\mathbf{I}$ preserve the rank structure. Second, we verify empirically that $\mathbf{R}_{\text{ref}} = \mathbf{I}$ yields equal or better numerical conditioning than recomputing $\mathbf{U}_\perp$ at the initial 5-point estimate, producing a success rate above 85% for rotation errors below 1° on synthetic data. A pose-dependent $\mathbf{U}_\perp$ would moreover break the linear-in- $\mathbf{t}$ factorization that yields the clean $\mathbf{G}_{\text{poly}}$ structure of (15).

Polynomial formulation. The 9 consistency conditions (14) involve the rotation $\mathbf{R}$ and translation $\mathbf{t}$, which we now parameterize to obtain a polynomial system. We use the Cayley parameterization $\mathbf{R}(\mathbf{s}) = (\mathbf{I} - [\mathbf{s}]_\times)(\mathbf{I} + [\mathbf{s}]_\times)^{-1}$, where $\mathbf{s} = (s_1, s_2, s_3) \in \mathbb{R}^3$, and write the translation as $\mathbf{t} = (1, u, v)^\top$ (up to scale; this assumes $t_x \neq 0$, which holds generically and can be ensured by a coordinate permutation when $|t_x|$ is small).

A crucial simplification comes from the fact that $\mathbf{r}_0$ depends *linearly* on $\mathbf{t}$. This means (14) can be factored as

$$\mathbf{G}_{\text{poly}}(\mathbf{s}) \begin{pmatrix} 1 \\ u \\ v \end{pmatrix} = \mathbf{0}, \quad (15)$$

where $\mathbf{G}_{\text{poly}}(\mathbf{s})$ is a 9×3 matrix whose entries are degree-2 polynomials in $\mathbf{s}$. (The Cayley parameterization introduces a common denominator $d(\mathbf{s}) = 1 + \|\mathbf{s}\|^2$, which we absorb into $\mathbf{G}_{\text{poly}}$ to clear fractions.) We determine $\mathbf{G}_{\text{poly}}$ numerically: each entry has 10 monomials $\{1, s_1, s_2, s_3, s_1^2, s_1 s_2, \dots, s_3^2\}$, so evaluating (14) at 20 random $\mathbf{s}$ -samples and solving a 20×10 least-squares system per entry recovers the polynomial coefficients (to machine precision).

Eliminating the translation via minors. The system (15) states that $(1, u, v)^\top$ lies in the null space of the 9×3 matrix $\mathbf{G}_{\text{poly}}(\mathbf{s})$. A nontrivial null vector exists

if and only if $\mathbf{G}_{\text{poly}}$ has rank at most 2, *i.e.* all its 3×3 subdeterminants (minors) vanish. To work with a smaller set of minors, we first reduce from 9 rows to 5 by forming 5 random linear combinations (with i.i.d. Gaussian coefficients), yielding a 5×3 matrix $\hat{\mathbf{G}}(\mathbf{s})$ with degree-2 entries. By Bézout’s theorem [16], the solution set is independent of the (generic) choice of combination coefficients. We reduce to 5 rows rather than fewer because 5×3 is the smallest reduction whose minor ideal is zero-dimensional: a 4×3 reduction leaves a positive-dimensional solution set in $\mathbf{s}$. We then require all $\binom{5}{3} = 10$ maximal minors of $\hat{\mathbf{G}}$ to vanish. Each 3×3 determinant has degree $3 \times 2 = 6$ in $\mathbf{s}$. A key algebraic observation simplifies this system: every minor is divisible by the Cayley denominator $d(\mathbf{s}) = 1 + \|\mathbf{s}\|^2$. This occurs because $\mathbf{G}_{\text{poly}}$ was constructed by multiplying the rational matrix $\mathbf{G}(\mathbf{s})$ by $d(\mathbf{s})$; at the complex roots of $d(\mathbf{s}) = 0$, the Cayley rotation is undefined (it corresponds to a 180° rotation), and the original constraints impose no condition on (u, v) , so $\hat{\mathbf{G}}$ must drop rank there, forcing $d(\mathbf{s})$ to divide every maximal minor. We verify this computationally by evaluating each minor at random complex points on the variety $\{d(\mathbf{s}) = 0\}$, confirming vanishing to machine precision ($\sim 10^{-14}$), while $d(\mathbf{s})^2$ does *not* divide the minors.

