

Gravity-aware partially calibrated absolute pose estimation from affine- or rotation-covariant features

Marcus Valtonen Örnham¹, Alberto Jaenal², and Stefan Adalbjörnsson¹

¹ Ericsson Research, Lund, Sweden

{marcus.valtonen.ornham,stefan.adalbjornsson}@ericsson.com

² University of Zaragoza, Spain

ajaenal@unizar.es

Abstract. Inertial measurement units (IMUs) are now standard in most consumer devices, such as smartphones, drones, and extended reality (XR) headsets. By fusing visual and inertial data, localization systems gain significantly in speed and robustness compared to vision-only or IMU-only approaches. However, traditional pose estimation methods fail to utilize the local geometric information embedded in feature descriptors like SIFT. Recent work has proved the advantages of leveraging this information for relative and absolute pose estimation, but its application to partially calibrated absolute pose estimation remains unexplored. In this paper, we derive novel constraints for joint estimation of absolute pose and focal length, making use of a gravity vector obtained from IMU data and the feature-induced local geometry, which we use to construct two efficient solvers: UP1PfAC, that operates given a single affine correspondence and UP2PfORI, which requires two orientation-covariant features. Unlike traditional, semi-calibrated absolute pose methods requiring four point correspondences, our solvers benefit from fewer samples and lower computational cost, simplifying robust estimation in modern RANSAC-like frameworks. We evaluate the proposed solvers against the state-of-the-art on large-scale public datasets and demonstrate that our method achieves fast and accurate localization and focal length estimation.

Keywords: Absolute pose estimation · Structure-from-Motion · Camera calibration

1 Introduction

Future generations of mixed-reality applications demand faster and more accurate global localization to enable collaborative multi-user experiences. In fact, a rich flora of techniques from image-based localization [52], structure-from-motion (SfM) [54, 55], simultaneous localization and mapping (SLAM) [15], and visual odometry (VO) [43] are available in the current literature to build such large-scale pipelines, yet solutions do not always transfer directly across applications. Klein and Murray [36], for instance, proposed a method tailored to augmented

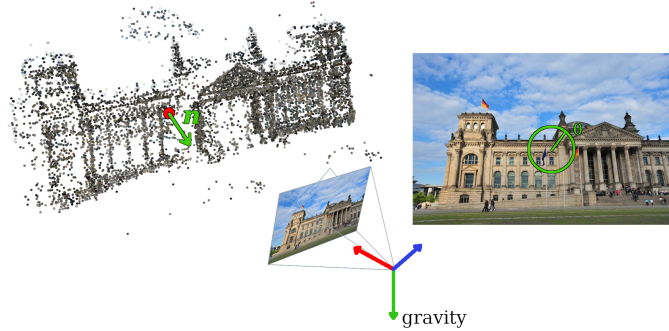


Fig. 1: *Overview of the problem.* We propose to jointly estimate the absolute pose and focal length of a query image from a 3D reconstruction. By leveraging IMU data and the intrinsic geometric properties of local descriptors, we are able to solve the problem using a single observation from an affine correspondence (UP1PfAC), or two orientation-covariant features (UP2PfORI) to a reference image. This greatly benefits the accuracy and execution times, compared to traditional point-based methods not utilizing IMU data or feature descriptor information, which require 3.5 points or more.

reality that deviates from typical SLAM by leveraging denser maps, lower-quality features, and decoupling tracking and mapping.

At the core of these applications lies the Perspective- n -Point (PnP) problem: estimating camera poses from 2D image features and 3D scene points. In practice, minimal solvers integrated into random sample consensus (RANSAC) [24] or its variants [8, 16] deal with noisy data and outliers, hence, traditionally, camera pose hypotheses are generated from three point correspondences using the P3P algorithm, for which efficient solutions are widely available [22, 26, 33, 48].

While point-based methods have been the undisputed approach for many years, several alternative approaches that utilize affine feature descriptors were proposed in the previous decade [2, 4, 5, 11, 13, 18, 23, 27–29, 49, 65]. The improvements in this field were summarized by Barath *et al.* [10], who argue for its continued use and benefits. Among the drawbacks of affine-based methods is the computational overhead of feature extraction and matching. For instance, the original Affine SIFT (ASIFT) [47] implementation reported a $13.5\times$ increase of computational complexity in extraction compared to regular SIFT features [45], and a corresponding matching step of $13.5^2 \approx 180\times$ greater complexity. With the introduction of deep learning-based descriptors, *e.g.*, AffNet [46], this gap has been reduced; however, the increased complexity remains a bottleneck.

Barath and Kukulova [6, 7] proposed to approximate a local affine feature $\mathbf{A}$ from a scale- and orientation-covariant feature, *e.g.*, SIFT and SURF [12]. They established the following constraints

$$a_1 c_{\text{ref}} + a_2 s_{\text{ref}} - q c_{\text{query}} = 0, \quad (1)$$

$$a_3 c_{\text{ref}} + a_4 s_{\text{ref}} - q s_{\text{query}} = 0, \quad (2)$$

where the relative scale is defined as $q = q_{\text{query}}/q_{\text{ref}}$ and the orientations α_{ref} and α_{query} are used as $c_i = \cos(\alpha_i)$, $s_i = \sin(\alpha_i)$. Here a_i denotes the elements

of A in row-major order. If only orientation-covariant features, *e.g.*, ORB [50], are used one may eliminate the scale by subtracting (1) from (2) to obtain

$$c_{\text{ref}}s_{\text{query}}a_1 + s_{\text{ref}}s_{\text{query}}a_2 - c_{\text{ref}}c_{\text{query}}a_3 - c_{\text{query}}s_{\text{ref}}a_4 = 0. \quad (3)$$

Since their discovery, these constraints have been used to derive new minimal solvers for homography estimation [3, 6, 19], relative pose estimation [7, 59], and calibrated absolute pose estimation [32, 61]. In this paper, we explore applications of affine- and orientation-covariant features to improve the overall localization accuracy and computational complexity. The contributions of this paper are:

- Derivation of constraints for absolute pose and unknown focal length estimation, using affine or orientation-covariant feature descriptors,
- Two novel polynomial solvers for the upright absolute pose problem with unknown focal length using affine or orientation-covariant features
- A rigorous comparison to state-of-the-art methods, discussing the benefits,
- A practical demonstration for real-world applications with focus on visual localization with respect to a global reference system shared by other users.

Notation. Following [60, 61], we use sans-serif capital A for matrices, bold $\mathbf{x}$ for vectors, and italic lower-case s for scalars. Subscripts indicate matrix and vector indexing, *e.g.*, $R_{1:2,1:2}$ refers to the 2×2 upper-left submatrix of R .

2 Related work

Absolute pose estimation with unknown focal length. The absolute pose problem with unknown focal length (PnPf) has been well-studied in both its minimal and non-minimal configurations. Since the problem has seven degrees of freedom, 3.5 point correspondences are necessary to solve the problem, *i.e.*, ignoring one of the equations obtained from the last point correspondences. Bujnak *et al.* [14] proposed one of the first general solvers for this case; however, the computational complexity rendered it infeasible for use in real-time applications. Zheng *et al.* [68] was able to solve the problem in a fast enough way to consider it for real-time applications, which was further improved upon by Wu [63]. Later, Larsson *et al.* [40] proposed a solver of a similar performance, using novel constraints for the particular setup. The overdetermined case was treated by Zheng *et al.* [68] and recently by Zhou *et al.* [69].

