DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Robotics, Cognition, Intelligence

Frame-To-Frame Rotation Estimation under Uncertain Feature Positions for Visual Odometry

Dominik Muhle

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Robotics, Cognition, Intelligence

Frame-To-Frame Rotation Estimation under Uncertain Feature Positions for Visual Odometry

Bild-zu-Bild Rotationsschätzung unter Unsicheren Featurepositionen für Visuelle Odometrie

Author:	Dominik Muhle
Supervisor:	Prof. Dr. Daniel Cremers
Advisor:	Lukas Koestler
	Nikolaus Demmel
Submission Date:	15.09.2021

I confirm that this master's thesis in robotics, cognition, intelligence is my own work and I have documented all sources and material used.

Munich, 15.09.2021 Dominik Muhle

Acknowledgments

I want to sincerely thank my supervisor Prof. Dr. Daniel Cremers, for his advice and for offering this thesis. Furthermore, I want to thank Prof. Dr. Florian Bernard for his valuable support. I want to thank Lukas Köstler and Nikolaus Demmel for their regular and continuous support throughout this thesis and for giving me crucial advice.

I want to thank my family, whose support and encouragement motivated me throughout my studies. I am deeply grateful to my partner for being there for me and their motivation and encouragement. I am much obliged to the Technical University of Munich for the opportunities and the brilliant courses offered to me.

Abstract

Relative pose estimation of two camera views is a fundamental problem in computer vision. While many algorithms to solve this problem have been proposed, almost all struggle with purely rotational motion given no additional information. Kneip *et al*. proposed the normal epipolar constraint (NEC) that allows for rotation estimation independent of the translation. However, their approach is highly dependent on accurate feature positions in both frames. This work presents the novel probabilistic normal epipolar constraint (PNEC) for relative pose estimation. The PNEC overcomes the NEC limitation by accounting for anisotropic and inhomogeneous uncertainties in the feature positions in the target frame. To this end, a novel objective function, along with an efficient optimization scheme, is derived that effectively estimates the rotation while maintaining real-time performance. Experiments on synthetic data demonstrate that the novel PNEC improves upon the original NEC and yields more accurate rotation estimates. Furthermore, this thesis shows the integration of the PNEC into a state-of-the-art monocular rotation-only odometry system. This new PNEC visual odometry consistently achieves improved results for the real-world KITTI dataset.

Contents

1. Introduction

Motivation Extracting 3D geometry from multiple images is a widely researched topic in computer vision. Its numerous applications include mobile robots, autonomous driving, and augmented or virtual reality. Essential for the extraction of 3D geometry is the identification of the camera pose of an image. This knowledge is often obtained by relative pose estimation between two images of the same scene. Relative pose estimation is the building block of many geometric vision algorithms like visual odometry (VO) or structure from motion (SfM). Estimating the relative rotation is especially important for VO systems since small errors quickly lead to a drift in estimation.

Most approaches to VO systems either use the essential or fundamental matrix [42, 50, 40] or maintain a 3D representation of the surrounding structure for relative pose estimation. Both approaches have deteriorating performance for purely rotational motion in common and require additional techniques like model selection or additional inertial information [32]. The normal epipolar constraint [33] for estimating rotation independent of the translation does not suffer for purely rotational motion. However, the normal epipolar constraint, like many other relative pose estimation algorithms, does not consider the quality of the features used. Apart from outliers in the feature correspondences that are removed, every match contributes equally to the final result. However, the two-dimensional error distribution of a feature correspondence is dependent on the image region and the method used for extraction. Therefore, each feature correspondence exhibits a different error distribution, which leads to unequal contributions to the problem. A feature located on an edge is not very accurately localized parallel to it but possesses a high accuracy perpendicular to the edge. An equal weighting of the feature correspondences loses this anisotropic information of the feature position. The benefit of incorporating uncertainty information into relative pose estimation has been shown for fundamental matrix estimation [6].

Problem Statement and Contributions The goal of this work is to incorporate uncertainty information about the feature position into the normal epipolar constraint (NEC) to achieve accurate rotation estimates even in cases of pure rotation. From the probabilistic description of a feature position in the image, we derive a novel energy function that accounts for the quality in the feature matches based on uncertainty information by propagating it to energy function residual. This new energy function is a weighted

version of NEC energy function. We show that optimizing the energy function of the PNEC is not a trivial task, and we break it down into two sub-problems for which efficient yet not optimal solutions are known. This thesis presents an efficient optimization scheme to minimize the energy function based on the eigenvalue-based solver for the NEC, the self-consistent-field (SCF) method for the sum of generalized Rayleigh quotients (GRQs), and a least-squares refinement that achieves real-time performance. We present further analysis of the geometry of the PNEC and its energy function. We focus on singularities that arise from the PNEC formulation. A regularization is proposed that removes these singularities from the PNEC energy function.

We evaluate our method on simulated data and the popular KITTI odometry dataset. We show that together with uncertainty extracted from Kanade-Lucas-Tomasi tracks the PNEC outperforms a NEC state-of-the-art rotation-only odometry system.

Outline Ch. 2 gives an overview over the related work in the field. We summarize popular techniques for relative pose estimation in Sec. 2.1, two approaches to feature extraction in Sec. 2.2, the usage of positional uncertainty in geometric vision in Sec. 2.3, and give an overview of visual odometry systems in Sec. 2.4. Furthermore, we introduce the work related to the sum of GRQs relevant to this work. In the method section Ch. 3 we first summarize the NEC (see Sec. 3.1) and from there derive the PNEC and its energy function in Sec. 3.2. Sec. 3.3 shows the optimization scheme for the PNEC used throughout this work, while Sec. 3.4 presents additional insight into the energy function from which we derive an effective regularization scheme for the PNEC used in practical applications. Experiments validating the effectiveness of the PNEC compared to the NEC are given in Ch. 4. Sec. 4.1 investigates the frame-to-frame rotation estimation of the PNEC with experiments on synthetic data. Sec. 4.2 evaluates the performance for real-world data in an odometry setting. We discuss the performance of the PNEC in Ch. 5 and present potential future work based on the experimental results. This thesis concludes with a summarization of the work presented (see Ch. 6).

2. Related Work

The focus of this thesis is the integration of uncertainty in the feature positions into the NEC. We aim to improve frame-to-frame rotation estimation for application in visual odometry. Therefore, the discussion of the related work is limited to the following aspects: giving an overview of the different methods for *relative pose estimation*; the usage of *uncertainty information* for relative pose estimation; the difference between *matching-based* and *tracking-based* approaches to correspondences extraction; an overview of different *visual odometry systems* with a focus on rotation-only algorithms. For an additional, broader overview, we refer the reader to the books by Szeliski [67] and by Hartley and Zisserman [24] as well as to the overview papers for bundle adjustment [69] and simultaneous localization and mapping [7], which is closely related to odometry.

2.1. Relative Pose Estimations

Relative pose estimation is the task of estimating the rotation and translation between two viewpoints. It is a long-standing problem in the field of computer vision, with the first known solution proposed in 1913 by Kruppa [34]. Most methods proposed to solve this task fall into two categories, *feature-based* [31, 47, 48, 71] or *direct* [14, 13]. Feature-based methods rely on previously computed feature correspondences extracted from the images. Direct methods use the intensity differences between the two images.

While direct methods have gained lots of popularity in the last years, they rely on *photometric consistency*. Therefore, they are sensitive to lighting and illumination changes in the scene [65]. In general problems, *e.g.* structure from motion or long-term relocalization, such appearance changes (lighting and weather) frequently violate the photometric consistency assumption [65]. Feature-based methods are considerably more robust to these effects. This work extends a feature-based rotation estimation method, and therefore this section will focus on these methods in the following.

