The Probabilistic Normal Epipolar Constraint for Frame-To-Frame Rotation Optimization under Uncertain Feature Positions

Dominik Muhle[*†], Lukas Koestler[*†], Nikolaus Demmel[†], Florian Bernard[‡] and Daniel Cremers[†]
[†]Technical University of Munich [‡]University of Bonn

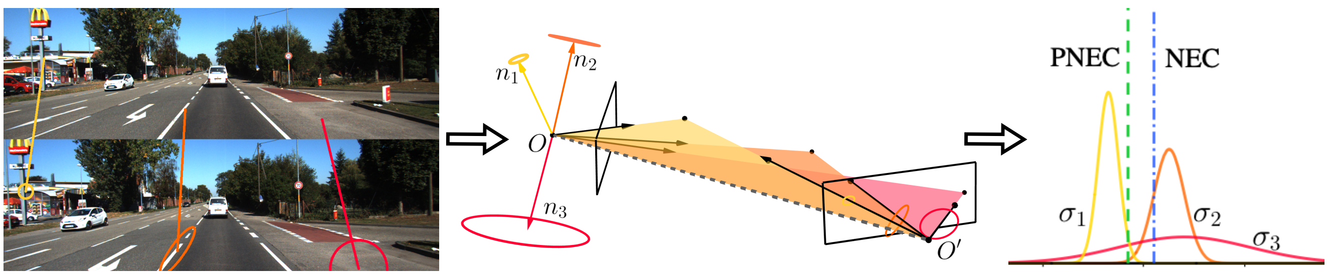


Figure 1. Illustration of the proposed method. The feature correspondences (*left: lines*) exhibit different position error distributions (*left: ellipses*). In contrast to the normal epipolar constraint (NEC) [34], our probabilistic normal epipolar constraint (PNEC) correctly accounts for the uncertainty and the geometry of the problem (*middle*) through a weighted (NEC) vs. unweighted (PNEC) averaging scheme (*right*). The NEC weighs all residuals equally (*right: dash-dotted blue line*), so that valuable information is lost. In contrast, our PNEC takes into account a proper weighting of individual residuals (*right: dashed green line*). Relative pose estimation with the PNEC instead of the NEC reduces the rotational-only versions of the RPE$_1$ and the RPE$_n$ error by up to 42% and 55% on the KITTI dataset [20], respectively.

Abstract

The estimation of the relative pose of two camera views is a fundamental problem in computer vision. Kneip et al. proposed to solve this problem by introducing the normal epipolar constraint (NEC). However, their approach does not take into account uncertainties, so that the accuracy of the estimated relative pose is highly dependent on accurate feature positions in the target frame. In this work, we introduce the probabilistic normal epipolar constraint (PNEC) that overcomes this limitation by accounting for anisotropic and inhomogeneous uncertainties in the feature positions. To this end, we propose a novel objective function, along with an efficient optimization scheme that effectively minimizes our objective while maintaining real-time performance. In experiments on synthetic data, we demonstrate that the novel PNEC yields more accurate rotation estimates than the original NEC and several popular relative rotation estimation algorithms. Furthermore, we integrate the proposed method into a state-of-the-art monocular rotation-only odometry system and achieve consistently improved results for the real-world KITTI dataset.

1. Introduction

Extracting the 3D geometry of a scene from images is a long-standing problem in computer vision and has numerous applications, including augmented and virtual reality, autonomous driving, or robots that can help with everyday life. One key component of many such approaches is the estimation of the relative pose between two viewpoints of a scene. For example, relative pose estimation is the foundation of geometric vision algorithms like structure from motion (SfM) or visual odometry (VO). Global SfM pipelines rely on accurate pairwise relative poses for use as fixed measurements in global motion averaging [23, 48]. In VO, relative pose estimation is used to construct a trajectory from a stream of images. Like for all odometry systems, small errors in the relative pose estimation lead to a drift in VO.

The most widely used concept for relative pose estimation is the essential matrix [43] in the calibrated case, or the fundamental matrix [25] in the general case. Respective approaches rely on correspondences between feature points, and are generally known to provide fast and accurate results [24]. However, approaches based on the essential matrix suffer from fundamental problems, with the most prominent being *solution multiplicity* [17, 25] and *planar degeneracy* [33]. To address such issues, often it is necessary to consider more involved solution strategies, which also lead to even more accurate relative poses as shown by

[*] equal contribution. Project page: https://go.vision.in.tum.de/pnec

Kneip *et al*. [33] and in this work.

To this end, Kneip *et al*. [34] proposed a constraint that avoids these problems. Their *epipolar plane normal coplanarity constraint* (in later works just *normal epipolar constraint*, NEC) allows the estimation of the rotation independent of the translation. A later work by Kneip and Lynen [33] provides a fast and reliable eigenvalue-based solver for the NEC, which allows for real-time relative pose estimation. This approach has been incorporated into rotation-only VO systems that estimate the rotation independent of the translation and has led to promising results [9, 39].

Yet, like many VO systems, neither of them considers the quality of the correspondences. After removing outliers from the feature matches, every match contributes equally to the final result. However, two-dimensional feature correspondences exhibit different error distributions depending on the content of the image and the specific method used to extract the correspondences, which can be seen in Fig. 2. A correspondence lying on an edge is accurately localized perpendicular to the edge and possesses higher position uncertainty parallel to it. This fine-grained information about the quality of the matches is completely ignored. It has been shown that considering the uncertainty is beneficial for fundamental matrix estimation [5]. While Kanazawa *et al*. [30] argue that the uncertainty needs to be sufficiently inhomogeneous to see the aforementioned benefit, our experiments show that the PNEC improves over the NEC even for homogeneous uncertainty due to the geometry of the problem.

The main objective of our work is to improve the accuracy of rotation estimation techniques. We achieve this based on the following technical contributions:

- We introduce the novel probabilistic normal epipolar constraint (PNEC), see Fig. 1, which for the first time makes it possible to incorporate uncertainty information into the normal epipolar constraint (NEC).
- We propose an efficient two-stage optimization strategy for the PNEC that achieves real-time performance.
- We analyse singularities in the PNEC energy function and address them with a simple regularization scheme.
- Experimentally, we compare our PNEC to several popular relative pose estimation algorithms, namely 8pt [24], 7pt [25], Stewenius 5pt [56], Nistér 5pt [51], and NEC [33], and demonstrate that our PNEC delivers more accurate rotation estimates. Moreover, we integrate our PNEC into a visual odometry system and achieve state-of-the-art results on real-world data.
- We publish the code for all experiments to facilitate future research.