After dividing out $d(\mathbf{s})$, we obtain 10 polynomial equations of degree 4 in $\mathbf{s} = (s_1, s_2, s_3)$. The number of isolated solutions is determined by the Hilbert function [16] of the ideal, which stabilizes at $H = 20$ (verified by computing the rank of Macaulay matrices at successive degrees). This establishes the algebraic degree of the system as 20.

Action matrix solver. We extract all 20 solutions using action matrices [16], a standard technique in computational algebraic geometry for solving zero-dimensional polynomial systems. The idea is to construct a matrix whose eigenvalues are the coordinates of the solutions.

We build the Macaulay matrix at extension degree $d_{\text{ext}} = 5$, which has dimensions 40×56 . This matrix is formed by multiplying each of the 10 degree-4 polynomials by the 4 monomials $\{1, s_1, s_2, s_3\}$ (degree ≤ 1), producing 40 rows; columns index all $\binom{5+3}{3} = 56$ monomials of degree ≤ 5 in (s_1, s_2, s_3) . The SVD of this matrix reveals a clear gap at the 36th singular value (ratio $\sigma_{36}/\sigma_{37} > 10^3$), confirming a 20-dimensional null space $\mathbf{V}_{\text{null}} \in \mathbb{R}^{56 \times 20}$ that represents the quotient algebra of the ideal. For each variable s_k , the action matrix is

$$\mathbf{M}_k = \mathbf{V}_{\text{null}}^\top \mathbf{S}_k \mathbf{V}_{\text{null}} \in \mathbb{R}^{20 \times 20}, \quad (16)$$

where $\mathbf{S}_k$ is the 56×56 monomial shift matrix that maps each basis monomial m to $s_k \cdot m$ (or to zero if $\deg(s_k \cdot m) > 5$). The eigenvalues of $\mathbf{M}_k$ are precisely the s_k -coordinates of the 20 solutions. Since the three action matrices commute, they can be simultaneously diagonalized: eigendecomposing $\mathbf{M}_1$ and applying the same eigenvector transform to $\mathbf{M}_2$ and $\mathbf{M}_3$ yields the complete triples $(s_1^{(i)}, s_2^{(i)}, s_3^{(i)})$ for $i = 1, \dots, 20$.

Solution recovery. For each real Cayley vector $\mathbf{s}^{(i)}$, the translation parameters (u, v) are recovered from the overdetermined system (15) via least squares, and the pose is computed as $\mathbf{R} = (\mathbf{I} - [\mathbf{s}]_\times)(\mathbf{I} + [\mathbf{s}]_\times)^{-1}$, $\mathbf{t} = (1, u, v)^\top / \|(1, u, v)\|$.

Solutions are ranked by the polynomial residual $\|\mathbf{G}_{\text{poly}}(\mathbf{s})(1, u, v)^\top\|$, which consistently identifies the correct solution. The RS parameters θ are then recovered from (13) by least squares.

Computational cost. The algebraic solver runs in ~ 1.2 ms in C++ (null space 0.04, $\mathbf{G}_{\text{poly}}$ fitting 0.2, minors and $d(\mathbf{s})$ division 0.02, 40×56 SVD 0.3, action matrices and eigendecomposition 0.6 ms). Though slower than the microsecond solvers ideal for RANSAC inner loops, the reduced sample size ($k=7$ vs. 20-44) more than compensates via far fewer iterations.

Degeneracies. Null-space elimination requires $\mathbf{J}$ to have full column rank (12); this fails for degenerate samples (*e.g.* all 7 ACs on one row, where the τ -dependent terms vanish). Such cases are rare in RANSAC and are detected via the SVD gap of the 40×56 Macaulay matrix – if absent, the solver returns nothing and RANSAC proceeds.

Remark: 6-AC solver. Each AC gives 3 constraints, so $N = 5$ ACs underdetermine the 17-DoF system while $N = 6$ overdetermine it by one, making 6 ACs theoretically sufficient. However, the resulting 6×3 matrix is too small for the minor-based elimination above; a hidden-variable alternative [23] instead suffers from a positive-dimensional spurious component (46-dimensional null space) that destabilizes extraction and costs ~ 55 ms per call – unsuitable for RANSAC (full derivation in the supplementary material).

4.3 Integration into RANSAC

We integrate the 7-AC solver into SuperRANSAC [5] with local optimization.