The feature-based methods proposed over the years differ in the geometric constraints used and the minimum number of points needed. Many methods [42, 50, 62, 40, 35] base their solution on the essential matrix in the case of a calibrated camera, or the fundamental matrix in the general case. Both matrices link the position of the feature correspondence pair in the epipolar geometry.

One of the earliest works to use the essential matrix is the eight-point-algorithm by Longuet-Higgins [42] that requires at least eight correspondence pairs and results in a linear solver. While the eight-point-algorithm has been criticized for its sensitivity to noise, the defense by Hartley [25] shows its good performance given careful normalization. However, this algorithm is not suitable in certain configurations, *e.g.* purely rotational motion of the camera or a coplanarity of the points used [32].

The latter scenario can be addressed by algorithms using the minimal required number of five correspondences. Nistér [50] proposed a solution using this minimal number by utilizing polynomials and root bracketing. A recent approach [15] with the same number of points uses quaternions for directly estimating the relative pose resulting in more noise resilience. However, neither of the essential matrix-based algorithms allow for relative pose estimation in the presence of noise if the translation between the two viewpoints is near to or equal to zero [32]. A common problem of these approaches is that given a small translation the essential matrix tends to zero.

A popular way to address this shortcoming, from which not only essential matrix-based methods suffer, is to use sensor fusion to incorporate inertial information into the relative pose estimation leading to visual-inertial odometry systems [38, 49, 64]. Recent works have also proposed different geometric constraints to address these problems without inertial information [41, 33].

The *antipodal-epipolar constraint* by Lim *et al.* [41] makes use of antipodal rays, decoupling the translational and rotational motion of the camera. Each component can then be solved in a lower dimension than the original. Antipodal rays require cameras with a field-of-view of over 180 degrees, and therefore the constraint cannot be used with a single pinhole camera.

As the *antipodal-epipolar constraint*, the normal epipolar constraint (NEC) by Kneip *et al.* [33] also decouples the rotational motion of the camera from the translational motion, resulting in a constraint that allows the estimation of the rotation independent of the translation given at least five correspondence pairs. It was first introduced by Kneip *et al.* [33] as the *epipolar plane normal coplanarity constraint* and later renamed to the normal epipolar constraint [10], which we will also use throughout this work. In their first paper, Kneip *et al.* [33] proposed a Gröbner basis solver for a system of polynomial equations leading to a complex optimization of the NEC for exactly five points. However, due to the complexity of this solver and its numerical instabilities for purely rotational motion, Kneip and Lynen [32] presented an eigenvalue-based solver that expands the use of the NEC to more than five points. While the NEC is recapped in more detail in Sec. 3.1 we refer the interested reader to [33] for a more detailed derivation of the constraint, to [32] for an elegant optimization scheme of the eigenvalue-based solver, and to [36] for an in-depth investigation into the geometry of the NEC.

2.2. Feature Correspondence Generation

Feature-based VO systems like PTAM [31] and ORB-SLAM [47, 48] need feature correspondence pairs between images. This section presents different methods to generate feature correspondences between a pair of images. It focuses on *feature extraction and matching* and *feature tracking* approaches to emphasize their differences. Both, *feature extraction and matching* [47, 48] and *feature tracking* [71] methods find their use in VO systems.

Feature extraction and matching is one of the most commonly used approaches to generate feature correspondences [66, 75]. It is a two-step method of first finding and describing keypoints in the image and then matching the extracted keypoints of two images based on their description to find correspondences. While this core mechanic stays the same, a variety of algorithms have been proposed that differ in the kind of features used and the method of matching them. Some of the most popular keypoint extraction algorithms are SIFT [43], SURF [3], and ORB [56]. A more detailed overview of the different algorithms and techniques can be found in the book by Nixon *et al*. [52].

Unlike feature extraction and matching, *feature tracking* does not need to extract keypoint in both images. A widely used example of a feature tracker is the Kanade-Lucas-Tomasi (KLT) tracker [44, 68, 58]. Features extracted in one image are tracked in another image using a directed search by optimizing a cost function. Over the years, different approaches to the optimized cost function have been made, resulting in a wide variety of formulations. The KLT tracker of [71] also used in this work makes use of the *inverse compositional formulation* [2] resulting in efficient computation of the Hessian needed for optimization. An overview of the most popular formulations can be found in [1] with additional information in [44, 68] and a multi-paper series starting with the excellent paper by Baker and Matthews [2].

2.3. Uncertainty for Feature Correspondences

The accurate determination of feature positions in images is difficult. Due to image noise, feature positions in the image cannot be localized with perfect accuracy. Estimating this positional uncertainty has been an extensively researched topic in computer vision [17, 61, 57, 77]. Different methods have been proposed that incorporate the aspects of the different feature extractors, that include the Förstner corner detector [61] and the Harris corner detector [6, 54], to obtain better results. Zeisl *et al*. [76] have presented a method for uncertainty estimation for the SIFT [43] and SURF [3] detectors with an emphasis on anisotropic and inhomogeneous covariances information, so that not every feature follows the same error distribution. The importance of uncertainty

information, however, is not only limited to feature extraction. Dorini and Goldenstein [12] have shown how to directly integrate this positional uncertainty into KLT tracking.

Integrating this positional uncertainty into the alignment problem has been of interest in the photogrammetry community [46], as well as in the computer vision community [6, 30], and has been investigated from a statistical perspective [28, 29]. Early works have debated the usefulness of uncertainty information for estimating the parameters of the fundamental matrix. Brooks *et al*. [6] used covariance information from the Harris corner detector beneficially and investigated the effects of correctly extracted covariance matrices of the error distribution for estimating the fundamental matrix. Kanazawa *et al*. [30] also showed the theoretical benefits of uncertainty information but questioned its practical use. Given that covariance matrices are too similar, nearly isotropic and homogeneous, no substantial information can be gained for estimating the fundamental matrix. Kanazawa *et al*. argued that corner detectors are designed to extract keypoints with a similar structure resulting in similar covariance matrices, limiting their use. However, Zeisl *et al*. [76] have demonstrated the possibility to extract anisotropic and inhomogeneous covariances in practical situations.

2.4. Visual Odometry Systems

Visual odometry is the task of recreating the trajectory of a camera motion from a stream of images. The basis of all visual odometry (VO) systems is the relative pose estimation between a pair of images or multiple images. This section gives an overview of VO systems with a focus on rotation-only odometry systems due to their relevance to this thesis.

The different approaches to relative pose estimation as presented in Sec. 2.1 lead to different VO systems. Algorithms that use feature-based relative pose estimation include PTAM [31] and ORB-SLAM [47, 48] with feature extraction and matching but also approaches with KLT tracks [71]. In recent years direct methods like LSD-SLAM [14] and DSO [13] have gained popularity.

A common problem of VO systems is a drift in the trajectory estimation over time. Small errors in the estimated rotation and translation accumulate, worsening the overall estimate with time. Reducing the error of the relative pose between two images also reduces this drift. Furthermore, additional methods allow further reduction of long-term drift in visual odometry. The following will give examples of some of the most popular approaches.

Bundle adjustment, for example, used in [59], tries to jointly optimize the position of 3D structure and the poses of multiple camera viewpoints by exploiting covisibility of the same features in multiple views. The advantage over two-view relative pose

estimation is the additional information that can be inferred about a feature in 3D space due to the different viewpoints. Additionally, it allows for the inclusion of different feature types like points, lines, and curves. An introduction to bundle adjustment and an overview of different approaches can be found in the excellent survey by Triggs *et al*. [69].

Including multiple views in the optimization increases the necessary computational resources. This is common to approaches that consider the covisibility of features between more than two directly subsequent images [59, 13, 10]. To keep the optimization as a constant time algorithm, Sibley *et al*. [59] proposed a sliding window approach for closely approximating the all-time maximum likelihood estimate of all images. Similar approaches that keep the number of images optimized constant over the whole stream have been made for pose graph optimization [10] and with a more sophisticated keyframe-based approach for direct methods [13].