Upright solvers. Kukulova *et al.* [38] derived a minimal gravity-aware two point solver solving the absolute pose problem (UP2P) also considering an upright three point solver with unknown focal length and radial distortion coefficient (UP3Pfk). Sweeney *et al.* [57] applied a similar approach to the UP2P solver, and demonstrated its superior performance for localization in AR applications. Recently, an upright solver requiring only a single scale- and orientation-covariant feature was proposed in [61]. This solver, called UP1SIFT, was derived from (1)–(2), and the theory for local affine features, discussed in the next section. Beyond absolute pose, gravity alignment has been leveraged for other problems such as globally optimal relative pose estimation [20] and 3D camera registration [25].

Affine correspondences and unknown focal length. The methods utilizing affine correspondences to estimate focal length in some way do not treat the absolute pose problem; however, it has been studied for the relative pose problem with unknown focal length. Barath *et al.* [11] use two affine correspondences to solve it. Hraby *et al.* [32] showed that it is possible to use a single affine feature and estimated monocular depth. Furthermore, Ding *et al.* [21] treated different cases of fundamental matrix estimation using relative depths, which can be obtained from affine features, learned scales or depths. Recently, Yu *et al.* [66] proposed a method to solve the relative pose problem with unknown focal length and a known gravity vector by utilizing two affine features. Their approach leads to a polynomial eigenvalue problem, which can be solved efficiently and robustly, showing that correct utilization of affine features in combination with IMU data can significantly aid in simultaneous localization and camera calibration.

3 Method

We extend the work by Ventura *et al.* [60,61] by introducing an unknown focal length parameter. We propose to solve the problem using a single affine correspondence (UP1PfAC) or two orientation-covariant correspondences (UP2PfORI) with known gravity direction, see Fig. 1.

We aim to compute the unknown parameters of a query camera w.r.t. a world coordinate system. Following the notation introduced in [61], the known rotational components are encoded in R_{xz} , which allows for the parameterization of the full query rotation matrix $R_{\text{query}} = R_y R_{xz}$. The unknown query translation is denoted $\mathbf{t}_{\text{query}}$. The PnP problem requires a pre-built 3D reconstruction from a set of reference camera images. As such, the known reference camera poses used to construct the map also consist of a rotation and translation denoted R_{ref} and $\mathbf{t}_{\text{ref}}$, respectively. In this setting, the relative rotation between the reference camera and the query camera is given by $R = R_{\text{query}} R_{\text{ref}}^T$, and the corresponding relative translation is given by $\mathbf{t} = \mathbf{t}_{\text{query}} - R\mathbf{t}_{\text{ref}}$.

Let $\mathbf{p}_{\text{ref}}$ and $\mathbf{p}_{\text{query}}$ denote two corresponding inhomogeneous image point correspondences in reference and query images, respectively. Since we aim to utilize their point-based pixel coordinates to estimate the query pose and unknown focal length, we utilize the relationship between a local affine transformation and the pose of the query image relative to the reference image, as originally introduced in [60] and later modified for the world coordinate system in [61]:

$$\mathbf{A} = \frac{d}{m} \left(b\mathbf{R}_{1:2,1:2} - \mathbf{p}'_{\text{ref}1:2} \mathbf{n}_{\text{ref}1:2}^T - \mathbf{p}_{\text{query}} (b\mathbf{R}_{3,1:2} - \mathbf{p}'_{\text{ref}3} \mathbf{n}_{\text{ref}1:2}^T) \right), \quad (4)$$

where d is the depth of the point $\mathbf{p}_{\text{ref}}$ in the reference image, $\tilde{\mathbf{p}}_{\text{ref}}$ is the homogeneous representation of $\mathbf{p}_{\text{ref}}$, $\mathbf{n}_{\text{ref}} = R_{\text{ref}} \mathbf{n}$ is the normal vector transformed to the coordinate system of the reference frame, $b = \mathbf{n}_{\text{ref}}^T \tilde{\mathbf{p}}_{\text{ref}}$, $\mathbf{p}'_{\text{ref}} = R \tilde{\mathbf{p}}_{\text{ref}}$, and $m = \mathbf{n}_{\text{ref}}^T \tilde{\mathbf{p}}_{\text{ref}} (d(R_{3,:} \tilde{\mathbf{p}}_{\text{ref}}) + t_3)$. Directly working with the world coordinate system enables the usage of correspondences in different reference images, thereby increasing the number of possible samples for RANSAC. This is often utilized in modern localization frameworks, *e.g.*, HLOC [51].

3.1 Incorporating the IMU data

As opposed to previous work, we assume that two degrees of freedom can be reduced due to a known gravity direction, typically obtained using an IMU sensor. In this situation, the camera rotation can be decomposed as $R = R_y R_{xz}$, where the known rotation is encoded in R_{xz} , aligning the remaining unknown yaw angle with the y -axis in the coordinate system. Let θ denote the unknown angle of rotation about the y axis of the query camera, then

$$R_y = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}, \quad (5)$$

The tangent half-angle parameterization $r = \tan(\theta/2)$, results in

$$\cos \theta = \frac{1 - r^2}{1 + r^2} \quad \text{and} \quad \sin \theta = \frac{2r}{1 + r^2}, \quad (6)$$

which is suitable for polynomial solvers. Furthermore, due to the scale ambiguity, it is possible to work with scaled rotations, *i.e.* $R^T R = sI$, where $s \neq 0$, further simplifying the equations. The tangent half-angle parameterization cannot represent a 180° rotation; however, this can be addressed by applying a random rotation to the variables [31].

3.2 Novel constraints

The unknown focal length f_{query} can be introduced by substituting $\mathbf{p}_{\text{query}} \mapsto \mathbf{p}_{\text{query}}/f_{\text{query}}$ in (4), and since the reference focal length is assumed to be known, we assume that $\tilde{\mathbf{p}}_{\text{ref}}$ is already normalized by it. Furthermore, the affine features are affected $A \mapsto f_{\text{ref}}/f_{\text{query}}A$. Due to space constraints, we omit the subindex of the unknown query focal length and simply write f . This yields

$$\frac{a_1}{f} = \frac{d}{m}(br_{11} - p'_{\text{ref}_1}n_{\text{ref}_1} - \frac{p_{\text{query}_1}}{f}(br_{31} - p'_{\text{ref}_3}n_{\text{ref}_1})), \quad (7)$$

$$\frac{a_2}{f} = \frac{d}{m}(br_{12} - p'_{\text{ref}_1}n_{\text{ref}_2} - \frac{p_{\text{query}_1}}{f}(br_{32} - p'_{\text{ref}_3}n_{\text{ref}_2})), \quad (8)$$

$$\frac{a_3}{f} = \frac{d}{m}(br_{21} - p'_{\text{ref}_2}n_{\text{ref}_1} - \frac{p_{\text{query}_2}}{f}(br_{31} - p'_{\text{ref}_3}n_{\text{ref}_1})), \quad (9)$$

$$\frac{a_4}{f} = \frac{d}{m}(br_{22} - p'_{\text{ref}_2}n_{\text{ref}_2} - \frac{p_{\text{query}_2}}{f}(br_{32} - p'_{\text{ref}_3}n_{\text{ref}_2})). \quad (10)$$