Hypothesis generation. For each RANSAC iteration, we draw 7 ACs and run: (1) Nistér’s 5-point algorithm [29] on the first 5 point coordinates, selecting the best $\mathbf{E}$ via the 6th and 7th ACs and decomposing into $(\mathbf{R}, \mathbf{t})$; then (2) the action-matrix solver (Sec. 4.2), producing up to 20 candidates filtered by cheirality and physical RS parameter bounds. A GS fallback ($\theta = \mathbf{0}$) is always included. Models are scored via the RS Sampson distance [17] $d_S^2 = ((\mathbf{q}_2^\top \tilde{\mathbf{E}}_i \mathbf{q}_1)^2) / (\|\tilde{\mathbf{E}}_i \mathbf{q}_1\|_{1:2}^2 + \|\tilde{\mathbf{E}}_i^\top \mathbf{q}_2\|_{1:2}^2)$, where $\tilde{\mathbf{E}}_i = \tilde{\mathbf{E}}(\tau_{1,i}, \tau_{2,i})$ is the essential matrix of pair i .

Non-minimal refinement. The best hypotheses are refined via joint Levenberg–Marquardt (LM) over all 17 parameters on the inlier set, minimizing

$$C = \sum_{i=1}^{N_{\text{in}}} (r_{0,i}^2 + w_{\text{ac}}^2 r_{1,i}^2 + w_{\text{ac}}^2 r_{2,i}^2), \quad (17)$$

where $r_{0,i}$ is the epipolar residual (5) and $r_{1,i}, r_{2,i}$ the affine residuals (8)–(9) of pair i , each evaluated with its own $\tilde{\mathbf{E}}_i$ under the full nonlinear RS model. All three residuals are computed on normalized image coordinates $(\mathbf{K}^{-1}\mathbf{q})$ and thus share a common scale; we keep the affine terms in algebraic form since, unlike the epipolar residual, they have no direct point-to-line distance analogue. The weight w_{ac} balances the affine against the epipolar terms: all reported results use $w_{\text{ac}} = 1$, and the pose AUC is insensitive to this choice for $w_{\text{ac}} \in [0.5, 2]$; a

principled, data-driven alternative sets w_{ac} to the ratio of the held-out standard deviations of the epipolar and affine residuals.

Handling the $\mathbf{v}$ - $\mathbf{t}$ coupling. The translational velocity $\mathbf{v}_k$ is poorly separated from $\mathbf{t}$ through epipolar constraints alone [1, 17], making the LM Jacobian ill-conditioned along $\mathbf{t}$. We apply mild damping $\lambda_v \mathbf{I}$ on the $\mathbf{v}$ -blocks, biasing toward small velocities. The affine residuals additionally help alleviate this coupling, as they depend on $\partial \tilde{\mathbf{E}} / \partial \tau_k$ and thus provide information to distinguish $\mathbf{v}_k$ from $\mathbf{t}$: while epipolar constraints measure only the projection of motion along epipolar lines – where $\mathbf{v}$ and $\mathbf{t}$ produce nearly identical effects – the AC constraints capture the *spatial gradient* of this projection across rows, which varies differently with $\mathbf{v}$ (row-dependent) than with $\mathbf{t}$ (constant).

4.4 Physically Feasible RS Parameters

The RS motion parameters ω_k and $\mathbf{v}_k$ represent the *total* rotation and translation accumulated during the sensor readout, so their physical magnitudes are tightly constrained. Since the model parameters relate to their physical counterparts as $\|\omega\|_{\text{model}} = \|\omega\|_{\text{phys}} \cdot t_{\text{ro}}$ and $\|\mathbf{v}\|_{\text{model}} = \|\mathbf{v}\|_{\text{phys}} \cdot t_{\text{ro}}$, even an aggressively maneuvering drone (angular velocity ~ 10 rad/s, translational velocity ~ 30 m/s) with a typical readout time $t_{\text{ro}} \approx 30$ ms yields $\|\omega\| \leq 0.3$ rad ($\sim 17^\circ$) and $\|\mathbf{v}\| \leq 0.9$. For handheld cameras, the values are 1–2 orders of magnitude smaller. Notably, while $\|\omega\|$ remains small across all scenarios, $\|\mathbf{v}\|$ can reach ~ 0.9 m for fast-moving platforms, underscoring the importance of accurately estimating $\mathbf{v}$ – which, as we show in Sec. 5.1, is poorly conditioned from point correspondences alone. This has two important consequences for our solver.