Leutenegger *et al*. [38] argue for the usage of *keyframes* for visual odometry. Not every frame provides useful information to the odometry problem that was not already included in previous frames. Discarding these frames results in a sparse graph of keyframes and landmarks for optimization. To keep the optimization fast old keyframes are marginalized out, leaving an optimization window. In contrast to previous sliding window approaches, the time between two keyframes can be arbitrarily large. The keyframe approach has also found its application in direct methods with DSO [13].

Pose graph optimization is a concept not only found in visual odometry but also in odometry and SLAM in general. It estimates the set of poses from pairwise relative pose estimates [8]. A special case of pose graph estimation is rotation averaging, where, instead of optimizing a graph of poses, only a graph of relative rotations is optimized. Averaging out the inaccuracies of the single relative rotation estimations leads to an improved result. Dellaert *et al*. [11] showed how to obtain globally optimal solutions for rotation averaging for rotations in any dimension by relaxing the problem to even higher dimensions. Chatterjee and Govindu [9] propose an iteratively reweighted least-squares approach to rotation averaging resulting in an algorithm suited for large-scale problems. MRO [10] combines this approach with a sliding window on the graph optimization to obtain a constant time rotation-only odometry algorithm.

Loop closure allows the elimination of large-scale drift in the pose estimations by revisiting locations. Given that a camera sees the same structure at different times of the image stream, loops in the trajectory can be identified (loop detection), leading to new constraints for the camera motion.

Loop detection requires place recognition in order to work. A popular technique used in [47, 10] is to use a bag-of-words [51] to describe images based on their features in order to find similar images over a large time difference. LDSO [18], a direct method also utilizes a bag-of-words in order to identify loops in the trajectory. A more thorough

overview of different approaches for loop detection can be found in [72].

In ORB-SLAM [47] constraints between subsequent images, sharing observations of enough points, and loop constraints found by loop detection are stored in a covisibility graph. These constraints are used for pose graph optimization as in [63] to reduce long-term drift. Further examples that use pose graph optimization for loop closure for direct methods can be found in [14] and [18].

Common among these approaches to visual odometry is that without additional information, their performance deteriorates for purely rotational motion, given no additional information. Sec. 2.1 already illustrated the problems with essential or fundamental matrix approaches. But not only these approaches are badly constrained for pure rotations. Methods that have a 3D representation of the feature position like the direct method DSO [13] find it infeasible to maintain them. Panoramic representations without depth information are needed as a fallback [19, 55] but make the initialization of new points in 3D difficult.

A popular method to address purely rotational motion is to incorporate complementary data in the form of inertial measurements into the relative pose estimations leading to visual-inertial odometry systems [38, 49, 64]. The NEC introduced by Kneip *et al.* [33] allows for VO systems to handle pure rotations without inertial information. It is used as a relative pose estimation basis for several rotation-only odometry systems [37, 10].

Rotation-Only Bundle Adjustment (ROBA) [37] expands the eigenvalue-based relative rotation estimation of the NEC [32] to multiple views. It combines the ideas of the NEC with bundle adjustment that allows for rotation-averaging with optimization directly on the image features. ROBA optimizes the relative rotations over a covisibility graph using a gradient-based optimization algorithm.

Monocular Rotational Odometry (MRO) [10] also builds a covisibility graph over relative rotations obtained by the eigenvalue-based optimization of the NEC [32] using ORB features. Absolute rotations are obtained by a windowed rotation averaging version of [9]. MRO also includes a bag-of-words loop closure to address long-term drift.

2.5. Sum of Generalized Rayleigh Quotients

The sum of generalized Rayleigh quotients (GRQs) is a generalization of the Rayleigh Quotient. While recent applications in data science and wireless communications lead to optimization problems over the sum of GRQs, the energy function derived in Sec. 3.2 leads to a similar optimization. While the sum of GRQs is explained in more detail in Sec. 3.3.2, this section gives an overview of recent advances made in the optimization

of the sum of GRQs.

Optimizing the sum of GRQs over the unit sphere has been of interest recently in the community of data science and wireless communications [78, 79, 4]. While this thesis is in neither of these fields of study, the resulting energy function obtained by incorporating uncertainty into the NEC is a sum of GRQs in the translation. Therefore, the advances made in the optimization of the sum of GRQs are of interest to this work.

Zhang [78] shows how to optimize the sum of two GRQs which was later expanded to an arbitrary number of GRQs by Zhang and Chang [79]. Due to its relation to eigenvector and eigenvalues, Zhang and Chang [79] proposed a numerical optimization with the self-consistent-field (SCF) method. The SCF method finds frequent use in the field of electronic structure calculations but also lends itself to optimize the sum of GRQs over the unit sphere. Since the sum of GRQs exhibits many local minima, the SCF method outperforms generic manifold optimization. However, convergence to a global optimum is not guaranteed. A more thorough investigation into the sum of GRQs can be found in [4] which also expands the SCF to a trust-region SCF for better convergence.

3. Method

The following chapter starts by stating the notation used throughout this work. Then we introduce the NEC by Kneip *et al*. [33] (see Sec. 3.1). Sec. 3.2 explains how to integrate uncertainty information into the NEC. We show how to propagate the uncertainty through the NEC to derive the PNEC energy function. Sec. 3.3 shows that the optimization scheme proposed by Kneip and Lynen [32] for the NEC cannot be applied naively to the PNEC. We propose an optimization scheme tailored to the PNEC energy function. The chapter closes with further investigations into the PNEC energy function, with respect to singularities, in Sec. 3.4. We also present implementation details for covariance information extraction from KLT tracks.

Since this thesis is also the topic of a submission by multiple authors most figures, tables, and algorithms are identical. An overview over parts not created by the author of this thesis can be found in Appendix C.

Notation The following notation is used in this work. Vectors (*e.g.* $\boldsymbol{f} \in \mathbb{R}^n$) are denoted by bold lowercase letters and matrices (*e.g.* $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$) by bold uppercase letters. The superscript $^\top$ applied to a vector or matrix denotes the respective transposed. $\|\cdot\|$ is the Euclidean norm of a vector. 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 allows the cross product between two vectors to be written as a matrix-vector product, *i.e.* $\boldsymbol{u} \times \boldsymbol{v} = \hat{\boldsymbol{u}}\boldsymbol{v}$. When talking about absolute or relative camera poses a rigid-body transformation is used to describe its orientation and position. It is represented by a rotation matrix $\boldsymbol{R} \in SO(3)$ and a unit length translation $\boldsymbol{t} \in \mathbb{R}^3$ for the orientation and position, respectively. $\|\boldsymbol{t}\| = 1$ is imposed since the two-view problem is scale invariant. Any additional notation used in this work is introduced when appropriate.

3.1. Normal Epipolar Constraint

This section summarizes the main idea of the NEC proposed in [33] and derives the energy function used in [37]. For a more detailed explanation of the NEC and more insight into the geometry of the problem the reader is referred to [33, 32, 37, 36].