Multiplying (7)–(10) by mf yields polynomial equations. Similarly, the orientation-covariant constraint (3) becomes

$$\begin{aligned} & c_{\text{ref}s_{\text{query}}}(br_{11} - p'_{\text{ref}_1}n_{\text{ref}_1} - \frac{p_{\text{query}_1}}{f}(br_{31} - p'_{\text{ref}_3}n_{\text{ref}_1})) \\ & + s_{\text{ref}s_{\text{query}}}(br_{12} - p'_{\text{ref}_1}n_{\text{ref}_2} - \frac{p_{\text{query}_1}}{f}(br_{32} - p'_{\text{ref}_3}n_{\text{ref}_2})) \\ & - c_{\text{ref}c_{\text{query}}}(br_{21} - p'_{\text{ref}_2}n_{\text{ref}_1} - \frac{p_{\text{query}_2}}{f}(br_{31} - p'_{\text{ref}_3}n_{\text{ref}_1})) \\ & - s_{\text{ref}c_{\text{query}}}(br_{22} - p'_{\text{ref}_2}n_{\text{ref}_2} - \frac{p_{\text{query}_2}}{f}(br_{32} - p'_{\text{ref}_3}n_{\text{ref}_2})) = 0, \end{aligned} \quad (11)$$

which are polynomial after multiplication by f . Finally, the two point-based constraints are given by (multiplied after by f to become polynomial)

$$\frac{p_{\text{query}_1}}{f}(d\mathbf{R}_{3,:}\tilde{\mathbf{p}}_{\text{ref}} + t_3) - (d\mathbf{R}_{1,:}\tilde{\mathbf{p}}_{\text{ref}} + t_1) = 0, \quad (12)$$

$$\frac{p_{\text{query}_2}}{f}(d\mathbf{R}_{3,:}\tilde{\mathbf{p}}_{\text{ref}} + t_3) - (d\mathbf{R}_{2,:}\tilde{\mathbf{p}}_{\text{ref}} + t_2) = 0. \quad (13)$$

3.3 Derivation of UP1PfAC

For the case of known gravity direction and unknown focal length with affine features, we use (7)–(10) and (12)–(13), which leaves us with a total of six equations and five unknowns: one rotational component (due to known reference direction), three translational components, and the focal length. In general, the problem is overdetermined, and therefore, we discard one of the equations. Experimentally, we found that discarding the last affine equation (10) works well. This equation, however, can be utilized later, as will be discussed. Note that all equations (7)–(13) are linear in $\mathbf{t}_{\text{query}}$, hence we may write the complete system of equations in the compact form

$$\mathbf{M}(r, f) \begin{bmatrix} \mathbf{t}_{\text{query}} \\ 1 \end{bmatrix} = \mathbf{0}, \quad (14)$$

where $\mathbf{M}$ is a 5×4 matrix. The same methods used in [60, 61] cannot be applied, as the unknown query focal length increases the number of monomials. Instead, we note that $\mathbf{M}$ must have a non-trivial nullspace; hence, it follows that all 4×4 subdeterminants of $\mathbf{M}$ must vanish. Furthermore, the structure of $\mathbf{M}$ can be utilized, as it may be written as

$$\mathbf{M} = \begin{bmatrix} -\xi & 0 & m_{13} & m_{14} \\ 0 & -\xi & m_{23} & m_{24} \\ 0 & 0 & m_{33} & m_{34} \\ 0 & 0 & m_{43} & m_{44} \\ 0 & 0 & m_{53} & m_{54} \end{bmatrix}, \quad (15)$$

where $\xi = f_{\text{ref}}f(1+r^2)$. It follows that only three subdeterminants $\text{subdet}(\mathbf{M})_{i,j}$ are non-trivially zero and these can be written as

$$\text{subdet}(\mathbf{M})_{i,j} = (1+r^2)^2 f^2 h_{i,j}(r, f), \quad (16)$$

where $h_{i,j}$ are polynomials in the monomial basis $\mathcal{B} = \{r^4 f, r^4, r^3 f, r^3, r f, r, f, 1\}$. As $1+r^2 \neq 0$, due to our choice of parameterization, we can simply ignore these roots. Similarly, $f = 0$ can be discarded as a physically improbable focal length. Additionally, only two of the three equations are linearly independent, and we may discard the second equation. The resulting system of equations is linear in f , hence we may repeat the same argument as before, and write

$$(1+r^2)^2 \bar{\mathbf{M}}(r) \begin{bmatrix} f \\ 1 \end{bmatrix} = \mathbf{0}, \quad (17)$$

where $\bar{\mathbf{M}}$ is a 2×2 matrix. Again, $\bar{\mathbf{M}}$ has a non-trivial nullspace. The resulting equation $\bar{h}(r) = \det(\bar{\mathbf{M}})(r)$, is quartic, hence an analytical solution is available. In general, there are four solutions, and it can be shown that the original system of equations also has four solutions. Next, we use (17) to extract f using backsubstitution, followed by extracting $\mathbf{t}_{\text{query}}$ from (14) in a similar manner. Now, the previously unused constraint (10) can be used to identify the correct solution, selecting the one that minimizes its residual. This guarantees that a single solution is obtained, which is beneficial in a RANSAC-like framework, as it limits the number of hypotheses to test against the full consensus set.

3.4 Derivation of UP2PforI

We treat the case of two orientation-covariant features in a similar manner. Again, there are six equations and only five unknowns, and we will use only one equation of the form (11). Note that (11) is independent of $\mathbf{t}_{\text{query}}$. We use the point-based equations (12)–(13), which give us four equations which are linear in $\mathbf{t}_{\text{query}}$, and proceed as before, by starting from the (partial) system of equations (14) where $\mathbf{M}$ now is a 4×4 matrix. Again, it has a non-trivial nullspace, hence $\det(\mathbf{M}) = 0$. The structure of $\mathbf{M}$ for this case is

$$\mathbf{M} = \begin{bmatrix} -\xi_1 & 0 & m_{13} & m_{14} \\ 0 & -\xi_1 & m_{23} & m_{24} \\ -\xi_2 & 0 & m_{33} & m_{34} \\ 0 & -\xi_2 & m_{43} & m_{44} \end{bmatrix}. \quad (18)$$

where $\xi_i = f_{\text{ref},i}f(1+r^2)$. Repeating the same argument previously used, the nullspace is non-trivial, and $\det(\mathbf{M}) = (1+r^2)^2 f^2 g(r, f)$, where

$$\begin{aligned} g(r, f) = & f_{\text{ref},2}^2(m_{13}m_{24} - m_{14}m_{23}) + f_{\text{ref},1}^2(m_{33}m_{44} - m_{34}m_{43}) \\ & + f_{\text{ref},1}f_{\text{ref},2}(m_{14}m_{43} + m_{23}m_{34} - m_{24}m_{33} - m_{13}m_{44}). \end{aligned} \quad (19)$$

Furthermore, it can be shown that $g(r, f) = (1+r^2)\bar{g}(r, f)$, due to the structure of the elements m_{ij} . The resulting equation $\bar{g}(r, f) = 0$ can now be used in conjunction with one of the orientation covariant constraints (11) to form a system of equations with two equations in the two remaining unknowns r and f . This system is again linear in f , and we may write it as

$$\bar{\mathbf{M}}(r) \begin{bmatrix} f \\ 1 \end{bmatrix} = \mathbf{0}, \quad (20)$$

hence $\det(\bar{\mathbf{M}}) = 0$ gives a final equation in the unknown r . This is again a quartic equation and can be solved analytically. It can be shown that the original system also has four solutions, and we may use the second (previously unused) equation (11) to single out the best hypotheses among these solutions.