2. Related Work

The focus of this paper is the integration of feature position uncertainties into frame-to-frame rotation estimation and the application to visual odometry. Hence, we restrict our discussion of related work to *relative pose estimation*, *uncertainty for feature correspondences*, and *visual odometry*. For a broader overview we refer the reader to the excellent books by Szeliski [57] and by Hartley and Zisserman [25] and to more topic-specific overview papers [6, 59].

Relative Pose Estimation. Estimating the relative pose between two viewpoints is a long-standing problem in computer vision with the first known solution proposed in 1913 by Kruppa [35]. Most methods either rely on previously computed feature correspondences (*feature-based*) or directly consider the intensity differences between the two images (*direct*). While direct methods have recently shown promising results [14, 15], they are currently limited to images that exhibit photo consistency and hence cannot be used for general problems, e.g. structure from motion. Feature-based methods are considerably more robust to viewpoint and appearance changes. Therefore, we use feature correspondences within this paper.

Given feature correspondences, many methods [36, 41, 43, 51, 55] estimate the essential matrix in the case of a calibrated camera, or the fundamental matrix in the general case. Nistér [51] proposes a minimal solution using polynomials and root bracketing, while the solver proposed by Longuet-Higgins [43] is linear and requires careful normalization for good performance [24]. Alternatively, the relative pose can be estimated directly using quaternions [16].

The essential matrix constraint deteriorates in zero-translation situations without noise, due to it being a zero matrix. Most essential-matrix-based algorithms estimate the correct motion only implicitly [33]. To address this problem, recent works have proposed algorithms that can estimate the rotation independent of the translation [34, 42]. Our work is based on the normal epipolar constraint (NEC) proposed by Kneip *et al*. [34] and the direct optimization scheme proposed in a follow-up paper [33]. Briales *et al*. [4] show how to obtain the global minimum for the NEC, however, their Shor relaxation is not applicable to our non-polynomial energy function.

Uncertainty for Feature Correspondences. Kanade-Lucas-Tomasi (KLT) tracks [45, 58] are widely used, and the position uncertainty has been extensively investigated [19, 53, 54, 65]. Based on the unscented transform [60], the position uncertainty has also been integrated directly into the KLT tracking [13]. Zeisl *et al*. [64] have shown a method to obtain anisotropic and inhomogeneous covariances for SIFT [44] and SURF [1] features.

The integration of the position uncertainty into the alignment problem has been studied from a statistical perspective [28, 29], in the photogrammetry community [47], as well as in the computer vision community [5, 30]. Brooks *et al*. [5] show that covariance information can be used beneficially if the estimated covariance is sufficiently accurate. Kanazawa *et al*. [30] question the practical use of covari-

Figure 2. Covariance ellipses for position uncertainties in KITTI seq. 07. The tracks are generated with KLT tracking. Our PNEC correctly considers such anisotropic inhomogeneous error distributions. For visualization purposes only a sub-image with subsampled and enlarged covariance ellipses is shown.

ance information if the covariance matrices are too similar and nearly isotropic, however, Fig. 2 shows clearly that the covariance matrices for real-world data are highly inhomogeneous and anisotropic.

Visual Odometry Systems. Most VO approaches utilize 3D-to-2D correspondences together with a sliding window formulation. Still, relative pose estimation is commonly used during initialization [49] and has been shown to provide excellent results on its own [11, 12]. Based on different solutions to the correspondence problem many visual odometry systems have been proposed. They include PTAM [31] and ORB-SLAM [49,50] with indirect features, approaches with KLT-tracks [61], as well as direct methods like LSD-SLAM [15] and DSO [14]. Common among these approaches is a deteriorating performance for pure rotation without inertial data.

Multiple rotation-only approaches that are robust against pure rotation have been proposed recently [9,10,39]. Chon-cha *et al*. [10] use the NEC to initialize a scale-consistent map even for purely rotational motion. Chng *et al*. [9] and Lee and Civera [39] use rotation averaging and rotation-only bundle adjustment on the basis of the NEC to further improve their results, respectively. While these approaches show promising results on existing datasets, neither of them utilizes uncertainty information of feature correspondences in their system. We base our VO evaluation on MRO [9].

3. Probabilistic Normal Epipolar Constraint

The normal epipolar constraint (NEC) [34] enforces the coplanarity of *epipolar plane normal vectors* constructed from feature correspondences. However, feature correspondences exhibit different error distributions that are not accounted for in the NEC. For example, an edge-like feature is well-localized perpendicular to the edge but not parallel to it, which is also known as the aperture problem. Fig. 2 clearly shows the anisotropicity and inhomogeneity of the position distributions. Moreover, it is well-known that correspondences in all areas of the image are required to sufficiently constrain the 3D geometry [14, 49] and thus we cannot simply discard feature points. To address this problem, we propose the probabilistic normal epipolar constraint (PNEC), which is able to take into account the uncertainty of feature positions by associating an anisotropic covariance matrix to each feature point.

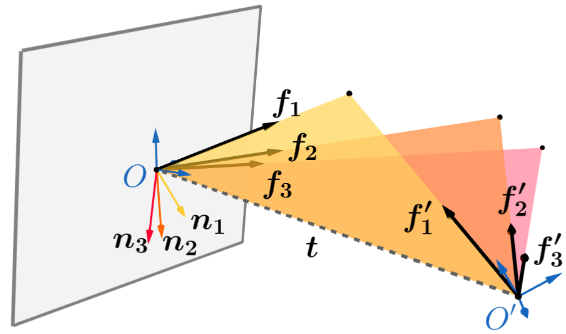


Figure 3. Geometry of the normal epipolar constraint (NEC) [34]. Feature correspondences are given by pairs of unit bearing vectors $\boldsymbol{f}_i$ and $\boldsymbol{f}'_i$ in the host frame (O) and target frame (O'), respectively. Each pair of bearing vectors spans an epipolar plane (*yellow, orange, red*), and has an associated normal vector $\boldsymbol{n}_i$ given in Eq. 1. All epipolar planes intersect in the line defined by the translation $\boldsymbol{t}$ (*dashed line*). The normal vectors span the *epipolar normal plane* (*gray*) that is orthogonal to $\boldsymbol{t}$. For visual clarity we show only three feature correspondences.