Justification of linearization. Since the RS parameters are small, the nonlinear terms in the epipolar constraint – which scale as $\tau^2 \|\omega\|^2 \leq 0.25 \cdot 0.02 = 5 \times 10^{-3}$ – are 2–3 orders of magnitude below typical measurement noise. The linearization in θ used by our solver (Sec. 4.2) is therefore an excellent approximation, not merely a practical heuristic.

Early rejection. The physical constraints on θ provide a useful sanity check during RANSAC: candidate models with $\|\omega\| > 0.5$ rad or unreasonably large $\|\mathbf{v}\|$ can be discarded, avoiding unnecessary scoring over the full correspondence set.

5 Experiments

We evaluate the proposed RS-7AC solver in three settings: (i) controlled synthetic experiments that isolate the effect of individual noise sources on solver accuracy, (ii) real-world experiments on the TUM RS dataset [33] within a full RANSAC pipeline, and (iii) a test on the global-shutter EuRoC MAV dataset [13] – the special case of zero readout time – to verify the method does not degrade in the absence of RS distortion. We compare against GS-5PC (Nistér’s 5-point algorithm [29], which ignores RS), GS-2AC (the 2-AC essential matrix solver of Barath and Hajder [7], which also ignores RS but uses affine correspondences),

and the RS-20PC and RS-44PC solvers of Dai et al. [17] (linearized and GN-refined).

5.1 Synthetic Experiments

We generate synthetic RS two-view scenes (50 random 3D points; two RS cameras with $f_x = f_y = 500$ px, 640×480 ; random relative pose with rotation $\leq 10^\circ$ and forward-dominant translation; RS motion scaled by a factor α applied to both ω_k and $\mathbf{v}_k$). Ground-truth affine frames follow from central-finite-difference differentiation of the full RS projection (including the iterative row solve): $\mathbf{A}_c = (\partial \mathbf{q}_2 / \partial \mathbf{X})(\partial \mathbf{q}_1 / \partial \mathbf{X})^\dagger$.

Each solver receives its minimal sample (5/20/44 PCs for GS-5PC/RS-20PC/RS-44PC, 7 ACs for RS-7AC) with oracle model selection; since minimal models are too noisy to compare directly, the best is refined on all 50 correspondences by joint LM, measuring each solver’s quality as a RANSAC initialization. RS-7AC refines with the combined cost (17) ($w_{ac} = 1$), the others with epipolar residuals only. We report median errors over 500 trials.

We sweep (1) point noise $\sigma_p \in [0, 3]$ px (fixed $\alpha = 0.5$), (2) affine noise $\sigma_a \in [0, 0.1]$ (fixed $\sigma_p = 0.5$ px, $\alpha = 0.5$), and (3) RS magnitude $\alpha \in [0, 3]$ (fixed $\sigma_p = 0.5$ px, $\sigma_a = 0.01$), where α scales both ω_k and $\mathbf{v}_k$ simultaneously. Affine noise is additive Gaussian on each entry of the 3×2 affine matrix. The swept range is calibrated to real data: $\sigma_a \approx 0.05$ matches the median AC residual of AffNet and $\sigma_a \approx 0.1$ that of RoMa on TUM-RS (supplementary material), so the tested noise levels reflect those of practical matchers.

Figure 3 shows the results. For *pose estimation* (top two rows), RS-7AC achieves competitive pose accuracy despite using only 7 correspondences versus 20-44. At moderate noise ($\sigma_p = 0.5$ px, $\alpha = 0.5$), the median rotation error is 0.65° for RS-7AC versus 0.46° for RS-44PC and 4.21° for GS-5PC. RS-44PC benefits from its larger sample in this oracle setting, but the gap is small and, as we show in Sec. 5.2, is more than offset by the RANSAC advantage of the smaller sample size in real-world conditions. The RS-aware methods degrade gracefully with increasing RS magnitude, whereas GS-5PC, which ignores RS, accumulates systematic bias. For *RS parameter estimation* (bottom two rows), RS-7AC exhibits $\sim 2\times$ higher angular velocity error ω than the point-based baselines, as the 7-AC algebraic solver provides a less accurate ω initialization than the overdetermined linear systems of RS-20PC and RS-44PC. However, the translational velocity $\mathbf{v}$ reveals the opposite trend: RS-20PC and RS-44PC, which rely solely on point correspondences, suffer from the $\mathbf{v}$ - $\mathbf{t}$ coupling discussed in Sec. 4.3, yielding median velocity errors of ~ 25 – far worse than predicting zero. RS-7AC, leveraging affine constraints in both estimation and refinement, achieves $\sim 3\times$ lower velocity error (~ 10), confirming that ACs provide information about $\mathbf{v}$ that PCs alone cannot resolve.