4 Experiments

In our experiments we include the P4Pf solver by Kukelova *et al.* [39] and the P3.5Pf solver by Larsson *et al.* [40]. The method by Wu [63] is reported to perform similarly to the latter and is therefore omitted. Among the gravity-aware solvers we include the recent UP1SIFT solver by Ventura *et al.* [61] and two solvers from [38]: the point-based minimal two-point, UP2P, and a semi-calibrated point-based UP2.5Pf solver, which was re-implemented from the UP3Pfk solver, assuming the radial distortion parameter $k = 0$. Furthermore, we include the affine P1AC solver [60]. Note that among these competing state-of-the-art methods, only the P4Pf, P3.5Pf and UP2.5Pf solvers estimate the unknown focal length, while the others assume this known and accounted for. Furthermore, among the methods using more than one correspondence, only the proposed UP2PfORI is capable of utilizing correspondences across multiple reference images simultaneously.

4.1 Synthetic experiments

To understand the characteristics of the proposed solvers we generate a synthetic scene by adopting the methodology used in [23, 60, 61]. Camera centers are generated for reference and query cameras, with the query positioned 2 units from the world coordinate system origin. The camera orientation is randomly generated to face the scene and the known gravity direction is obtained by decomposing it. A single reference camera is positioned at the origin with unit orientation, and random 3D points within a cube of side length 2 are generated. The unknown query focal length is chosen uniformly random in the range [200, 1200], and the 3D points are projected through the cameras to generate the image point correspondences. Following [4], we generate local affine features by first estimating a homography corresponding to each of the 3D world points, which requires a 3D normal unit vector that is randomly drawn from a normal distribution. When the local affine transformation matrix is computed, the scales and orientations are extracted, as outlined in [61]. Physically improbable configurations, where 3D points are behind the cameras, the affine transformation does not have a positive determinant, or the rotation between query and reference cameras is greater than 180° are discarded. The following metrics are used:

Rotation error: The standard angular distance (in degrees) between the estimated $\mathbf{R}_{\text{est}}$ and the ground truth rotation $\mathbf{R}_{\text{gt}}$ is used,

$$\mathbf{R}_{\text{err}} := \arccos\left(\frac{\text{tr}(\mathbf{R}_{\text{gt}}^T \mathbf{R}_{\text{est}}) - 1}{2}\right), \quad (21)$$

which can be seen as the minimum angle that aligns the rotations [30, 53].

Position error: The distance between the estimated camera center $\mathbf{c}_{\text{est}}$ and the ground truth rotation $\mathbf{c}_{\text{gt}}$, defined as $\mathbf{t}_{\text{err}} := \|\mathbf{c}_{\text{gt}} - \mathbf{c}_{\text{est}}\|$.

Focal length error: The normalized focal length error $f_{\text{err}} := |f_{\text{gt}} - f_{\text{est}}|/f_{\text{gt}}$.

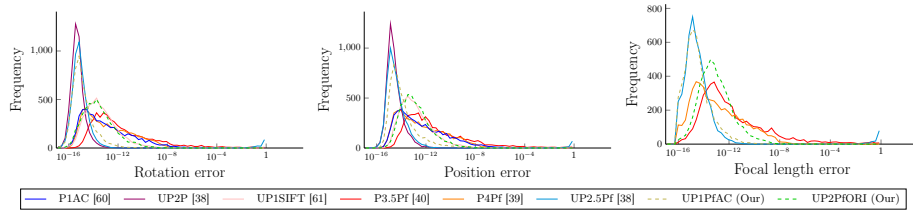


Fig. 2: Numerical stability. Frequency histogram (5000 trials) of $\log_{10}$ errors in rotation, translation, and focal length estimation across different solvers, evaluated on noise-free data. A more stable method results in a distribution shifted further to the left. Best viewed in color.

Numerical stability We generated 5000 random configurations of the synthetic environment described above, and statistics for each method are collected on them all. No noise is added to demonstrate the numerical stability of each solver, and the results are shown in Fig. 2. All solvers should be considered numerically stable, with the majority of errors less than 10^{-12} . Among the methods estimating focal length, the proposed UP1PfAC solver and UP2.5Pf are the most stable.

Sensitivity to noise Next, we inject Gaussian noise with zero mean and varying standard deviation to the image points to simulate matching errors naturally occurring in real localization pipelines. Image point noise is added with a constant standard deviation of 1.2 pixels; this value was chosen to represent the actual real-world scenarios, as most modern features typically exhibit errors up to approximately 1.2 pixels [56], although methods achieving subpixel accuracy exist [35]. In another experiment to simulate IMU noise, the gravity vector is perturbed with a random vector drawn from a normal distribution with zero mean and varying standard deviation. Modern consumer devices have an IMU error of up 0.2 degrees, but high-end products may be below 0.05 degrees [38], which justifies the ranges of the varying noise input.

The results are shown in Fig. 3, where the median value over 1000 random test instances is reported for each noise level and method. Of particular importance is the focal length errors, as the proposed methods are the first among the category of affine-based and orientation-covariant solvers to incorporate unknown focal length. Regarding sensitivity to image noise, the proposed methods perform superior, and even with the addition of extra IMU noise, the proposed methods show an advantage for reasonable consumer device levels of IMU noise.

Execution time All solvers are implemented in C++, compiled with the same optimization flags, and executed with a standard laptop equipped with an 11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60GHz. We use public implementations where available: the P4Pf [39] and UP2P [38] are included in PoseLib [42]³,