The NEC is a constraint on the *epipolar plane normal vectors* created from feature

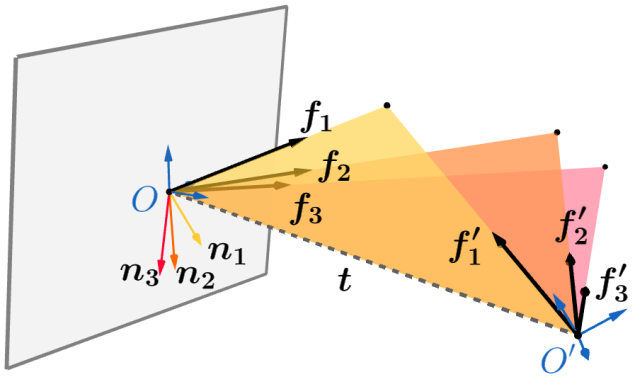


Figure 3.1.: Geometry of the NEC [33]. Each observed feature results in a correspondence pair represented by unit bearing vectors f_i and f'_i in the host frame (O) and the target frame (O'), respectively. Each pair, together with translation spans an epipolar plane (*yellow, orange, red*) with a corresponding normal vector n_i given by Eq. 3.4. All epipolar planes intersect in the translation (*dashed line*). The normal vectors span the *epipolar normal plane* (*gray*). For visual clarity only three feature correspondences are shown.

correspondences of two camera frames. Given a host frame and a target frame observing the same scene, an object generates a feature correspondences (x_i, x'_i) in the coordinate systems of the camera views. The relationship

$$x_i = Rx'_i + t \tag{3.1}$$

between the object position in the target frame x'_i and in the host frame x_i is given by the relative rotation R and translation t between the two frames.

The position of the feature correspondence can be described by a pair of two unit-bearing vectors (f_i, f'_i). In the host frame by

$$f_i = \frac{x_i}{\|x_i\|} \tag{3.2}$$

and in the target frame by

$$f'_i = \frac{x'_i}{\|x'_i\|}, \tag{3.3}$$

respectively. Each of the bearing vector pairs, together with the translation vector spans

an epipolar plane. Each plane is represented by its normal vector

$$\boldsymbol{n}_i = \boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i, \tag{3.4}$$

the *epipolar plane normal vector*. Since the translation vector spans the epipolar plane, all normal vectors are orthogonal to $\boldsymbol{t}$, making them coplanar. Together they span a plane that has $\boldsymbol{t}$ as its normal vector. We refer to this plane as the *epipolar normal plane* in this work. Fig. 3.1 shows the geometry of the NEC with the epipolar planes, the normal vectors, and the *epipolar normal plane*.

The rotation between two camera frames can be estimated by enforcing the coplanarity of the *epipolar plane normal vectors*. For an estimated rotation $\boldsymbol{R}$ the normal vector constructed from the rotation and a feature correspondence pair will not necessarily be orthogonal to the translation $\boldsymbol{t}$ and therefore not lie in the *epipolar normal plane*. The residual is given by the normalized epipolar error [37]

$$e_i = |\boldsymbol{t}^\top \boldsymbol{n}_i|, \tag{3.5}$$

i.e. the Euclidean distance of the normal vector to the *epipolar normal plane*. A least squares energy function

$$E(\boldsymbol{R}, \boldsymbol{t}) = \sum_i e_i^2 = \sum_i |\boldsymbol{t}^\top (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)|^2 \tag{3.6}$$

is constructed over all feature correspondences pairs. Sec. 3.3 shows how this energy function is minimized in order to estimate the rotation.

3.2. Probabilistic Normal Epipolar Constraint

This section expands on the idea of the NEC by correctly accounting for uncertainty information of feature positions in the constraint. Sec. 3.2.1 presents how to derive uncertainty information of the unit bearing vectors from the uncertainty in the image plane features using the unscented transform. We show in Sec. 3.2.2 how to propagate this uncertainty even further to the residual to obtain the weighted PNEC energy function.

3.2.1. Uncertainty Propagation

The PNEC accounts for error in the feature position of feature correspondence pairs. The formulation of the PNEC presented in this work considers the error to be entirely in the target frame (see Fig. 3.2). An equivalent formulation can be derived analogously

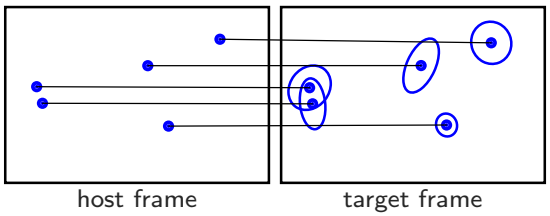


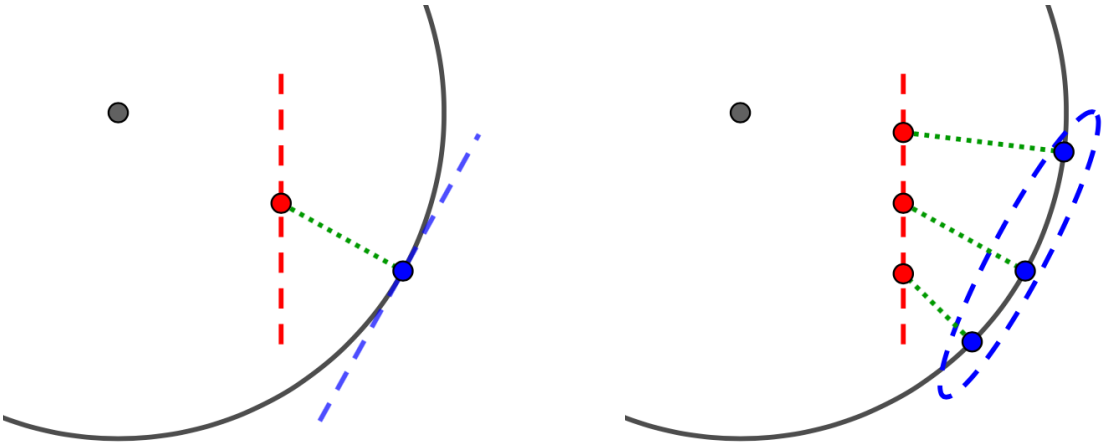
Figure 3.2.: Illustration of feature correspondences with feature position uncertainties. The feature position error is considered to be in the target frame. The probabilistic normal epipolar constraint (PNEC) expands the NEC by incorporating this uncertainty of this error. The PNEC assumes a Gaussian error distribution represented by the covariance ellipses in the target frame.

for a position error in the host frame.

The PNEC assumes that the error distribution follows as 2D Gaussian distribution on the image plane. For each correspondence pair the error distribution is characterized by a 2D covariance matrix $\mathbf{\Sigma}_{2D,i}$. To derive the PNEC energy function this covariance is propagated through to the NEC residual Eq. 3.5 to estimate the distribution of the residual.

The first step obtains a 3D error distribution $\mathbf{\Sigma}_i$ of the bearing vector $\boldsymbol{f}'_i$ by using the unscented transform [70] to project $\mathbf{\Sigma}_{2D,i}$ onto the unit sphere. The following paragraphs give an overview over the unscented transform and show how it is used to obtain $\mathbf{\Sigma}_i$ for omnidirectional and pinhole cameras.

The Unscented Transform The unscented transform approximates the mean and covariance of a Gaussian distribution after applying a non-linear transformation to it. Given an initial distribution $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_X, \mathbf{\Sigma}_X)$ and a non-linear function $\boldsymbol{y} = f(\boldsymbol{x})$ the unscented transform computes a Gaussian approximation $\boldsymbol{Y}_u \sim \mathcal{N}(\boldsymbol{\mu}_Y, \mathbf{\Sigma}_Y)$ of the distribution of $\boldsymbol{Y}$. The unscented transform computes the mean and covariance from selected points to which the non-linear transformation is applied. Fig. 3.3 illustrates the difference of using the unscented transform to a linear approximation for a pinhole camera in 2D. Given a covariance matrix $\mathbf{\Sigma}_X$ with rank n the unscented transforms



(a) linear approximation (b) unscented transform

Figure 3.3.: Illustration of the difference between linear approximation (a) and unscented transform (b) for the projection of the image plane covariance onto the bearing vector with unit length (in 2D). The linear approximation of the projection gives a covariance tangential to the unit-sphere—the covariance matrix does not have full rank. The unscented transform projects multiple points onto the unit sphere and captures the non-linearity of the projection—the covariance matrix has full rank.