Notation. Vectors are denoted by bold lowercase letters (*e.g.* $\boldsymbol{f}$) and matrices by bold uppercase letters (*e.g.* $\boldsymbol{\Sigma}$). The hat operator applied to a vector $\boldsymbol{u} \in \mathbb{R}^3$ gives a skew-symmetric matrix $\hat{\boldsymbol{u}} \in \mathbb{R}^{3\times3}$ that computes the cross product between two vectors, i.e. $\boldsymbol{u} \times \boldsymbol{v} = \hat{\boldsymbol{u}}\boldsymbol{v}$. The superscript $^\top$ denotes the transpose. A rigid-body transformation is represented by a rotation matrix $\boldsymbol{R} \in SO(3)$ and a unit length translation $\boldsymbol{t} \in \mathbb{R}^3$ ($\|\boldsymbol{t}\| = 1$ is imposed since the two-view problem is scale-invariant).

3.1. Background – NEC

In the following we summarize the main idea of the NEC proposed in [34]. Given are a host frame and a target frame that observe at least five feature correspondences that are defined by pairs of unit bearing vectors $\boldsymbol{f}_i$ and $\boldsymbol{f}'_i$ in the host and target frame, respectively (see Fig. 3). A 3D point $\boldsymbol{x}'$ in the target frame is transformed into the host frame by applying the relative rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ s.t. $\boldsymbol{x} = \boldsymbol{R}\boldsymbol{x}' + \boldsymbol{t}$. In the ideal, error-free case, a single feature correspondence, together with the two viewpoints, creates an epipolar plane, represented by its normal vector

$$\boldsymbol{n}_i = \boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i. \tag{1}$$

All normal vectors are orthogonal to the translation and they span the *epipolar normal plane*.

The rotation is estimated by enforcing the coplanarity of the normal vectors. The residual of the model is given by the normalized epipolar error

$$e_i = |\boldsymbol{t}^\top \boldsymbol{n}_i| \,, \tag{2}$$

i.e. the Euclidean distance of a normal vector to the *epipolar normal plane*. An energy function

$$E(\boldsymbol{R}, \boldsymbol{t}) = \sum_i e_i^2 = \sum_i |\boldsymbol{t}^\top (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)|^2 \tag{3}$$

is constructed from the residuals. For a more detailed derivation, we kindly refer the reader to the original paper [34] or the recent paper by Lee and Civera [38], which offers numerous geometric interpretations of the NEC.

3.2. Deriving the PNEC

The probabilistic normal epipolar constraint (PNEC) extends the NEC by incorporating uncertainty. To be more specific, the PNEC allows the use of the anisotropic and inhomogeneous nature of the uncertainty of the feature position in the energy function. The feature position error is considered in the target frame as shown in Fig. 1 and we assume that the position error follows a 2D Gaussian distribution in the image plane with a known covariance matrix $\boldsymbol{\Sigma}_{\text{2D},i}$ per feature. In the supplementary we show how the covariance matrix can be extracted for KLT tracks from the KLT energy function using Laplace's approximation [3].

Given the 2D covariance matrix of the feature position in the target frame $\boldsymbol{\Sigma}_{\text{2D},i}$, we propagate it through the unprojection function using the unscented transform [60] in order to obtain the 3D covariance matrix $\boldsymbol{\Sigma}_i$ of the bearing vector $\boldsymbol{f}'_i$. Using the unscented transform ensures full-rank covariance matrices after the transform. We derive the details of the unscented transform in the supplementary material and show qualitative examples.

Propagating this distribution to the normalized epipolar error gives the probabilistic distribution of the residual. Due to the linearity of the transformations, the distribution of the residual is a univariate Gaussian distribution $\mathcal{N}(0, \sigma_i^2)$, with variance

$$\sigma_i^2(\boldsymbol{R}, \boldsymbol{t}) = \boldsymbol{t}^\top \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i{}^\top \boldsymbol{t} \,. \tag{4}$$

We integrate this variance into the cost-function of the NEC so that the Euclidean distance becomes the Mahalonobis distance [8] and define the **PNEC Energy Function**

$$E_P(\boldsymbol{R}, \boldsymbol{t}) = \sum_i \frac{e_i^2}{\sigma_i^2} = \sum_i \frac{|\boldsymbol{t}^\top (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)|^2}{\boldsymbol{t}^\top \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i{}^\top \boldsymbol{t}} \,, \tag{5}$$

which results in a weighted optimization problem. In the supplementary material we show a geometric interpretation of the above derivation.

Algorithm 1: PNEC Optimization Scheme

Initialize weights $\tilde{\sigma}_{i,0} \leftarrow 1 \; \forall i$
for $s \leftarrow 1$ **to** S **do**
  Optimize over $\boldsymbol{R}$ (cf. Sec. 4.3)
  $\boldsymbol{R}_s \leftarrow \text{Opt}_{\boldsymbol{R}} \; \lambda_{\text{min}}(\boldsymbol{M}_P(\boldsymbol{R}; \{\tilde{\sigma}_{i,s-1}\}_i))$
  Optimize over $\boldsymbol{t}$ (cf. Sec. 4.2)
  $\boldsymbol{t}_s \leftarrow \text{Opt}_{\boldsymbol{t}} \; E_P(\boldsymbol{R}_s, \boldsymbol{t})$
  Update the weights (cf. Eq. 4)
  $\tilde{\sigma}_{i,s} \leftarrow \sigma_i(\boldsymbol{R}_s, \boldsymbol{t}_s) \; \forall i$
end
Joint Refinement (cf. Sec. 4.4) using $(\boldsymbol{R}_S, \boldsymbol{t}_S)$ as starting value
$\boldsymbol{R}^*, \boldsymbol{t}^* \leftarrow \text{Opt}_{\boldsymbol{R},\boldsymbol{t}} \, E_P(\boldsymbol{R}, \boldsymbol{t})$

4. Optimization

To optimize the PNEC energy function Eq. 5, we propose a two-stage optimization scheme consisting of an alternating iterative optimization and a joint refinement. We propose this two-stage approach since the eigenvalue-based optimization of the NEC [33] cannot be naively applied to our derived PNEC energy function, which we show in Sec. 4.1. The whole PNEC optimization is given in Alg. 1 and we detail the first stage in Sec. 4.2 & Sec. 4.3, and the second stage in Sec. 4.4.