A direct comparison of raw solver outputs is not meaningful here, as the methods consume different inputs (7 ACs vs. 20 or 44 PCs); the oracle setup instead measures each solver’s quality as an initialization for refinement. Under real RANSAC with outliers (Tab. 1), RS-7AC outperforms the baselines on *both*

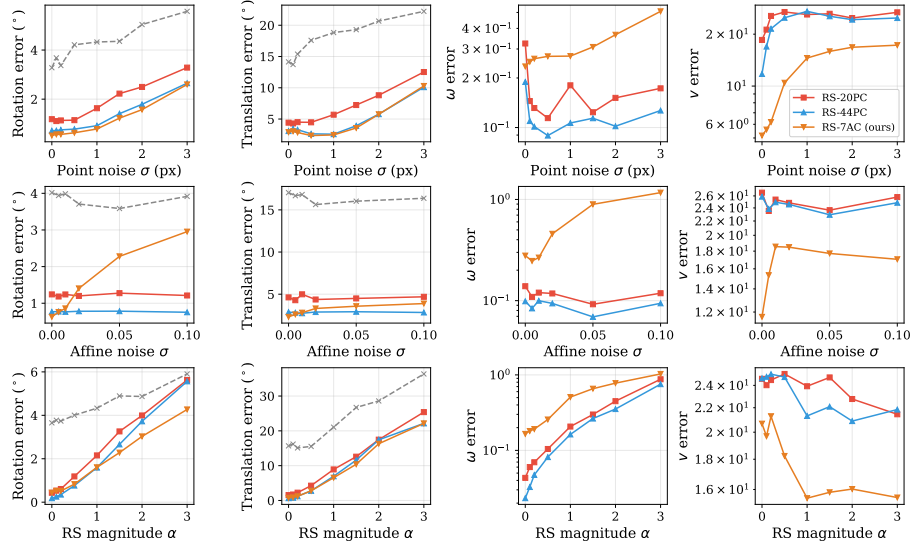


Fig. 3: Synthetic solver evaluation. Median errors over 500 trials for four metrics (rows: rotation, translation, ω error, $\mathbf{v}$ error) as a function of three noise parameters (columns: point noise, affine noise, RS magnitude α). GS-5PC is excluded from the RS parameter rows as it does not estimate RS parameters. RS-7AC achieves pose accuracy competitive with RS-44PC while using $6\times$ fewer correspondences, and is the only method with meaningful translational velocity estimates.

ω and $\mathbf{v}$, as the large samples of RS-20PC and RS-44PC cripple clean-sample selection.

5.2 Real-World Experiments

We evaluate on the TUM-RS dataset [33] (10 RS fisheye sequences with motion-capture ground truth), testing all 10 sequences at strides of 10 and 20 frames (increasing baseline and RS effect). We extract ACs with two feature types: (i) **Key.Net** [11] + **AffNet** [28] + **HardNet** [27], a classical pipeline (Key.Net keypoints, AffNet local affine shape, HardNet descriptor, mutual-NN ratio matching), with $\mathbf{A}_c$ from the matched affine shapes; and (ii) **RoMa** [18], a dense matcher whose pixel-level warp field we differentiate (central finite differences) to get a per-point 2×2 Jacobian. Since every warp-field sample yields one AC, $\#ACs = \#PCs$ at no extra cost – ACs are thus not fewer than PCs for dense matchers; their per-point accuracy is quantified in the supplementary material. All methods run in the same SuperRANSAC [5] pipeline (PROSAC sampling [15], MAGSAC scoring [9], nested local optimization). We report pose AUC at 5° , 10° , and 20° , the per-pair error being $\max(\epsilon_R, \epsilon_t)$.