selects $2n+1$ points as well as corresponding weights around the mean using

$$\begin{aligned} \xi_0 &= \mu_X \\ w_0 &= \frac{\kappa}{n+\kappa} \\ \xi_{i,i+n} &= \mu_X \pm \sqrt{n+\kappa}\boldsymbol{C}_i \quad i = 1 \dots n \\ w_{i,i+n} &= \frac{1}{2(n+\kappa)} \qquad\quad i = 1 \dots n, \end{aligned} \tag{3.7}$$

where $\boldsymbol{C}_i$ is the i-th column of the matrix $\boldsymbol{C}$ such that $\boldsymbol{\Sigma}_x = \boldsymbol{C}\boldsymbol{C}^\top$. κ controls the spread of the points, which is set to its default value $\kappa = 1$ in this work. A popular way to compute $\boldsymbol{C}$ is using the Cholesky-decomposition of $\boldsymbol{\Sigma}_X$. The non-linear function $f(x)$ is applied to the points

$$\zeta_i = f(\xi_i) \tag{3.8}$$

resulting in $2n+1$ points of the new distribution. The mean and covariance of this new distribution are computed

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{Y}} &= \sum_{i=0}^{2n+1} w_i \boldsymbol{\zeta}_i, \\ \boldsymbol{\Sigma}_{\boldsymbol{Y}} &= \sum_{i=0}^{2n+1} w_i (\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{\boldsymbol{Y}})(\boldsymbol{\zeta}_i - \boldsymbol{\mu}_{\boldsymbol{Y}})^\top \end{aligned} \tag{3.9}$$

using the weights. For the PNEC we project the 2D covariance $\boldsymbol{\Sigma}_{2\mathrm{D}}$ of the target frame feature in the image onto the unit-sphere in 3D.

The Unscented Transform for Omnidirectional Cameras Feature points for omnidirectional cameras are not located on a 2D plane, but on a sphere in 3D. Their covariance, still assumed to have rank 2, is therefore embedded in 3D space. The covariance matrix $\boldsymbol{\Sigma}_{spherical} \in \mathbb{R}^{3\times 3}$ does not have full rank. Since it is not positive definite, the Cholesky-decomposition is not defined for it. In order to still use the unscented transform for omnidirectional camera the Cholesky-decomposition of a 2×2 sub-matrix $\boldsymbol{C}\boldsymbol{C}^\top = \boldsymbol{\Sigma}_{2\mathrm{D}}$ of the form

$$\begin{pmatrix} \boldsymbol{\Sigma}_{2\mathrm{D}} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 0 \end{pmatrix} = \boldsymbol{R}\boldsymbol{\Sigma}_{spherical}\boldsymbol{R}^\top. \tag{3.10}$$

is used. A matrix $\boldsymbol{R}$ that gives us such a form is the rotation matrix

$$\boldsymbol{R} = \frac{1}{\|\boldsymbol{\mu}\|} \begin{pmatrix} \|\boldsymbol{\mu}\| - \frac{\mu_1^2}{\|\boldsymbol{\mu}\|+\mu_3} & -\frac{\mu_1\mu_2}{1+\mu_3} & -\mu_1 \\ -\frac{\mu_1\mu_2}{1+\mu_3} & \|\boldsymbol{\mu}\| - \frac{\mu_2^2}{\|\boldsymbol{\mu}\|+\mu_3} & -\mu_2 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix}, \tag{3.11}$$

where μ_i denotes the i-th entry of the vector $\boldsymbol{\mu}$. $\boldsymbol{R}$ aligns the feature point with the z-axis and the covariance with the xy-plane.

Using this realignment of the covariance matrix the points for the unscented transform are selected as

$$\begin{aligned} \boldsymbol{\xi}_0 &= \boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\xi}_{i,i+n} &= \boldsymbol{\mu}_{\boldsymbol{X}} \pm \sqrt{n+\kappa}\boldsymbol{R}^\top \begin{pmatrix} \boldsymbol{C}_i \\ 0 \end{pmatrix} \quad i = 1 \dots n\,. \end{aligned} \tag{3.12}$$

For omnidirectional cameras we use the non-linear function

$$f(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \tag{3.13}$$

the projection onto the unit sphere.

The Unscented Transform for Pinhole Cameras Feature points for pinhole cameras are located on the image plane with 2D covariance matrices for which the Cholesky-decomposition is defined. Points for the unscented transform are selected according to Eq. 3.7. The non-linear function

$$f(\boldsymbol{x}) = \frac{\boldsymbol{K}^{-1}\boldsymbol{x}}{\|\boldsymbol{K}^{-1}\boldsymbol{x}\|} \tag{3.14}$$

is given by the unprojection of the point with the inverse camera matrix [24] and subsequent projection onto the unit sphere.

3.2.2. The Probabilistic Normal Epipolar Constraint Energy Function

Because the uncertainty in the image is only a few pixels, the approximation using the unscented transform is reasonable since the non-linear function is locally well approximated by a linear function. The covariance of the bearing vector $\boldsymbol{f}'_i$, obtained by the unscented transform, is then propagated through the linear functions of the normalized epipolar error (Eq. 3.4 and Eq. 3.5) to give the univariate Gaussian distribution of the residual error $\mathcal{N}(0, \sigma_i^2)$, with the variance

$$\sigma_i^2(\boldsymbol{R}, \boldsymbol{t}) = \boldsymbol{t}^\top \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i{}^\top \boldsymbol{t}\,. \tag{3.15}$$

The variance is incorporated into the energy function by using a Mahalanobis distance instead of the Euclidean distance. We obtain the PNEC energy function

$$E_P(\boldsymbol{R}, \boldsymbol{t}) = \sum_i \frac{e_i^2}{\sigma_i^2} = \sum_i \frac{|\boldsymbol{t}^\top (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)|^2}{\boldsymbol{t}^\top \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i{}^\top \boldsymbol{t}}, \tag{3.16}$$

a weighted version of the NEC energy function. The following discusses the optimization of the PNEC energy function for rotation estimation.

3.3. Optimization

Kneip and Lynen [32] presented an elegant eigenvalue-based optimization scheme for the NEC. Sec. 3.3.1 shows why we cannot naively apply this eigenvalue-based optimization to the PNEC. The rest of this section proposes a scheme to estimate the rotation and translation using the PNEC by optimizing its energy function. Similar to the NEC optimization is split up into two sub-problems, over the translation (see

Sec. 3.3.2) and the rotation (see Sec. 3.3.3). Algorithms tailored to each sub-problem are presented. They are employed alternately to get an iterative optimization scheme. Furthermore, a refinement is proposed to improve the results (see Sec. 3.3.4).

3.3.1. Eigenvalue Based Optimization

The optimization of the NEC energy function can be split up into two sub-problems, an optimization over the rotation and an analytical optimization over the translation. Solving the rotation-only sub-problem leads to an eigenvalue-based optimization scheme independent of the translation. Following [37], the energy function of the NEC Eq. 3.6 can be rewritten as $E(\boldsymbol{R}, \boldsymbol{t}) = \boldsymbol{t}^\top \boldsymbol{M}(\boldsymbol{R})\boldsymbol{t}$ using the symmetric and positive-semi-definite Gramian matrix

$$\boldsymbol{M}(\boldsymbol{R}) = \sum_i (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)(\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)^\top. \tag{3.17}$$

In this form, the energy function is quadratic in $\boldsymbol{t}$ such that the optimization over the translation can be done analytically. The solution is given by the eigenvector corresponding to the smallest eigenvalue $\lambda_{\min}$ of $\boldsymbol{M}(\boldsymbol{R})$.