4.1. Background – Optimizing the NEC

Following [33], the NEC energy function Eq. 3 is rewritten as $E(\boldsymbol{R}, \boldsymbol{t}) = \boldsymbol{t}^\top \boldsymbol{M}(\boldsymbol{R}) \boldsymbol{t}$ using the (symmetric and positive-semidefinite) Gramian matrix

$$\boldsymbol{M}(\boldsymbol{R}) = \sum_i (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)(\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)^\top \,. \tag{6}$$

Because the energy is a quadratic form in the unit vector $\boldsymbol{t}$, the optimization over the translation $\boldsymbol{t}$ can be carried out analytically, i.e.

$$\min_{\substack{\boldsymbol{R} \in \text{SO}(3) \\ \boldsymbol{t}: \, \|\boldsymbol{t}\|=1}} \boldsymbol{t}^\top \boldsymbol{M}(\boldsymbol{R}) \boldsymbol{t} \; = \min_{\boldsymbol{R} \in \text{SO}(3)} \lambda_{\text{min}}(\boldsymbol{M}(\boldsymbol{R})) \,. \tag{7}$$

The eigenvector corresponding to the smallest eigenvalue λ_{min} of $\boldsymbol{M}(\boldsymbol{R})$ minimizes the Rayleigh quotient $\boldsymbol{t}^\top \boldsymbol{M}(\boldsymbol{R}) \boldsymbol{t}$ over all unit-length vectors $\boldsymbol{t}$. The constructed sub-problem is then optimized over the rotation $\boldsymbol{R}$ using the Levenberg-Marquardt algorithm [40, 46], whereas the translation $\boldsymbol{t}$ is obtained by solving an eigenvalue problem.

4.2. Optimizing the PNEC - Translation

The PNEC energy function Eq. 5 is the sum of generalized Rayleigh quotients (GRQs) in the translation $\boldsymbol{t}$, and thus the optimum is not simply given by an eigenvalue as for the NEC. Optimizing the sum of GRQs over the unit sphere has recently been studied in the context of data science and wireless communications [2, 66, 67], and it has been shown by Zhang *et al*. [67] that the self-consistent-field (SCF) algorithm [26] outperforms generic manifold optimization methods.

Since the sum of GRQs can exhibit many local minima [2], and thus the SCF iteration is not guaranteed to converge to a global optimum, we propose a simple, yet effective globalization strategy. To this end, we make use of the intrinsic low dimension of the unit sphere in $\mathbb{R}^3$ by sampling evenly distributed initial points $\boldsymbol{t}_k$ efficiently using the Fibonacci lattice [21]. We then pick the point with the lowest objective function and apply the SCF iteration for N steps. Due to the inherent parallelism, the resulting optimization procedure can be implemented efficiently. We present the effectiveness of the globalization strategy, as well as technical details for the SCF iteration, in the supplementary material.

4.3. Optimizing the PNEC - Rotation

Kneip and Lynen [33] have shown how to optimize Eq. 7 efficiently using the Levenberg-Marquardt algorithm with the rotation parametrized based on the Cayley transformation [7]. To account for the weights in the PNEC energy function Eq. 5, we employ an optimization scheme similar to the popular iteratively reweighted least squares (IRLS) algorithm [37]. Specifically, given a previous estimate of the rotation and translation $(\boldsymbol{R}_p, \boldsymbol{t}_p)$, we compute fixed weights $\tilde{\sigma}_i = \sigma_i(\boldsymbol{R}_p, \boldsymbol{t}_p)$ for all i, and define the weighted matrix

$$\boldsymbol{M}_P(\boldsymbol{R}; \{\tilde{\sigma}_i\}_i) = \sum_i \frac{(\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)(\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)^\top}{\tilde{\sigma}_i^2} \tag{8}$$

that depends only on the rotation $\boldsymbol{R}$. The rotation is obtained by finding $\boldsymbol{R}$ such that the smallest eigenvalue of $\boldsymbol{M}_P(\boldsymbol{R}; \{\tilde{\sigma}_i\}_i)$ is minimal. After doing so based on the optimizer of Kneip and Lynen [33], the weights $\{\tilde{\sigma}_i\}_i$ are updated with new $\boldsymbol{R}, \boldsymbol{t}$.

4.4. Optimizing the PNEC - Joint Refinement

After the first stage we improve the result using joint refinement. Specifically, we use a least-squares optimization strategy, which is effective for finding a local optimum of the energy function given a good starting point [52]. For the PNEC we optimize over

$$E_P(\boldsymbol{R}, \boldsymbol{t}) = \sum_i \left(\frac{\boldsymbol{t}^\top (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)}{\sqrt{\boldsymbol{t}^\top \hat{\boldsymbol{f}}_i \boldsymbol{R}\boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i{}^\top \boldsymbol{t}}} \right)^2, \tag{9}$$

the least-squares formulation of the constraint. The Levenberg-Marquardt algorithm optimizes the objective function in the rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ simultaneously and uses the solution of the first stage as the starting value. Because $\boldsymbol{R}$ is a rotation matrix, we use manifold optimization [27] to optimize over the special orthogonal group $\mathrm{SO}(3)$. For the translation $\boldsymbol{t}$ we use spherical coordinates with the radius fixed to one in order to ensure that $\|\boldsymbol{t}\| = 1$ holds. We would like to highlight that this joint refinement is different from bundle adjustment. Most notably, it does not need to calculate the 3D position of the features.

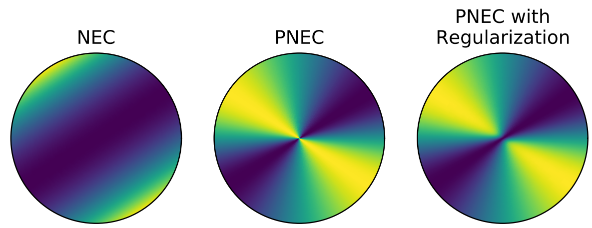


Figure 4. Visualization of the NEC, the PNEC, and the PNEC with the regularization proposed in Sec. 4.5. The plot shows $\boldsymbol{t}$ in a neighborhood of $\boldsymbol{f}$ (in polar coordinates), where the center of the circle corresponds to $\boldsymbol{t} = \boldsymbol{f}$. For $\boldsymbol{t} = \boldsymbol{f}$, the PNEC shows a finite discontinuity, for which the limit depends on the direction. Our regularization eliminates this singularity while also maintaining the overall shape of the energy function.