Table 1 shows the results. The proposed RS-7AC solver consistently outperforms all baselines by a significant margin. At stride 10, RS-7AC achieves an

Table 1: Pose and RS parameter estimation on the TUM-RS dataset [33]. All methods use SupeRANSAC [5] with MSAC scoring and 5 000 iterations. Pose accuracy is measured by AUC of the cumulative error curve at 5° , 10° , and 20° thresholds, where the error of each pair is $\max(\epsilon_R, \epsilon_t)$. RS parameter accuracy is the median ℓ_2 error of the angular (ϵ_ω , rad) and translational (ϵ_v , m) velocities summed over both cameras. GS-5PC and GS-2AC do not estimate RS parameters (–). Two feature extractors are compared: AffNet [28] (sparse, ~ 500 ACs) and RoMa [18] (dense, $\sim 10\,000$ ACs). **Bold** marks the best result per column and extractor.

		stride = 10					stride = 20					t (secs) $\downarrow$
Method		AUC@5° $\uparrow$	@10° $\uparrow$	@20° $\uparrow$	ϵ_ω $\downarrow$	ϵ_v $\downarrow$	AUC@5° $\uparrow$	@10° $\uparrow$	@20° $\uparrow$	ϵ_ω $\downarrow$	ϵ_v $\downarrow$	
AffNet	RS-20PC [17]	0.101	0.181	0.258	0.177	11.260	0.097	0.161	0.228	0.183	8.861	12.7
	RS-44PC [17]	0.267	0.366	0.451	0.097	13.110	0.240	0.319	0.389	0.102	10.752	25.2
	GS-5PC [29]	0.428	0.535	0.626	—	—	0.380	0.472	0.544	—	—	1.5
	GS-2AC [7]	0.309	0.490	0.660	—	—	0.288	0.447	0.600	—	—	0.3
	Proposed	0.502	0.609	0.687	0.047	0.056	0.476	0.573	0.645	0.045	0.054	3.7
RoMa	RS-20PC [17]	0.769	0.869	0.924	0.063	8.011	0.687	0.773	0.826	0.071	5.595	20.9
	RS-44PC [17]	0.870	0.929	0.960	0.050	9.537	0.772	0.827	0.857	0.050	7.721	46.3
	GS-5PC [29]	0.861	0.908	0.934	—	—	0.757	0.803	0.830	—	—	5.3
	GS-2AC [7]	0.859	0.926	0.962	—	—	0.772	0.838	0.876	—	—	1.3
	Proposed	0.897	0.948	0.973	0.043	0.051	0.810	0.860	0.886	0.044	0.053	7.5

AUC@ 5° of 0.502, improving over GS-5PC (0.428) by 17% relative and over RS-44PC (0.267) by 88%. The advantage persists at stride 20, where the larger baseline amplifies the RS effect.

The strong performance of the GS methods over the Dai et al. solvers stems from sample size: drawing an all-inlier sample of size k at outlier ratio ϵ has probability $(1 - \epsilon)^k$, which at $\epsilon = 30\%$ drops from 0.08 for $k = 7$ to $\sim 10^{-7}$ for $k = 44$, crippling RANSAC for the 20- and 44-PC solvers. GS-2AC [7] benefits from an even smaller sample ($k=2$) but lacks RS modeling. Our 7-AC sample thus combines small-sample RANSAC efficiency with RS-aware modeling.

Runtime. RS-20PC and RS-44PC are by far the slowest (13–46 s) due to their large samples; RS-7AC runs in 3.7 s (AffNet) and 7.5 s (RoMa) – only $\sim 2\times$ slower than GS-5PC despite a 17-DoF vs. 5-DoF problem, thanks to its small sample. GS-2AC is fastest (~ 0.3 –1.3 s) but lacks RS modeling.

Global-Shutter Special Case. To verify that the RS solver does not degrade without RS distortion, we evaluate on the EuRoC MAV dataset [13] (six indoor global-shutter sequences, 752×480 , 20 fps, three from Vicon Room 1 and three from Machine Hall, with Vicon ground truth), using RoMa features at strides 10 and 20. As the camera is global-shutter, the RS parameters should be near-zero, and pose accuracy measures how gracefully the method handles their absence.

Table 2 shows the results. The proposed solver achieves similar pose accuracy to the GS baselines, demonstrating that it gracefully handles the GS special case. Notably, it correctly recovers near-zero RS parameters ($\epsilon_\omega \approx 0.01$ rad, $\epsilon_v \approx 0.03$), confirming that the solver does not introduce spurious distortions when RS effects are absent. In contrast, the Dai et al. solvers yield much higher

Table 2: Pose estimation on EuRoC MAV [13] (global-shutter cameras). Same setup as Tab. 1 but on GS images. RS parameters should be near-zero; ϵ_ω and ϵ_v measure how well each RS solver handles the GS case. **Bold** marks the best result per column. The proposed solver achieves comparable performance to GS solvers.