This leaves the sub-problem of optimizing the matrix $\boldsymbol{M}(\boldsymbol{R})$ such that the overall energy is minimized. Because $\boldsymbol{t}$ is chosen to be the eigenvector, the minimization

$$\begin{aligned} \min_{\substack{\boldsymbol{R}\in \mathrm{SO}(3) \\ \boldsymbol{t}:\, \|\boldsymbol{t}\|=1}} \boldsymbol{t}^\top \boldsymbol{M}(\boldsymbol{R})\boldsymbol{t} &= \min_{\boldsymbol{R}\in \mathrm{SO}(3)} \boldsymbol{t}^\top \lambda_{\min}(\boldsymbol{M}(\boldsymbol{R}))\boldsymbol{t} \\ &= \min_{\boldsymbol{R}\in \mathrm{SO}(3)} \lambda_{\min}(\boldsymbol{M}(\boldsymbol{R})) \end{aligned} \tag{3.18}$$

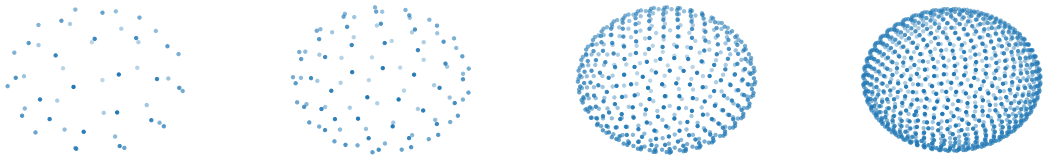
is reduced to minimizing $\lambda_{\min}$ over $\boldsymbol{R}$. Kneip and Lynen [32] give an interpretation of this sub-problem as optimization over the normal vectors as a point cloud and $\boldsymbol{M}(\boldsymbol{R})$ as a second order momentum matrix. Furthermore, they propose optimizing over $\boldsymbol{R}$ using a Levenberg-Marquardt algorithm [39, 45].

While the energy of the NEC in $\boldsymbol{t}$ is given by a Rayleigh quotient of the form

$$E(\boldsymbol{R}, \boldsymbol{t}) = \frac{\boldsymbol{t}^\top \boldsymbol{M}(\boldsymbol{R})\boldsymbol{t}}{\boldsymbol{t}^\top \boldsymbol{t}}, \tag{3.19}$$

the energy function of the PNEC in $\boldsymbol{t}$ is given by the sum of generalized Rayleigh quotients (GRQs) of the form

$$E_P(\boldsymbol{R}, \boldsymbol{t}) = \sum_i \frac{\boldsymbol{t}^\top \boldsymbol{A}_i \boldsymbol{t}}{\boldsymbol{t}^\top \boldsymbol{B}_i \boldsymbol{t}} \tag{3.20}$$



(a) 50 points (b) 100 points (c) 500 points (d) 1000 points

Figure 3.4.: Fibonacci lattice point generation for different number of points in 3D. The Fibonacci Lattice distributed a given number of points equally on the surface of a sphere. This distribution is used a sampling for different starting points of the SCF algorithm. See Alg. 2 for more details.

and thus the optimization over $\boldsymbol{t}$ is not simply given by an eigenvalue as for the NEC.

3.3.2. Optimizing over t

Optimizing the sum of GRQs has been of interest for data science and wireless communications [78, 79, 4]. Recent advances by Zhang *et al.* [79] have shown the self-consistent-field (SCF) [26] outperforming generic manifold optimization. The following shows how the SCF algorithm is used to optimize the PNEC energy function over $\boldsymbol{t}$.

Self-Consistent-Field Algorithm The self-consistent-field (SCF) algorithm for the sum of GRQs optimizes energy functions of the following form

$$E_P(\boldsymbol{R}, \boldsymbol{t}) = \sum_i \frac{\boldsymbol{t}^\top \boldsymbol{A}_i \boldsymbol{t}}{\boldsymbol{t}^\top \boldsymbol{B}_i \boldsymbol{t}} + \boldsymbol{t}^\top \boldsymbol{D} \boldsymbol{t}\,, \tag{3.21}$$

where $\boldsymbol{A}_i, \boldsymbol{D}$ are symmetric matrices and $\boldsymbol{B}_i$ are symmetric, positive definite matrices. For the PNEC the matrices are given by

$$\begin{aligned} \boldsymbol{A}_i &= \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{f}_i' \boldsymbol{f}_i'^\top \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i^\top\,, \\ \boldsymbol{B}_i &= \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i^\top + c\boldsymbol{I}_3\,, \\ \boldsymbol{D} &= \boldsymbol{0}\,, \end{aligned} \tag{3.22}$$

where $\boldsymbol{B}_i$ needs a regularization term to make it positive definite. The SCF algorithm is an iterative process for optimizing the energy function. The main step of the SCF algorithm is to compute the $\boldsymbol{E}$-matrix [4, Eqn. (2.3)], for the PNEC a 3×3 symmetric

Algorithm 1: SCF Optimization w/ Globalization

Data: Fixed rotation $\tilde{\boldsymbol{R}}$
Result: Optimized translation $\boldsymbol{t}^*$
Sample the Fibonacci Lattice with K points (cf. Alg. 2)
$\{\bar{\boldsymbol{t}}_k\}_k \leftarrow \text{FibonacciLattice}(K)$
Select the starting point with minimal Energy (cf. Eq. 3.16)
$\boldsymbol{t}_0 \leftarrow \arg\min_k E_P(\tilde{\boldsymbol{R}}, \bar{\boldsymbol{t}}_k)$
for $s \leftarrow 1$ **to** S **do**
  Construct the $\boldsymbol{E}$-matrix (cf. Eq. 3.23)
  $\boldsymbol{E}_s \leftarrow \boldsymbol{E}(\tilde{\boldsymbol{R}}, \bar{\boldsymbol{t}}_{s-1})$
  Eigendecompose $\boldsymbol{E}_s \in \mathbb{R}^{3\times 3}$ using $\boldsymbol{E}_s = \boldsymbol{E}_s^\top$
  $\lambda_1, \lambda_2, \lambda_3, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \leftarrow \text{eig}(\boldsymbol{E}_s)$ *s.t.* $\lambda_1 \leq \lambda_2 \leq \lambda_3$
  Set $\boldsymbol{t}_s$ as the eigenvector with maximal eigenvalue
  $\boldsymbol{t}_s \leftarrow \boldsymbol{v}_3$
end

matrix given by

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{R}, \boldsymbol{t}) &= \sum_i w_i \cdot \left(\boldsymbol{t}^\top \boldsymbol{B}_i \boldsymbol{t} \cdot \boldsymbol{A}_i - \boldsymbol{t}^\top \boldsymbol{A}_i \boldsymbol{t} \cdot \boldsymbol{B}_i \right) , \\ w_i &= (\boldsymbol{t}^\top \boldsymbol{B}_i \boldsymbol{t})^{-2} \cdot \prod_j \boldsymbol{t}^\top \boldsymbol{B}_j \boldsymbol{t} . \end{aligned} \tag{3.23}$$

The weights of the form $w_i = (\boldsymbol{t}^\top \boldsymbol{B}_i \boldsymbol{t})^{-2}$ are used to avoid numerical instabilities arising from the common factor $\prod_j \boldsymbol{t}^\top \boldsymbol{B}_j \boldsymbol{t}$. The translation $\boldsymbol{t}$ for the next iteration is given by the eigenvector to the maximal eigenvalue of $\boldsymbol{E}$. Alg. 1 summarizes the steps of the SCF optimization.

Although the SCF algorithm outperforms generic manifold optimization, it is not guaranteed to converge to the global optimum. To improve the results of the SCF, we use a simple yet effective globalization strategy in this work. Optimizing the PNEC energy function for $\boldsymbol{t}$ over the unit sphere in $\mathbb{R}^3$ has low dimensionality. Sampling evenly distributed points on the unit sphere using the Fibonacci lattice [21] gives initial points $\boldsymbol{t}_k$. The initial point with the lowest energy value is picked and used as the starting value for the SCF-Iteration. Alg. 2 details the Fibonacci-lattice-based point generation on the sphere in $\mathbb{R}^3$.