³ <https://github.com/PoseLib/PoseLib>

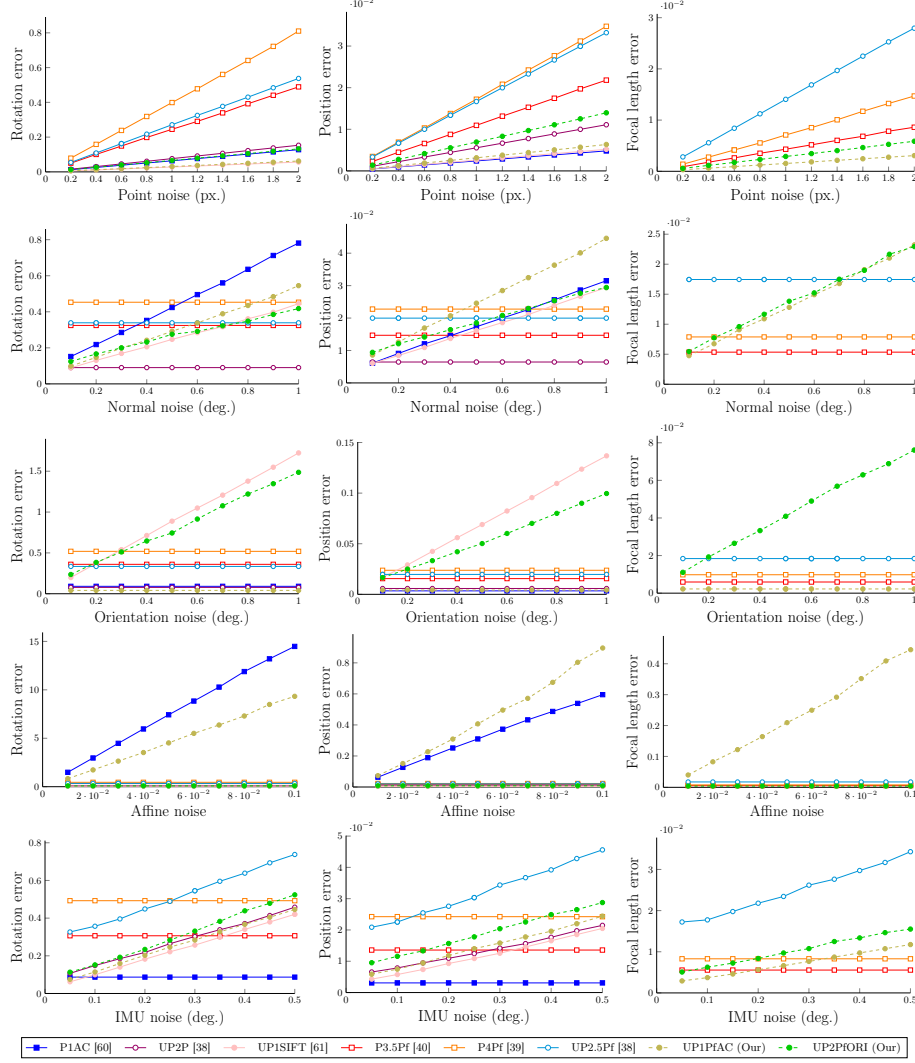


Fig. 3: Noise sensitivity. Median error w.r.t. image noise, normal vectors, feature orientations, affine features, and gravity vector. All measurements report the median value over 1000 random test instances per noise level, where the x -axis shows the standard deviation of the added Gaussian noise. *Visualization code:* circular and square markers indicate upright and non-upright orientations, respectively; filled markers denote point-based solvers, while white markers indicate alternative approaches; and dashed lines highlight our contributions. Semi-calibrated approaches are in the right column. Best viewed in color.

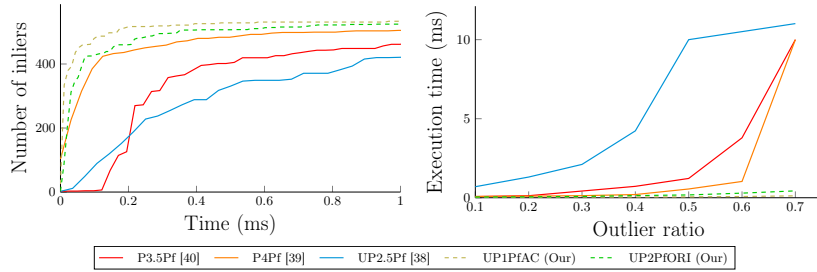


Fig. 4: *Robust estimation.* (Left): Number of inliers found *vs* execution time in a RANSAC framework with a 50 % outlier ratio. (Right): Outlier ratio *vs* total execution time to reach 80 % of the true inlier set. Best viewed in color.

and P1AC [60]⁴ and UP1SIFT [61]⁵ are available directly from the authors’ repositories. The P3.5Pf [40] solver was re-implemented based on the automatic Gröbner basis solver [41] proposed by the first author.

Table 1: Median timings (in nanoseconds).

P1AC [60]	UP2P [38]	UP1SIFT [61]	P3.5Pf [40]	P4Pf [39]	UP2.5Pf [38]	UP1PfAC	UP2PfORI
2740	484	1448	19118	3179	642	2586	2149

Integration with robust estimation frameworks In this section, we integrate the solvers in a standard RANSAC framework to demonstrate how they cope with outliers. We only consider solvers where the focal length is assumed unknown, as this is the intended application of the proposed solvers. We generate 1000 correspondences in each scene and use a 1.2 pixel noise and an IMU noise of 0.2 degrees. Furthermore, outliers are injected by corrupting the data randomly. First, we consider a scenario with a 50 % outlier ratio and compare the total number of inliers found by the respective solvers as a function of time. To distinguish outliers from inliers a threshold of 5 pixels is used. Fig. 4 reports the median values over 100 random problem instances. Both of the proposed solvers display advantageous performance compared to the state-of-the-art.

Furthermore, we vary the outlier ratio and report the time it takes the corresponding solvers to find an inlier set containing at least 80 % of the true inliers. The results are shown to the right in Fig. 4. Here, the benefits of using more than point-based data become clearer, where the increasing exponential relationship is due to the minimal sample sizes of the corresponding solvers.

4.2 Real data experiments

We evaluate the proposed solvers in a visual localization scenario based on real data, where a query device must localize in the map created beforehand. For this, we use two well-known datasets: the Cambridge Landmarks [34] and the

⁴ <https://github.com/jonathanventura/P1AC>

⁵ <https://github.com/danini/absolute-pose-from-oriented-and-scaled-features>

Aachen Day-Night v1.1 [53]. The Cambridge dataset consists of images gathered from video while walking at five different scenes in Cambridge (UK). Each scene has a train and a test split, using the first for the reconstruction and the latter as localization queries. We re-triangulate the maps by using ground-truth from VisualSfM [62], where the intrinsic and extrinsic calibration is obtained for each camera. This is a relatively easy dataset because images are captured from the same device, the scene scale is medium and visual conditions are similar. We report median errors for position, rotation and focal length, as well as the median execution time. The Aachen Dataset consists of images captured at the historical center of Aachen (Germany), a larger scale environment recorded with several different devices, which resembles more the XR scenario. It considers changes in visual appearance, making it a more challenging benchmark, especially for semi-calibrated methods. We triangulate the models based on the poses from [67], and assume all query images to be upright. We follow the evaluation protocol and report the percentage of images localized within the thresholds ($0.25\text{m}/2^\circ$, $0.5\text{m}/5^\circ$, $5.0\text{m}/10^\circ$), as well as the median focal length estimation error.

Robust estimation method We integrate our and the other solvers in the Graph-Cut-RANSAC framework [9], a locally optimized RANSAC alternating graph-cut and model re-fitting in the local optimization step. For the non-minimal refinement in the case of partially calibrated absolute pose estimation, we use an initial estimate from Direct Linear Transform (DLT) from the inliers of the minimal solver, which is then refined by 10 iterations of Levenberg-Marquardt.