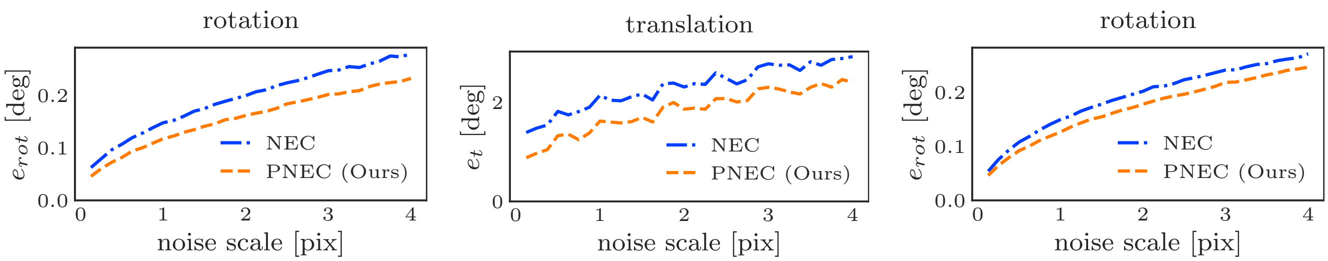
4.5. Singularities of the PNEC

The PNEC energy function Eq. 5 has a singularity if the translation $\boldsymbol{t}$ is parallel to a bearing vector $\boldsymbol{f}_i$ because the variance σ_i^2 vanishes due to $\hat{\boldsymbol{f}}_i \boldsymbol{t} = \boldsymbol{f}_i \times \boldsymbol{t} = \boldsymbol{0}$. On the other hand, the numerator involves the same term and thus the energy function is bounded and possesses a finite discontinuity, as illustrated in Fig. 4. In the supplementary material we present the derivation of the directional limit of the energy function.

While the discontinuity is finite and less problematic than an infinite discontinuity, it still poses challenges. First, in contrast to the function values, the derivatives of the energy function are not bounded, which is problematic for the joint refinement. Second, the matrix $\boldsymbol{M}_P$, unlike the energy function, includes $\hat{\boldsymbol{f}}_i \boldsymbol{t}$ only in its denominator not the numerator. Hence $\boldsymbol{M}_P$ tends to infinity for $\boldsymbol{t} \to \boldsymbol{f}_i$. To address these issues, we consider a variance of the form $\sigma_i'^2 = \sigma_i^2 + c$ with regularization constant $c > 0$. Fig. 4 shows the effect that the regularizer has on the energy function.

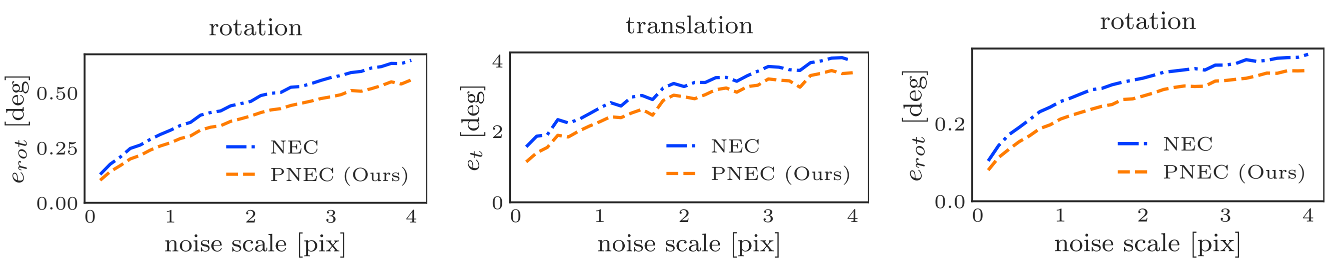
5. Evaluation

We evaluate the performance of the PNEC and compare it to the original NEC on simulated data as well as in a visual odometry setting on real world data. On the simulated data, the proposed PNEC achieves better results than the



(a) Experiment w/ translation: e_{rot} (b) Experiment w/ translation: e_{t} (c) Experiment w/o translation: e_{rot}

Figure 5. Experiments for omnidirectional cameras. Results are averaged over 10 000 random instantiations for anisotropic inhomogeneous noise over different noise intensities. Our PNEC consistently leads to smaller errors compared to the NEC [33] for all noise levels. This holds for rotation and translation estimates in the general case in Fig. 5a and Fig. 5b, respectively, as well as the rotation in the zero-translation case in Fig. 5c.



(a) Experiment w/ translation: e_{rot} (b) Experiment w/ translation: e_{t} (c) Experiment w/o translation: e_{rot}

Figure 6. Experiments for pinhole cameras. Results are averaged over 10 000 random instantiations for anisotropic inhomogeneous noise over different noise intensities. As for an omnidirectional camera in Fig. 5, our PNEC consistently leads to smaller errors compared to the NEC [33]. For a pinhole camera the average errors are higher for both methods in comparison to an omnidirectional camera. The experiments show that our PNEC is viable for the two most common camera types.

NEC and several other popular relative pose estimation algorithms. On KITTI we compare our approach to the MRO algorithm [9] that uses the NEC for rotation estimation. For evaluating the PNEC we replace the ORB features in MRO with Kanade-Lucas-Tomasi (KLT) tracks [45,58], which allow for uncertainty extraction as discussed in Sec. 3.2, and the NEC with the PNEC.

In the supplementary material, a more detailed analysis including translational errors shows that compared to the NEC the PNEC is not only significantly more accurate on average, but also more consistent. An ablation study on KITTI demonstrates that all stages of our optimization scheme are essential for best results. We furthermore detail all hyperparameter choice and detail the experimental protocol for reproducibility.

5.1. Frame-to-Frame Simulation

With the simulated experiments we evaluate the performance of the PNEC in a frame-to-frame setting. The experiments consist of randomly generated problems of two frames with known correspondences. We use

$$e_{\text{rot}} := \angle(\boldsymbol{R}^\top \tilde{\boldsymbol{R}}), \text{ and} \tag{10}$$

$$e_{\text{t}} := \arccos(\boldsymbol{t}^\top \tilde{\boldsymbol{t}}) \tag{11}$$

as error metrics between the ground truth $\boldsymbol{R}, \boldsymbol{t}$ and the estimated values $\tilde{\boldsymbol{R}}, \tilde{\boldsymbol{t}}$, where $\angle(\cdot)$ returns the angle of the rotation matrix.

Omnidirectional Camera. In this experiment we follow the experimental outline proposed by Kneip and Lynen [33] closely. We differ from the original experiments in the following ways: we only add noise to the points in the second frame; to compensate for the lack of noise in the first frame, we scale the standard deviation by a factor of 2; we recreate the experiment with different noise types based on the classification by Brooks *et al.* [5] and generate individual covariance matrices for each point. A detailed description of how the matrices are generated can be found in the supplementary material. To show the effectiveness of the PNEC still holds even for pure rotation, we repeat the experiment with the translational difference fixed to zero.

Fig. 5 shows the results for anisotropic inhomogeneous noise for both experiments. The PNEC achieves consistently better results for the rotation over all noise levels in both experiments.