Algorithm 2: Fibonacci Lattice Point Generation

Data: Number of points K
Result: Points on the sphere $\{\bar{\boldsymbol{t}}_k = (x_k, y_k, z_k)\}_k$
Compute the golden ratio angle
$\phi \leftarrow \pi \cdot (3 - \sqrt{5})$
for $k \leftarrow 1$ **to** K **do**
  Compute the k^{th} y-coordinate $y_k \in [-1, 1]$
  $y_k \leftarrow 1 - 2 \cdot \frac{k-1}{K-1}$
  Compute the radius in the x-z-plane
  $r_{xz} \leftarrow \sqrt{1 - y_k^2}$
  Compute the remaining coordinates x_k, z_k for $\bar{\boldsymbol{t}}_k$
  $x_k \leftarrow r_{xz} \cdot \cos((k-1)\phi)$
  $z_k \leftarrow r_{xz} \cdot \sin((k-1)\phi)$
end

3.3.3. Optimizing over R

The eigenvalue-based optimizer, proposed by Kneip and Lynen [32], efficiently estimates the rotation using the NEC with a Levenberg-Marquardt approach. Estimating the rotation using the PNEC energy function can be done similarly. While the rotation estimation cannot be decoupled completely from the translation, an optimization scheme similar to the popular iteratively reweighted least squares (IRLS) is employed that reuses the idea of Kneip and Lynen [32].

Given an estimate of the rotation and translation $(\boldsymbol{R}_p, \boldsymbol{t}_p)$ a fixed weight $\tilde{\sigma}_i = \sigma_i(\boldsymbol{R}_p, \boldsymbol{t}_p)$ is computed for all feature correspondences pairs. Instead of optimizing the eigenvalue of the unweighted matrix $\boldsymbol{M}(\boldsymbol{R})$ (see Eq. 3.17) the eigenvalue of the weighted matrix

$$\boldsymbol{M}_P(\boldsymbol{R}; \{\tilde{\sigma}_i\}_i) = \sum_i \frac{(\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}_i')(\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}_i')^\top}{\tilde{\sigma}_i^2} \tag{3.24}$$

is optimized. Due to the fixed weights the matrix only depends on the rotation $\boldsymbol{R}$, such that the optimizer of Kneip and Lynen [32] can be employed.

3.3.4. Least Squares Refinement

The previously proposed optimization techniques are efficient in estimating the rotation and translation. However, they are not guaranteed to find a minimum of the PNEC energy function. We use a least-squares refinement strategy to further improve the

Algorithm 3: 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. 3.3.3)

$\boldsymbol{R}_s \leftarrow \mathrm{Opt}_{\boldsymbol{R}} \; \lambda_{\min}(\boldsymbol{M}_P(\boldsymbol{R}; \{\tilde{\sigma}_{i,s-1}\}_i))$

Optimize over $\boldsymbol{t}$ (cf. Sec. 3.3.2)

$\boldsymbol{t}_s \leftarrow \mathrm{Opt}_{\boldsymbol{t}} \; E_P(\boldsymbol{R}_s, \boldsymbol{t})$

Update the weights (cf. Eq. 3.15)

$\tilde{\sigma}_{i,s} \leftarrow \sigma_i(\boldsymbol{R}_s, \boldsymbol{t}_s) \; \forall i$

end

Least-Squares Refinement (cf. Sec. 3.3.4) using $(\boldsymbol{R}_S, \boldsymbol{t}_S)$ as starting value

$\boldsymbol{R}^*, \boldsymbol{t}^* \leftarrow \mathrm{Opt}_{\boldsymbol{R},\boldsymbol{t}} \, E_P(\boldsymbol{R}, \boldsymbol{t})$

results. It is effective in finding a local minimum of the energy function given a starting point [53].

Using the Levenberg-Marquardt algorithm the least squares formulation of the PNEC with the residuals

$$r_i(\boldsymbol{R}, \boldsymbol{t}) = \frac{\boldsymbol{t}^\top (\boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}'_i)}{\sqrt{\hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i^\top}} \tag{3.25}$$

is optimized over the rotation $\boldsymbol{R}$ and the translation $\boldsymbol{t}$ simultaneously. The results of the previously presented iterative optimization scheme are used as the starting values. Manifold optimization [27] ensures that neither the constraints for the rotation nor the translation are violated. The special orthogonal group $\mathrm{SO}(3)$ is used for the rotation and for the translation spherical coordinates with the radius fixed to 1 ensure that $\|\boldsymbol{t}\| = 1$. Alg. 3 shows the complete optimization scheme with the iterative optimization and the least squares refinement.

3.4. Further Investigations

This section gives further insight into the PNEC energy function as well as implementation details. We use uncertainty information provided by KLT tracks on visual odometry datasets throughout this work. Sec. 3.4.1 looks at implementation details for extracting the covariance in the image from these KLT tracks and can be skipped by readers not interested in the implementation details. Sec. 3.4.2 shows that the PNEC leads to singularities in certain geometrical configurations. We investigate them and present a simple yet effective regularization to avoid them. Furthermore, Sec. 3.4.3

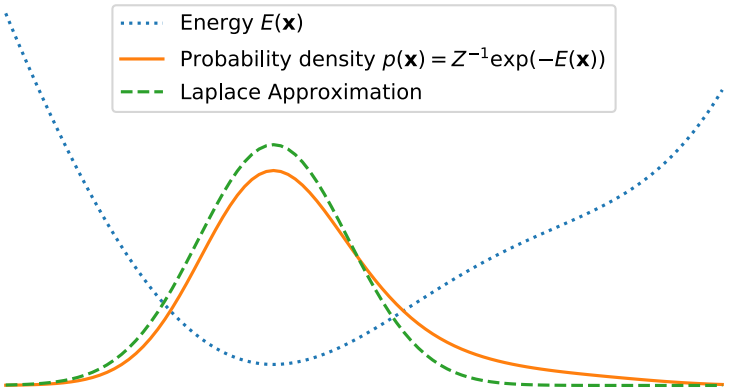


Figure 3.5.: Illustration of the relation between the energy function $E(\boldsymbol{x})$ (*dotted blue*), the normalized distribution $p(\boldsymbol{x})$ (*orange*), and the Laplace approximation centered on the mode $\boldsymbol{x}^*$ of $p(\boldsymbol{x})$ (*dashed green*). The Laplace approximation gives a Gaussian distribution based on an energy function. It is used to approximate the error distribution of the KLT tracks for the PNEC.

gives a geometric interpretation of the energy function and its singularities.

3.4.1. Extracting Covariances in the Image from Kanade-Lucas-Tomasi Tracks

In order to use the PNEC the uncertainty of a feature position has to be determined. This section presents a method to extract such information in the form of a covariance matrix for a KLT tracking system used in [71]. We use this tracking in the odometry dataset experiments of this work. The covariance matrix is based on the Gaussian approximation of a Boltzmann distribution derived from an energy function.