Visual Localization Procedure We followed the pipeline proposed by HLOC [51]. First, we triangulated the models from the available poses using COLMAP [54]. For that, we used image retrieval to recover a subset of covisible database images for each database image, using the top-20 DenseVLAD [58] retrievals. Then, we extracted and matched the features from the images in the database, using two possible configurations: RootSIFT [1]+Nearest Neighbor or SuperPoint [17]+LightGlue [44]. Then, we retriangulate the model to obtain the 3D reconstructions. For localization, we retrieve the top-10 most similar database images for each query. For each of those retrieved database images, we lift the 2D-2D correspondences to 2D-3D correspondences, picking the 3D points from the SfM model related to each 2D matched keypoint in the database image. From all database images, we select the estimate with the highest number of inliers.

In the case of SIFT, the Difference-of-Gaussians (DoG) extractor provides an estimated scale and orientation. For SuperPoint, we estimate them through S3Esti [64]. These parameters can then be used to approximate the affine matrix A_i by $A_i \approx S_{q_i} R_{\alpha_i}$, with the diagonal matrix $S_{q,i}$ scaled uniformly by the detected scale factor q_i , and the rotation matrix $R_{\alpha_i} \in \text{SO}(2)$, by the estimated angle $\alpha_i \in [0, 2\pi)$. Normals were estimated using the closest 200 nearest neighbors.

Localization results Tab. 2 and Tab. 3 report performance in Cambridge dataset using the SuperPoint and the SIFT pipelines, respectively. The results show that, for both features, our solvers improve the estimation errors committed by

the other semi-calibrated solvers in all scenes, while also improving the time. Compared with the calibrated methods, our solvers improve performance in rotation estimation for all scenes but gets worse translation errors due to the focal length ambiguity. We observe a slight increase in terms of time, attributed to the different nonlinear refinement method used for the case of unknown focal length. Results show a better performance in general with SuperPoint, but the performance gap gets smaller when the solver is accurate enough, as in our case. Tab. 4 shows performance for both SuperPoint and SIFT setups, following a similar trend than in Cambridge. As translation and rotation errors are compacted in a single metric, results show a performance degradation in case of partially calibrated methods, with our methods being better. In general, we observed that the non-linear refinement method homogenizes the final performance of the partially calibrated solvers, the main difference lying in the time.

Table 2: Cambridge Landmarks Median errors in position, rotation and focal length, and median times for the tested solvers in GC-RANSAC, using SP+LG and S3Esti, best method highlighted differently for *calib* and **semi-calib** methods.

	GreatCourt				KingsCollege				OldHospital				ShopFacade				StMarysChurch				
	cm	°	f_{err}	ms	cm	°	f_{err}	ms	cm	°	f_{err}	ms	cm	°	f_{err}	ms	cm	°	f_{err}	ms	
Known f	P1AC [60]	32.6	0.16	-	23.3	16.6	0.28	-	32.4	21.2	0.36	-	35.2	6.7	0.32	-	36.8	11.2	0.35	-	34.3
	UP2P [38]	31.6	0.15	-	23.1	16.1	0.28	-	33.4	21.1	0.35	-	31.8	6.5	0.30	-	34.1	10.8	0.33	-	33.9
	UP1SIFT [61]	45.9	0.20	-	22.5	20.4	0.32	-	34.0	44.5	0.64	-	31.2	8.4	0.38	-	33.5	17.0	0.55	-	29.1
	P4Pf [39]	61.5	0.14	0.010	26.9	36.3	0.30	0.011	48.3	58.7	0.42	0.015	40.9	14.6	0.30	0.011	39.9	23.9	0.32	0.014	42.3
Unknown f	P3.5Pf [40]	61.3	0.14	0.010	28.0	36.0	0.30	0.011	45.2	58.0	0.42	0.015	38.6	14.6	0.30	0.011	41.4	23.9	0.32	0.014	41.2
	UP2.5Pf [38]	61.6	0.14	0.010	23.8	35.9	0.29	0.011	41.1	57.2	0.42	0.015	38.0	14.4	0.30	0.011	35.4	23.7	0.32	0.013	37.0
	UP1PfAC	57.7	0.13	0.009	22.6	35.6	0.28	0.012	39.9	52.3	0.38	0.013	34.2	10.6	0.26	0.008	33.9	22.6	0.31	0.014	34.3
	UP2PfORI	57.9	0.13	0.009	24.0	35.3	0.28	0.012	41.7	52.2	0.37	0.014	37.1	10.9	0.26	0.008	34.3	22.9	0.31	0.014	38.2

Table 3: Cambridge Landmarks Median errors in position, rotation and focal length, and median times the tested solvers in GC-RANSAC, using SIFT and Nearest Neighbors, best method highlighted differently for *calib* and **semi-calib** methods.

	GreatCourt				KingsCollege				OldHospital				ShopFacade				StMarysChurch				
	cm	°	f_{err}	ms	cm	°	f_{err}	ms	cm	°	f_{err}	ms	cm	°	f_{err}	ms	cm	°	f_{err}	ms	
Known f	P1AC [60]	59.5	0.21	-	15.2	17.8	0.28	-	34.5	26.6	0.45	-	20.1	8.1	0.30	-	21.1	14.9	0.42	-	22.4
	UP2P [38]	55.1	0.19	-	16.0	17.4	0.28	-	34.8	25.1	0.42	-	20.5	6.8	0.28	-	20.8	14.8	0.40	-	22.5
	UP1SIFT [61]	58.4	0.20	-	15.4	18.6	0.29	-	36.3	29.4	0.48	-	18.9	7.6	0.30	-	20.3	15.4	0.42	-	21.3
	P4Pf [39]	75.8	0.16	0.011	23.3	37.9	0.31	0.012	41.9	51.1	0.39	0.017	25.9	11.2	0.28	0.008	23.6	28.7	0.37	0.018	28.4
Unknown f	P3.5Pf [40]	75.8	0.16	0.011	22.7	38.0	0.32	0.012	42.7	51.7	0.39	0.017	27.8	11.2	0.28	0.008	25.8	28.6	0.37	0.018	31.2
	UP3Pf [38]	75.6	0.16	0.011	20.5	37.8	0.31	0.012	40.1	51.9	0.39	0.017	24.2	11.2	0.28	0.008	20.6	28.4	0.38	0.018	26.4
	UP1PfAC	70.5	0.15	0.012	16.7	37.6	0.30	0.012	35.5	48.4	0.33	0.018	21.0	11.1	0.27	0.008	20.8	27.7	0.38	0.017	24.2
	UP2PfORI	70.8	0.15	0.011	17.7	37.8	0.30	0.012	35.4	50.4	0.36	0.018	23.8	11.0	0.27	0.008	21.6	28.1	0.38	0.018	27.1

5 Discussion

In general, we observed that the 1-point UP1PfAC works better than UP2PfORI. It should be noticed that the affine constraints encode significantly more information than the orientation-covariant constraint alone. It therefore acts as a

Table 4: Aachen Dataset Performance for various solvers and features using GC-RANSAC, with percentage error recalls for thresholds $0.25\text{m}/2^\circ$, $0.5\text{m}/5^\circ$, $5.0\text{m}/10^\circ$ and f_{err} . Best method highlighted differently for *calib* and **semi-calib** methods.