Method	stride = 10						stride = 20					
	AUC@5° ↑	@10° ↑	@20° ↑	ϵ_ω ↓	ϵ_v ↓		AUC@5° ↑	@10° ↑	@20° ↑	ϵ_ω ↓	ϵ_v ↓	
RS-20PC [17]	0.426	0.631	0.765	0.075	11.762		0.421	0.580	0.692	0.072	11.785	
RS-44PC [17]	0.499	0.687	0.800	0.041	19.296		0.467	0.622	0.726	0.035	14.763	
GS-5PC [29]	0.521	0.704	0.813	–	–		0.492	0.644	0.741	–	–	
GS-2AC [7]	0.517	0.702	0.814	–	–		0.491	0.647	0.749	–	–	
Proposed	0.512	0.701	0.816	0.016	0.035		0.484	0.641	0.744	0.012	0.028	

RS parameter errors ($\epsilon_\omega \approx 0.04$ - 0.08 rad, $\epsilon_v \approx 10$ - 20), as their large minimal samples make RANSAC less likely to find clean inlier sets even in the GS regime.

6 Conclusion

We introduced affine correspondences into rolling shutter two-view geometry and developed a linearized 7-AC solver (degree 20, 1.8ms) that reduces the RANSAC sample from 20 to 7, making RS-aware robust estimation practical. On TUM-RS, the method achieves the best pose accuracy among all baselines and substantially improves translational velocity estimation, which is poorly conditioned from point correspondences alone. On the global-shutter EuRoC MAV dataset, the solver handles the GS special case gracefully, being comparable to the GS 5-point algorithm while correctly recovering near-zero RS parameters.

References

1. Ait-Aider, O., Berry, F.: Structure and kinematics triangulation with a rolling shutter stereo rig. In: IEEE International Conference on Computer Vision (ICCV). pp. 1835–1840 (2009). <https://doi.org/10.1109/ICCV.2009.5459336>
2. Albl, C., Kukulova, Z., Larsson, V., Pajdla, T.: Rolling shutter camera absolute pose. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(6), 1439–1452 (2020). <https://doi.org/10.1109/TPAMI.2019.2894395>
3. Albl, C., Kukulova, Z., Pajdla, T.: R6P – rolling shutter absolute pose problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2292–2300 (2015). <https://doi.org/10.1109/CVPR.2015.7298842>
4. Albl, C., Kukulova, Z., Pajdla, T.: Rolling shutter absolute pose problem with known vertical direction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3355–3363 (2016). <https://doi.org/10.1109/CVPR.2016.365>
5. Barath, D.: SupeRANSAC: One RANSAC to rule them all. arXiv preprint arXiv:2506.04803 (2025)
6. Barath, D., Hajder, L.: A theory of point-wise homography estimation. Pattern Recognition Letters **94**, 7–14 (2017). <https://doi.org/10.1016/j.patrec.2017.05.007>