Pinhole Camera. Since most cameras are modeled as pinhole cameras we also repeat the previous experiments for pinhole cameras. The generation of the frames stays the same. Points are sampled in viewing direction of the coordi-

	Omnidirectional									Pinhole									
	w/ T						w/o T			w/ T						w/o T			
Noise level [px]	0.5		1.0		1.5		0.5	1.0	1.5	0.5		1.0		1.5		0.5	1.0	1.5	
Metric [degree]	e_{rot}	e_{t}	e_{rot}	e_{t}	e_{rot}	e_{t}	e_{rot}	e_{rot}	e_{rot}	e_{rot}	e_{t}	e_{rot}	e_{t}	e_{rot}	e_{t}	e_{rot}	e_{rot}	e_{rot}	
7pt [25]/8pt [24]	0.19	<u>1.76</u>	0.26	2.36	0.33	2.64	<u>0.10</u>	<u>0.15</u>	0.20	0.62	4.64	0.89	6.09	1.07	6.74	<u>0.17</u>	<u>0.23</u>	<u>0.28</u>	+95%
Stewenius 5pt [56]	0.23	2.34	0.30	2.97	0.37	3.38	0.18	0.30	0.33	0.61	3.14	0.71	4.04	0.89	4.57	0.46	0.64	0.77	+137%
Nistér 5pt [51]	1.61	6.82	1.64	7.55	1.92	8.42	0.29	0.39	0.42	3.26	8.72	3.46	9.80	3.76	10.39	0.51	0.67	0.83	+643%
NEC [33]	<u>0.11</u>	1.90	<u>0.15</u>	<u>2.10</u>	<u>0.17</u>	<u>2.11</u>	0.11	<u>0.15</u>	<u>0.18</u>	<u>0.25</u>	<u>2.41</u>	<u>0.34</u>	<u>2.78</u>	<u>0.41</u>	<u>2.91</u>	0.19	0.25	0.29	+24%
PNEC (Ours)	**0.08**	**1.29**	**0.12**	**1.60**	**0.14**	**1.66**	**0.09**	**0.13**	**0.15**	**0.20**	**2.06**	**0.28**	**2.38**	**0.34**	**2.54**	**0.15**	**0.21**	**0.25**	

Table 1. Rotation and translation error for different algorithms. Results for omnidirectional and pinhole cameras for experiments with and without translation over different noise levels for anisotropic and inhomogenous noise. Errors are averaged over 10 000 random problems each with 10 points. For experiments without translation (W/O T) only e_{rot} is reported, due to e_{t} not being defined for zero translation. For all other algorithms apart from our PNEC we use the implementations from OpenGV [32]. (7pt) falls back to (8pt) for the non-minimal number of 10 correspondences. Our PNEC consistently achieves the best results, outperforming the NEC and several popular relative pose estimation algorithms. The last column gives the average error increase compared to the PNEC.

nate system of the first frame. The points are projected into the world coordinate system and then into the two frames using a pinhole camera model. The noise offset is added in the image plane. As with the omnidirectional camera experiment, we repeat this experiment for pure rotation.

Fig. 6 shows the results for the pinhole camera experiment for anisotropic inhomogeneous noise. While the overall error for both methods is slightly higher for pinhole cameras than for omnidirectional cameras, the PNEC still outperforms the NEC consistently.

Tab. 1 gives a quantitative comparison of our PNEC to other relative pose estimation algorithms from the literature on the experiments presented in Fig. 5 and Fig. 6. Our PNEC consistently achieves the best result for both camera models for experiments with and without translation.

Additional experiments on other noise types show that our PNEC outperforms the NEC even in cases of isotropic and homogeneous noise. Although the covariance matrices are identical, the variance for each residual is different due to the geometry of the problem. Furthermore, our PNEC is robust against wrongly estimated noise parameters. We present the results to these experiments in the supplementary material.

5.2. Visual Odometry

Besides the simulated experiments, we also validate the PNEC on real world data, namely the highly popular KITTI odometry dataset [20]. We compare our results with the MRO algorithm by Chng *et al.* [9] that uses the optimization from [33]. For MRO and our algorithm we disable rotation averaging and loop closure to focus on local rotation estimation. Our approach differs from MRO in two ways.

First, we use the KLT-based tracking implementation also used in [61] to extract feature keypoints instead of ORB features. Second, we replace the NEC with our PNEC for relative rotation estimation. To capture the effects of both changes, we compare the rotation estimation of MRO, as reported in [9], KLT-NEC, using KLT tracks and the NEC, and KLT-PNEC, the proposed PNEC with KLT tracks. Both KLT-NEC and KLT-PNEC use the same KLT tracks for the relative rotation estimation.

The proposed PNEC can account for uncertainties in the feature correspondence positions that approximately follow a Gaussian distribution. To overcome outlier correspondences from failed KLT tracks, we use the same RANSAC [18] routine as the NEC for estimating the rotation in the first loop of Alg. 1.

Fig. 7 shows a trajectory generated from the rotation estimates of MRO and our approach. In Tab. 2 we compare the mean performance over 5 runs of all approaches in the rotation-only version of the *Relative Pose Error* (RPE) for n camera poses as defined in [9]. The RPE evaluates the root mean square error (RMSE) of rotational residuals over frame pairs. The residual for a "time-step" Δ is

$$E_i := \angle((\boldsymbol{R}_i^\top \boldsymbol{R}_{i+\Delta})^\top (\tilde{\boldsymbol{R}}_i^\top \tilde{\boldsymbol{R}}_{i+\Delta})). \tag{12}$$

The RMSE is calculated over $m := n - \Delta$ residuals

$$\text{RMSE}(\Delta) := \left(\frac{1}{m} \sum_{i=1}^{m} E_i^2 \right)^{\frac{1}{2}}. \tag{13}$$

For our evaluation we use

$$\text{RPE}_1 := \text{RMSE}(1), \text{ and} \tag{14}$$

$$\text{RPE}_n := \frac{1}{n} \sum_{\Delta=1}^{n} \text{RMSE}(\Delta), \tag{15}$$

to capture local frame-to-frame rotation error and long term drift, respectively.

The results show the following: With a single exception (seq. 01), using KLT tracks instead of ORB features is beneficial for relative rotation estimation with the NEC. PNEC

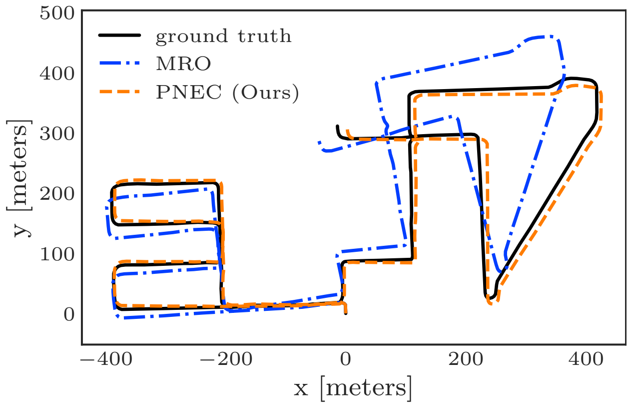


Figure 7. Qualitative trajectory comparison for KITTI seq. 08. The trajectory was generated with the estimated rotations of MRO [9] and PNEC, respectively, and are combined with the ground truth translations for visualization purposes. Relative rotations computed with the proposed PNEC lead to a significantly reduced drift.