		SuperPoint+LightGlue+S3Esti						SIFT+Nearest Neighbors									
		Day		Night		f_{err}	ms	Day		Night		f_{err}	ms				
Known f	P1AC [60]	69.5	89.6	99.2	70.7	88.0	99.0	-	15.7	59.8	76.3	87.7	21.5	25.7	27.7	-	29.1
	UP2P [38]	71.8	90.0	99.3	71.2	87.4	98.4	-	15.6	61.3	76.9	87.9	21.5	25.7	27.7	-	26.9
	UP1SIFT [61]	60.9	84.2	93.8	59.7	79.6	92.7	-	16.4	59.0	75.0	87.0	19.4	23.0	26.2	-	24.8
Unknown f	P4Pf [39]	46.1	70.0	96.8	54.5	74.9	95.8	0.010	16.7	37.4	57.0	85.9	19.4	20.9	26.2	0.019	37.9
	P3.5Pf [40]	45.6	69.7	96.7	54.5	74.9	95.3	0.010	20.3	37.7	57.8	85.7	19.4	21.5	26.2	0.018	34.8
	UP2.5Pf [38]	46.5	68.8	96.6	55.0	73.8	95.3	0.010	18.6	37.9	57.4	85.4	18.8	21.5	26.2	0.018	32.2
	UP1PiAC	48.5	70.8	97.9	60.2	80.1	97.4	0.009	16.7	38.8	58.6	86.3	18.3	22.0	27.2	0.017	28.0
	UP2Pfori	48.4	71.2	98.2	62.3	80.1	97.4	0.009	19.8	39.7	58.4	86.3	19.4	22.5	26.7	0.017	30.8

stronger geometric prior; but at a higher computational cost compared to, *e.g.*, ORB. The proposed 2-point solver effectively bridges this gap, offering a compelling balance between geometric expressiveness and computational efficiency.

While modern robust estimation frameworks are equipped with numerous features, it is often unclear what truly drives their performance. Historically, research emphasized the development of fast and stable minimal solvers, but this focus has recently been reassessed [18, 37]. Factors such as the number of generated solutions, the choice of nonlinear refinement, and sampling strategies may be equally important—yet the speed of the minimal solver remains critical.

Our proposed method is computationally efficient, but real-world data introduces multiple error sources, including image noise, normal noise from 3D reconstruction, descriptor mismatches, and auxiliary sensor inaccuracies (*e.g.*, from IMUs). Given these compounded errors, the reliability of any method is a valid concern. We demonstrate that, within a modern estimation framework, strong performance is still achievable: reasonably accurate hypotheses generated speedily, combined with effective nonlinear refinement, is sufficient to yield robust results. With this in mind, we believe the future of minimal solvers remains bright, provided they can continue to leverage diverse sources of information, including deep learning-based descriptors, geometric priors, and auxiliary sensor data. Integrating these cues will be key to pushing the boundaries of what is possible in robust estimation and the applications thereof.

6 Conclusions

In this paper, we have proposed two novel solvers for the absolute pose problem with unknown focal length. To this date, these are the first solvers available in the literature that incorporate affine- or orientation-covariant features for this problem configuration. In order to derive the solvers, novel polynomial constraints were derived and incorporated. Our results demonstrate that by utilizing more than point-based information extracted from the feature descriptors, we are able to significantly improve the execution speed in robust estimation frameworks compared to state-of-the-art methods, while maintaining superior or equal accuracy, both in synthetic environments and in localization tasks. Solvers of this kind pave the way for future generations of low-energy localization for constrained devices, *e.g.*, XR devices, UAVs, and embedded robotics platforms.

Acknowledgements

This publication is part of the grant JDC2024-055088-I, funded by MICIU/AEI/10.13039/501100011033 and the FSE+.

References

1. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2911–2918 (2012)
2. Banglei, G., Su, A., Li, Z., Fraundorfer, F.: Rotational alignment of imu-camera systems with 1-point ransac. In: Pattern Recognition and Computer Vision: Second Chinese Conference (PRCV) (2019)
3. Barath, D.: Making sift features affine covariant. *International Journal of Computer Vision (IJCV)* **131** (2023)
4. Barath, D., Hajder, L.: Novel ways to estimate homography from local affine transformations. In: Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. pp. 434–445 (2016)
5. Barath, D., Hajder, L.: Efficient recovery of essential matrix from two affine correspondences. *IEEE Transactions on Image Processing* **27**, 5328–5337 (2018)
6. Barath, D., Kukulova, Z.: Homography from two orientation- and scale-covariant features. In: International Conference on Computer Vision (ICCV) (2019)
7. Barath, D., Kukulova, Z.: Relative pose from SIFT features. In: European Conference on Computer Vision (ECCV) (2022)
8. Barath, D., Matas, J.: Graph-cut RANSAC. In: Conference on Computer Vision and Pattern Recognition (2018)
9. Barath, D., Matas, J.: Graph-cut ransac. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6733–6741 (2018)
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) (2020)
11. Barath, D., Toth, T., Hajder, L.: A minimal solution for two-view focal-length estimation using two affine correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
12. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: European Conference on Computer Vision (ECCV). pp. 404–417 (2006)
13. Bentolila, J., Francos, J.M.: Conic epipolar constraints from affine correspondences. *Computer Vision and Image Understanding* **122**, 105–114 (2014)
14. Bujnak, M., Kukulova, Z., Pajdla, T.: A general solution to the p4p problem for camera with unknown focal length. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (06 2008)
15. Campos, C., Elvira, R., Rodríguez, J.J.G., M. Montiel, J.M., D. Tardós, J.: Orbslam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics* **37**(6), 1874–1890 (2021)
16. Chum, O., Matas, J., Kittler, J.: Locally optimized ransac. In: Pattern Recognition. pp. 236–243 (2003)
17. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 224–236 (2018)