7. Barath, D., Hajder, L.: Efficient recovery of essential matrix from two affine correspondences. *IEEE Transactions on Image Processing* **27**(11), 5328–5337 (2018). <https://doi.org/10.1109/TIP.2018.2849866>
8. Barath, D., Matas, J.: Graph-cut RANSAC. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 6733–6741 (2018). <https://doi.org/10.1109/CVPR.2018.00704>
9. Barath, D., Noskova, J., Ivashechkin, M., Matas, J.: MAGSAC++, a fast, reliable and accurate robust estimator. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1301–1309 (2020). <https://doi.org/10.1109/CVPR42600.2020.00138>
10. Barath, D., Polic, M., Förstner, W., Sattler, T., Pajdla, T., Kukulova, Z.: Making affine correspondences work in camera geometry computation. In: *European Conference on Computer Vision (ECCV)*. *Lecture Notes in Computer Science*, vol. 12356, pp. 723–740. Springer (2020). https://doi.org/10.1007/978-3-030-58621-8_42
11. Barroso-Laguna, A., Riba, E., Ponsa, D., Mikolajczyk, K.: Key.net: Keypoint detection by handcrafted and learned CNN filters. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 5836–5844 (2019). <https://doi.org/10.1109/ICCV.2019.00593>
12. Bentolila, J., Francos, J.M.: Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding* **122**, 105–114 (2014). <https://doi.org/10.1016/j.cviu.2014.01.003>
13. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research (IJRR)* **35**(10), 1157–1163 (2016)
14. Chum, O., Matas, J., Kittler, J.: Locally optimized RANSAC. In: *DAGM Symposium on Pattern Recognition*. *Lecture Notes in Computer Science*, vol. 2781, pp. 236–243. Springer (2003). https://doi.org/10.1007/978-3-540-45243-0_31
15. Chum, O., Matas, J.: Matching with PROSAC – progressive sample consensus. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 220–226 (2005). <https://doi.org/10.1109/CVPR.2005.221>
16. Cox, D.A., Little, J., O’Shea, D.: *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. *Undergraduate Texts in Mathematics*, Springer, 4th edn. (2015). <https://doi.org/10.1007/978-3-319-16721-3>
17. Dai, Y., Li, H., Kneip, L.: Rolling shutter camera relative pose: Generalized epipolar geometry. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 4132–4140 (2016). <https://doi.org/10.1109/CVPR.2016.448>
18. Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: RoMa: Robust dense feature matching. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 19790–19800 (2024). <https://doi.org/10.1109/CVPR52733.2024.01871>
19. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381–395 (1981). <https://doi.org/10.1145/358669.358692>
20. Hahn, M.A., Kohn, K., Marigliano, O., Pajdla, T.: Order-one rolling shutter cameras. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 27007–27016 (2025). <https://doi.org/10.1109/CVPR52734.2025.02515>

21. Hedborg, J., Forssén, P.E., Felsberg, M., Ringaby, E.: Rolling shutter bundle adjustment. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1434–1441 (2012). <https://doi.org/10.1109/CVPR.2012.6247831>
22. Hruby, P., Pollefeys, M.: Single-scanline relative pose estimation for rolling shutter cameras. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2025)
23. Kukulova, Z., Bujnak, M., Pajdla, T.: Automatic generator of minimal problem solvers. In: European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, vol. 5304, pp. 302–315. Springer (2008)
24. Lee, C.R., Yoon, J.H., Park, M.G., Yoon, K.J.: Gyroscope-aided relative pose estimation for rolling shutter cameras. arXiv preprint arXiv:1904.06770 (2019)
25. Meingast, M., Geyer, C., Sastry, S.: Geometric models of rolling-shutter cameras. In: 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (OMNIVIS) (2005)
26. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision **60**(1), 63–86 (2004). <https://doi.org/10.1023/B:VISI.0000027790.02288.f2>
27. Mishchuk, A., Mishkin, D., Radenović, F., Matas, J.: Working hard to know your neighbor’s margins: Local descriptor learning loss. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 4826–4837 (2017)
28. Mishkin, D., Radenović, F., Matas, J.: Repeatability is not enough: Learning affine regions via discriminability. In: European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, vol. 11213, pp. 284–300. Springer (2018)
29. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(6), 756–770 (2004). <https://doi.org/10.1109/TPAMI.2004.17>
30. Perdoch, M., Matas, J., Chum, O.: Epipolar geometry from two correspondences. In: International Conference on Pattern Recognition (ICPR). vol. 4, pp. 215–219 (2006). <https://doi.org/10.1109/ICPR.2006.497>
31. Raposo, C., Barreto, J.P.: Theory and practice of structure-from-motion using affine correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5470–5478 (2016). <https://doi.org/10.1109/CVPR.2016.590>
32. Saurer, O., Pollefeys, M., Lee, G.H.: A minimal solution to the rolling shutter pose estimation problem. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1328–1334 (2015). <https://doi.org/10.1109/IROS.2015.7353540>
33. Schubert, D., Demmel, N., von Stumberg, L., Usenko, V., Cremers, D.: Rolling-shutter modelling for direct visual-inertial odometry. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 2462–2469 (2019). <https://doi.org/10.1109/IROS40897.2019.8968539>
34. Sun, P., Guan, B., Yu, Z., Shang, Y., Yu, Q., Barath, D.: Learning affine correspondences by integrating geometric constraints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27038–27048 (2025)
35. Zhuang, B., Cheong, L.F., Lee, G.H.: Rolling-shutter-aware differential SfM and image rectification. In: IEEE International Conference on Computer Vision (ICCV). pp. 948–956 (2017). <https://doi.org/10.1109/ICCV.2017.108>