	MRO [9]		KLT-NEC		KLT-PNEC (Ours)	
Seq.	RPE_1	RPE_n	RPE_1	RPE_n	RPE_1	RPE_n
00	0.360	8.670	0.125	5.922	**0.119**	**3.429**
01*	**0.290**	**16.030**	0.695	27.406	0.782	23.500
02	0.290	16.030	**0.093**	**6.693**	0.122	9.687
03	0.280	5.470	0.073	2.728	**0.059**	**1.411**
04	0.040	1.080	0.041	0.619	**0.038**	**0.463**
05	0.250	11.360	0.079	4.489	**0.070**	**3.203**
06	0.180	4.720	0.073	3.162	**0.042**	**2.322**
07	0.280	7.490	0.105	4.640	**0.074**	**2.065**
08	0.270	9.210	0.070	5.523	**0.060**	**3.347**
09	0.280	9.850	0.088	3.533	**0.080**	**3.514**
10	0.380	13.250	0.073	**3.959**	**0.072**	4.094

Table 2. Quantitative comparison for KITTI. The significant gap between MRO and KLT-NEC confirms the benefit of using KLT tracks. The difference between KLT-NEC and KLT-PNEC shows the effectiveness of our PNEC compared to the NEC. *In seq. 01 the KLT implementation of [61] fails and produces many wrong tracks with incorrect covariances due to self-similar structure.

outperforms NEC on 8 out of 11 sequences in both metrics, often significantly. Excluding seq. 01, the PNEC on average improves the RPE_1 for frame-to-frame rotational error by 10% and the RPE_n for long term drift by 19%.

5.3. Runtime

Tab. 3 shows the average frame processing time on the KITTI dataset for MRO, KLT-NEC and KLT-PNEC. The experiments were performed on a laptop with a 2.4 GHz Quad-Core Intel Core i5 processor and 8 GB of memory. For MRO we use the same configuration as for their demo. The results show the runtime advantage of KLT tracks that do not need feature matching like ORB features. While the proposed optimization scheme for the PNEC is slightly slower than the simpler NEC optimization algorithm, the odometry with KLT-PNEC runs in real-time on KITTI.

	MRO [9]	KLT-NEC	KLT-PNEC
feature creation	36	30	30
matching	120		
optimization	5	23	47
total time	161	53	77

Table 3. Average frame processing time in milliseconds. For MRO, most of the time is needed for matching. KLT-NEC and KLT-PNEC (Ours) achieve real-time performance on KITTI.

6. Discussion and Future Work

While the proposed optimization scheme effectively optimizes the PNEC energy function, it relies on two consecutive stages, and is thus more involved than the optimization scheme proposed for the NEC [33]. Further and more detailed limitations of the proposed approach are given in the supplementary material. Nevertheless, we have shown in Sec. 5.3 that the proposed algorithm is real-time capable. As we explain in Sec. 4.2, the optimization over the translation alone is an actively studied problem for which no simple solution is known. Nevertheless, investigating improved optimization schemes for our PNEC energy function is a promising direction for future work. Recent works have shown that deep learning can boost the performance of visual odometry algorithms [22,62,63]. However, the focus of our work is on the correct modelling of the uncertainty for relative pose estimation, similar to [33, 34]. As such, while in our work we do not consider deep learning, we believe that the integration of our ideas into learning systems may be an interesting direction for follow-up works. For example, the 2D feature position covariance matrices could be predicted by a neural network.

7. Conclusion

This paper shows how to utilise 2D feature position uncertainties to obtain more accurate relative pose estimates from a pair of images. To this end, we introduce the probabilistic normal epipolar constraint (PNEC), and we propose an effective optimization scheme that runs in real-time. In synthetic experiments, the PNEC gives more accurate rotation estimates than the NEC and several popular relative rotation estimation algorithms for different noise levels and for the pure-rotation case. The results on KITTI show, that the relative rotation estimation of the PNEC improves upon the NEC-based MRO, a state-of-the-art rotation-only VO system, and can be used *e.g.* for global initialization in SfM.

Acknowledgements

We express our appreciation to our colleagues and the WARP student team who have supported us. We thank Laurent Kneip for providing us with the suppl. for [33]. This work was supported by the ERC Advanced Grant SIMULACRON and by the Munich Center for Machine Learning.

References

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *ECCV*, 2006. 2

[2] Aohud Abdulrahman Binbuhaer. *On Optimizing the Sum of Rayleigh Quotients on the Unit Sphere*. PhD thesis, The University of Texas at Arlington, 2019. 5

[3] Christopher M. Bishop. *Pattern recognition and machine learning, 5th Edition*. Springer, 2007. 4

[4] Jesus Briales, Laurent Kneip, and Javier Gonzalez-Jimenez. A certifiably globally optimal solution to the non-minimal relative pose problem. In *CVPR*, 2018. 2

[5] M.J. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel. What value covariance information in estimating vision parameters? In *ICCV*, 2001. 2, 6

[6] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, José Neira, Ian Reid, and John J Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 32, 2016. 2

[7] Arthur Cayley. About the algebraic structure of the orthogonal group and the other classical groups in a field of characteristic zero or a prime characteristic. *Reine Angewandte Mathematik*, 32, 1846. 5

[8] Mahalanobis Prasanta Chandra et al. On the generalised distance in statistics. In *Proceedings of the National Institute of Sciences of India*, 1936. 4

[9] Chee-Kheng Chng, Álvaro Parra, Tat-Jun Chin, and Yasir Latif. Monocular rotational odometry with incremental rotation averaging and loop closure. *Digital Image Computing: Techniques and Applications (DICTA)*, 2020. 2, 3, 6, 7, 8

[10] Alejo Concha, Michael Burri, Jesús Briales, Christian Forster, and Luc Oth. Instant visual odometry initialization for mobile ar. *IEEE Transactions on Visualization and Computer Graphics*, 27, 2021. 3

[11] Igor Cvišić, Josip Ćesić, Ivan Marković, and Ivan Petrović. Soft-slam: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. *Journal of field robotics*, 2018. 3

[12] Igor Cvišić, Ivan Marković, and Ivan Petrović. Recalibrating the KITTI Dataset Camera Setup for Improved Odometry Accuracy. In *European Conference on Mobile Robots (ECMR)*, 2021. 3