18. Ding, Y., Astermark, J., Oskarsson, M., Larsson, V.: Noisy one-point homographies are surprisingly good. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5125–5134 (2024)
19. Ding, Y., Barath, D., Kukelova, Z.: Homography-based egomotion estimation using gravity and sift features. In: Asian Conference on Computer Vision (ACCV) (2020)
20. Ding, Y., Barath, D., Yang, J., Kong, H., Kukelova, Z.: Globally optimal relative pose estimation with gravity prior. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 394–403 (2021)
21. Ding, Y., Vávra, V., Bhayani, S., Wu, Q., Yang, J., Kukelova, Z.: Fundamental matrix estimation using relative depths. In: European Conference on Computer Vision (ECCV) (2024)
22. Ding, Y., Yang, J., Larsson, V., Olsson, C., Åström, K.: Revisiting the P3P problem. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4872–4880 (2023)
23. Eichhardt, I., Chetverikov, D.: Affine correspondences between central cameras for rapid relative pose estimation. In: European Conference on Computer Vision (ECCV) (2018)
24. Fischler, M., Bolles, R.: 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)
25. Fragoso, V., DeGol, J., Hua, G.: gDLS*: Generalized pose-and-scale estimation given scale and gravity priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2210–2219 (2020)
26. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(8), 930–943 (2003)
27. Guan, B., Zhao, J., Barath, D., Fraundorfer, F.: Minimal cases for computing the generalized relative pose using affine correspondences. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6068–6077 (2021)
28. Guan, B., Zhao, J., Li, Z., Sun, F., Fraundorfer, F.: Minimal solutions for relative pose with a single affine correspondence. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
29. Hajder, L., Barath, D.: Relative planar motion for vehicle-mounted cameras from a single affine correspondence. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 8651–8657 (2020)
30. Hartley, R.I., Trunf, J., Dai, Y., Li, H.: Rotation averaging. *International Journal of Computer Vision* **103**, 267 – 305 (2013)
31. Hesch, J.A., Roumeliotis, S.I.: A direct least-squares (DLS) method for PnP. In: International Conference on Computer Vision (ICCV). pp. 383–390 (2011)
32. Hruby, P., Pollefeys, M., Barath, D.: Semicalibrated relative pose from an affine correspondence and monodepth. In: European Conference on Computer Vision (ECCV) (2024)
33. Ke, T., Roumeliotis, S.I.: An efficient algebraic solution to the perspective-three-point problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4618–4626 (2017)
34. Kendall, A., Grimes, M., Cipolla, R.: PoseNet: A convolutional network for real-time 6-DoF camera relocalization. In: IEEE International Conference on Computer Vision (ICCV). pp. 2938–2946 (2015)
35. Kim, S., Pollefeys, M., Barath, D.: Learning to make keypoints sub-pixel accurate. In: European Conference on Computer Vision (ECCV) (2024)

36. Klein, G., Murray, D.: Parallel tracking and mapping for small ar workspaces. In: IEEE and ACM International Symposium on Mixed and Augmented Reality. pp. 225–234 (2007)
37. Kocur, V., Tzamos, C., Ding, Y., Haladova, Z.B., Sattler, T., Kukulova, Z.: Are minimal radial distortion solvers really necessary for relative pose estimation? *International Journal of Computer Vision* **134**(2), 48 (2026)
38. Kukulova, Z., Bujnak, M., Pajdla, T.: Closed-form solutions to minimal absolute pose problems with known vertical direction. In: Asian Conference on Computer Vision (ACCV). pp. 216–229 (2010)
39. Kukulova, Z., Bujnak, M., Pajdla, T.: Efficient intersection of three quadrics and applications in computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1799–1808 (2016)
40. Larsson, V., Kukulova, Z., Zheng, Y.: Making Minimal Solvers for Absolute Pose Estimation Compact and Robust. In: International Conference on Computer Vision (ICCV). pp. 2335–2343 (2017)
41. Larsson, V., Astrom, K., Oskarsson, M.: Efficient solvers for minimal problems by syzygy-based reduction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
42. Larsson, V., contributors: PoseLib - Minimal Solvers for Camera Pose Estimation (2020), <https://github.com/vlarsson/PoseLib>
43. Li, H., Chen, W., Zhao, J., Bazin, J.C., Luo, L., Liu, Z., Liu, Y.H.: Robust and efficient estimation of absolute camera pose for monocular visual odometry. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 2675–2681 (2020)
44. Lindenberger, P., Sarlin, P.E., Pollefeys, M.: Lightglue: Local feature matching at light speed. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 17627–17638 (2023)
45. Lowe, D.G.: Object recognition from local scale-invariant features. In: IEEE International Conference on Computer Vision (CVPR). vol. 2, pp. 1150–1157 (1999)
46. Mishkin, D., Radenović, F., Matas, J.: Repeatability is not enough: Learning affine regions via discriminability. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) European Conference on Computer Vision (ECCV). pp. 287–304 (2018)
47. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* **2**, 438–469 (2009)
48. Persson, M., Nordberg, K.: Lambda twist: An accurate fast robust perspective three point (p3p) solver. In: European Conference on Computer Vision (ECCV) (2018)
49. Raposo, C., Barreto, J.P.: Theory and practice of structure-from-motion using affine correspondences. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
50. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: International Conference on Computer Vision (ICCV). pp. 2564–2571 (2011)
51. Sarlin, P.E., Cadena, C., Siegwart, R., Dymczyk, M.: From coarse to fine: Robust hierarchical localization at large scale. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
52. Sattler, T., Leibe, B., Kobbelt, L.: Efficient & effective prioritized matching for large-scale image-based localization. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(9), 1744–1756 (Sep 2017)

53. Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., Kahl, F., Pajdla, T.: Benchmarking 6DOF outdoor visual localization in changing conditions. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8601–8610 (2018)
54. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4104–4113 (2016)
55. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph.* **25**(3), 835–846 (Jul 2006)
56. Suwanwimolkul, S., Komorita, S., Tasaka, K.: Learning of low-level feature keypoints for accurate and robust detection. In: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 2262–2271 (2021)
57. Sweeney, C., Flynn, J., Nuernberger, B., Turk, M., Höllerer, T.: Efficient computation of absolute pose for gravity-aware augmented reality. In: IEEE International Symposium on Mixed and Augmented Reality. pp. 19–24 (2015)
58. Torii, A., Arandjelovic, R., Sivic, J., Okutomi, M., Pajdla, T.: 24/7 place recognition by view synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1808–1817 (2015)
59. Valtonen Örnham, M., Jaenal, A.: Leveraging scale- and orientation-covariant features for planar motion estimation. In: European Conference on Computer Vision (ECCV) (2024)
60. Ventura, J., Kukelova, Z., Sattler, T., Baráth, D.: P1AC: Revisiting absolute pose from a single affine correspondence. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 19751–19761 (2023)
61. Ventura, J., Kukelova, Z., Sattler, T., Baráth, D.: Absolute pose from one or two scaled and oriented features. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20870–20880 (2024)
62. Wu, C.: Towards linear-time incremental structure from motion. In: International Conference on 3D Vision (3DV). pp. 127–134. IEEE (2013)
63. Wu, C.: P3.5P: Pose estimation with unknown focal length. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2440–2448 (2015)
64. Yan, P., Tan, Y., Xiong, S., Tai, Y., Li, Y.: Learning soft estimator of keypoint scale and orientation with probabilistic covariant loss. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19406–19415 (2022)
65. Yu, Z., Guan, B., Liang, S., Li, Z., Ye, S., Yu, Q.: Globally optimal relative pose estimation using affine correspondences with known vertical direction. *IEEE Transactions on Instrumentation and Measurement* **72**, 1–12 (2023)
66. Yu, Z., Ye, S., Jin, R., Liang, S., Liu, Z., Zhang, H., Guan, B.: A minimal solver for relative pose estimation with unknown focal length from two affine correspondences. *IEEE Robotics and Automation Letters* **11**(2), 2290–2297 (2026)
67. Zhang, Z., Sattler, T., Scaramuzza, D.: Reference pose generation for long-term visual localization via learned features and view synthesis. *International Journal of Computer Vision (IJCV)* **129**(4), 821–844 (2021)
68. Zheng, Y., Sugimoto, S., Sato, I., Okutomi, M.: A general and simple method for camera pose and focal length determination. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 430–437 (2014)
69. Zhou, B., Chen, Z., Liu, Q.: An efficient solution to the perspective-n-point problem for camera with unknown focal length. *IEEE Access* **8**, 162838–162846 (2020)