[13] Leyza Baldo Dorini and Siome Klein Goldenstein. Unscented feature tracking. *Computer Vision and Image Understanding*, 115, 2011. 2

[14] J. Engel, V. Koltun, and D. Cremers. Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 2, 3

[15] J. Engel, T. Schöps, and D. Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *ECCV*, 2014. 2, 3

[16] Kaveh Fathian, J Pablo Ramirez-Paredes, Emily A Doucette, J Willard Curtis, and Nicholas R Gans. Quest: A quaternion-based approach for camera motion estimation from minimal feature points. *IEEE Robotics and Automation Letters (RAL)*, 3, 2018. 2

[17] O.D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. In *Workshop on Visual Motion*, 1989. 1

[18] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24, 1981. 7

[19] Wolfgang Förstner and Eberhard Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *ISPRS intercommission conference on fast processing of photogrammetric data*, 1987. 2

[20] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *CVPR*, 2012. 1, 7

[21] Álvaro González. Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, 42, 2010. 5

[22] W. Nicholas Greene and Nicholas Roy. Metrically-scaled monocular SLAM using learned scale factors. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2020. 8

[23] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International journal of computer vision*, 103(3):267–305, 2013. 1

[24] Richard I Hartley. In defense of the eight-point algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 19, 1997. 1, 2, 7

[25] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 1, 2, 7

[26] Douglas R Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*, 1928. 5

[27] Christoph Hertzberg, René Wagner, Udo Frese, and Lutz Schröder. Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds. *Inf. Fusion*, 14, 2013. 5

[28] Kenichi Kanatani. For geometric inference from images, what kind of statistical model is necessary? *Systems and Computers in Japan*, 35, 2004. 2

[29] Kenichi Kanatani. Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis. *IJCV*, 80, 2008. 2

[30] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *ICCV*, 2001. 2

[31] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007. 3

[32] Laurent Kneip and Paul Furgale. Opengv: A unified and generalized approach to real-time calibrated geometric vision. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2014. 7

[33] Laurent Kneip and Simon Lynen. Direct optimization of frame-to-frame rotation. In *ICCV*, 2013. 1, 2, 4, 5, 6, 7, 8

[34] Laurent Kneip, Roland Siegwart, and Marc Pollefeys. Finding the exact rotation between two images independently of the translation. In *ECCV*, 2012. 1, 2, 3, 4, 8

[35] Erwin Kruppa. *Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung*. Hölder, 1913. 2

[36] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *BMVC*, 2008. 2

[37] Charles Lawrence Lawson. *Contribution to the theory of linear least maximum approximation*. PhD thesis, University of California, 1961. 5

[38] S. H. Lee and Javier Civera. Geometric interpretations of the normalized epipolar error. *ArXiv*, abs/2008.01254, 2020. 4

[39] Seong Hun Lee and Javier Civera. Rotation-only bundle adjustment. In *CVPR*, 2021. 2, 3

[40] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2, 1944. 4

[41] Hongdong Li and Richard Hartley. Five-point motion estimation made easy. In *IEEE International Conference on Pattern Recognition (ICPR)*, 2006. 2

[42] John Lim, Nick Barnes, and Hongdong Li. Estimating relative camera motion from the antipodal-epipolar constraint. *IEEE TPAMI*, 32, 2010. 2

[43] HC Longuet-Higgins. Readings in computer vision: issues, problems, principles, and paradigms. *A computer algorithm for reconstructing a scene from two projections*, 1987. 1, 2

[44] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60, 2004. 2

[45] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 1981. 2, 6

[46] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11, 1963. 4

[47] Jochen Meidow, Christian Beder, and Wolfgang Förstner. Reasoning with uncertain points, straight lines, and straight line segments in 2d. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 2009. 2

[48] Pierre Moulon, Pascal Monasse, and Renaud Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3248–3255, 2013. 1

[49] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós. Orb-slam: A versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 31, 2015. 3

[50] R. Mur-Artal and J. D. Tardós. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33, 2017. 3

[51] D. Nister. An efficient solution to the five-point relative pose problem. In *CVPR*, 2003. 2, 7

[52] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 1999. 5

[53] Sameer Sheorey, Shalini Keshavamurthy, Huili Yu, Hieu Nguyen, and Clark N Taylor. Uncertainty estimation for klt tracking. In *Asian Conference on Computer Vision*, 2014. 2

[54] R.M. Steele and C. Jaynes. Feature uncertainty arising from covariant image noise. In *CVPR*, 2005. 2

[55] Henrik Stewenius, Christopher Engels, and David Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60, 2006. 2

[56] Henrik Stewénius, David Nistér, Magnus Oskarsson, and Kalle Åström. Solutions to minimal generalized relative pose problems. *Workshop on omnidirectional vision*, 2005. 2, 7

[57] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010. 2

[58] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. *IJCV*, 9, 1991. 2, 6

[59] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, 1999. 2

[60] Jeffrey K Uhlmann. *Dynamic map building and localization: New theoretical foundations*. PhD thesis, University of Oxford Oxford, 1995. 2, 4

[61] Vladyslav Usenko, Nikolaus Demmel, David Schubert, Jörg Stückler, and Daniel Cremers. Visual-inertial mapping with non-linear factor recovery. *IEEE Robotics and Automation Letters (RAL)*, 5, 2020. 3, 7, 8

[62] Nan Yang, Lukas von Stumberg, Rui Wang, and Daniel Cremers. D3VO: deep depth, deep pose and deep uncertainty for monocular visual odometry. In *CVPR*, 2020. 8

[63] Nan Yang, Rui Wang, Jörg Stückler, and Daniel Cremers. Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry. In *ECCV*, 2018. 8

[64] Bernhard Zeisl, Pierre Georgel, Florian Schweiger, Eckehard Steinbach, and Nassir Navab. Estimation of location uncertainty for scale invariant feature points. In *BMVC*, 2009. 2

[65] Hongmou Zhang, Denis Grießbach, Jürgen Wohlfeil, and Anko Börner. Uncertainty model for template feature matching. In *Pacific-Rim Symposium on Image and Video Technology*, pages 406–420. Springer, 2017. 2

[66] Lei-Hong Zhang. On optimizing the sum of the rayleigh quotient and the generalized rayleigh quotient on the unit sphere. *Computational Optimization and Applications*, 54, 2013. 5

[67] Lei-Hong Zhang and Rui Chang. A nonlinear eigenvalue problem associated with the sum-of-rayleigh-quotients maximization. *Transactions on Applied Mathematics*, 2, 2021. 5