Given an energy function $E(\boldsymbol{x}) : \mathbb{R}^d \to \mathbb{R}$ its *Boltzmann distribution* [5, Eqn. (8.41)] is the associated probability distribution

$$p(\boldsymbol{x}) = Z^{-1} \exp(-E(\boldsymbol{x})) \tag{3.26}$$

with the normalization constant $Z > 0$, such that $\int_{-\infty}^{\infty} p(\boldsymbol{x}) = 1$. Around a minimum of $E(\boldsymbol{x})$, which is a maximum of the Boltzmann distribution, the probability distribution can locally be approximated as a Gaussian distribution using the *Laplace approximation* [5, Sec. 4.4]. Fig. 3.5 illustrates the relationship between the energy function, its Boltzmann distribution, and the Gaussian approximation. Around the local minimizer

Figure 3.6.: Covariance ellipses for position uncertainties on KITTI seq. 07. The tracks are generated using the Gaussian approximation of the KLT tracking energy function. The PNEC correctly considers the anisotropic inhomogeneous error distributions of the features. For visualization purposes the covariances are sub-sampled and enlarged. Only a sub-image is shown.

$\boldsymbol{x}^*$ of $E(\boldsymbol{x})$, the Gaussian approximation has the mean $\boldsymbol{\mu} = \boldsymbol{x}^*$ and inverse covariance

$$\boldsymbol{\Sigma}^{-1} = \frac{d^2}{d\boldsymbol{x}^2} E(\boldsymbol{x}) \bigg|_{\boldsymbol{x}=\boldsymbol{x}^*}, \tag{3.27}$$

where $\frac{d^2}{d\boldsymbol{x}^2}$ denotes the Hessian matrix.

The KLT tracking implementation [71] used in this work tracks a patch P, a pattern of pixels, from the host to the target frame. The energy function optimized for this tracking is

$$E_{KLT}(\boldsymbol{T}) = \sum_{\boldsymbol{p} \in P} \left(\frac{I_h(\boldsymbol{p})}{\bar{I}_h} - \frac{I_t(\boldsymbol{T}(\boldsymbol{p}))}{\bar{I}_t} \right)^2, \tag{3.28}$$

with the mean intensity in the host frame

$$\bar{I}_h = \frac{1}{|P|} \sum_{\boldsymbol{p} \in P} I_h(\boldsymbol{p}) \tag{3.29}$$

and the target frame

$$\bar{I}_t = \frac{1}{|P|} \sum_{\boldsymbol{p} \in P} I_t(\boldsymbol{T}(\boldsymbol{p})), \tag{3.30}$$

respectively. The transformation of a track between the host and target frame is given by $\boldsymbol{T}$, $|P|$ denotes the number of pixels in P. The implementation in [71] uses an *inverse compositional formulation* [2] for more efficient tracking. A proxy for Eq. 3.28

is optimized, and therefore, we get an approximation of the Hessian of the energy function. The Hessian has to be computed only once in the host frame. We refer a reader more interested in KLT tracking and its different formulations to [44, 68] and the excellent paper by Baker and Matthews [2], the first in a multi-paper series.

Due to the inverse compositional formulation only needing the host frame its subscript is dropped in the following such that the host frame is denoted by I. The Hessian of the energy function is computed with the Gauss-Newton approximation using the Jacobian. The system of equations to compute the Hessian is given by

$$\begin{aligned} \boldsymbol{J}_{\boldsymbol{p}_i} &= |P| \frac{\nabla I(\boldsymbol{p}_i) \sum_{\boldsymbol{p} \in P} I(\boldsymbol{p}) - I(\boldsymbol{p}_i) \sum_{\boldsymbol{p} \in P} \nabla I(\boldsymbol{p})}{\left(\sum_{\boldsymbol{p} \in P} I(\boldsymbol{p}) \right)^2} \\ \boldsymbol{J}_i &= \boldsymbol{J}_{\boldsymbol{p}_i} \begin{pmatrix} 1 & 0 & -p_{i,y} \\ 0 & 1 & p_{i,x} \end{pmatrix} \\ \boldsymbol{H}_{\mathrm{SE}(2)} &= \begin{pmatrix} \boldsymbol{J}_1 & \boldsymbol{J}_2 & \dots & \boldsymbol{J}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{J}_1 \\ \boldsymbol{J}_2 \\ \vdots \\ \boldsymbol{J}_n \end{pmatrix} \\ \boldsymbol{\Sigma}_{\mathrm{SE}(2)} &= \boldsymbol{H}_{\mathrm{SE}(2)}^{-1} \end{aligned} \tag{3.31}$$

where $\nabla I(\boldsymbol{p})$ is the image gradient and $\boldsymbol{J}_{\boldsymbol{p}_i}$ the Jacobian w.r.t. the pixel position.

All patches are tracked w.r.t. a $\mathrm{SE}(2)$ transform, that includes a 2D translation and rotation so the Hessian as well as the covariance matrix is of size 3×3. Only a 2×2 covariance sub-matrix of the patch in the host frame is relevant for the PNEC. The sub-matrix is given by the marginal over the translational coordinates. This is obtained by selecting the upper left 2×2 sub-matrix $\boldsymbol{\Sigma}_{2\mathrm{D},h}$ of $\boldsymbol{\Sigma}_{\mathrm{SE}(2)}$. To project the covariance matrix into the target frame and correctly account for the rotation in the tracking transformation $\boldsymbol{T}$ between the two frames the covariance matrix is rotated by the 2D rotation matrix of the transformation

$$\boldsymbol{\Sigma}_{2\mathrm{D},t} = \boldsymbol{R}_{\boldsymbol{T}} \boldsymbol{\Sigma}_{2\mathrm{D},h} \boldsymbol{R}_{\boldsymbol{T}}^{\top} . \tag{3.32}$$

$\boldsymbol{\Sigma}_{2\mathrm{D},t}$ is covariance matrix is used in the PNEC to calculate the variance (see Eq. 3.15).

3.4.2. Singularities of the Probabilistic Normal Epipolar Constraint

Switching from the Euclidean distance to the Mahalanobis distance for the energy function introduces singularities. This section investigates them in more detail and shows a simple yet effective regularization to avoid them.

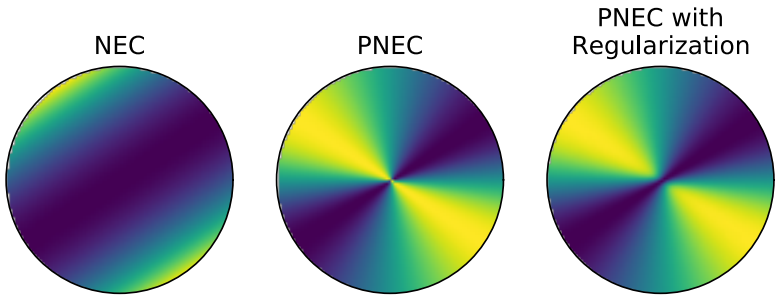


Figure 3.7.: Visualization of the energy functions of the NEC, the PNEC, and the PNEC with regularization proposed in Sec. 3.4.2. The plot visualizes the energy functions in a neighborhood of a feature $\boldsymbol{f}_i$ with regard to the translation $\boldsymbol{t}$ (in polar coordinates). The center of the circles depicts $\boldsymbol{t} = \boldsymbol{f}_i$. The PNEC has a finite discontinuity in the center for $\boldsymbol{t} = \boldsymbol{f}_i$ with only a directional limit exists (see Sec. 3.4.2). The proposed regularization removes the discontinuity and maintains the shape of the energy function otherwise.

The singularities of the PNEC arise if the variance σ_i^2 (Eq. 3.15) vanishes. This is the case if the translation is parallel to a bearing vector in the host frame $\boldsymbol{t} = \boldsymbol{f}_i$, because $\hat{\boldsymbol{f}}_i^\top \boldsymbol{t} = \boldsymbol{f}_i \times \boldsymbol{t} = 0$ makes the denominator 0. Due to the different optimization approaches of the eigenvalue based optimization and the least squares refinement the singularity behaves differently and has to be investigated separately. Nevertheless, a unified solution to avoid this singularity is shown.

The Singularity in the Eigenvalue-Based Optimization The matrix $\boldsymbol{M}_P$ used in the eigenvalue-based optimization of the rotation has no equivalent term to $\hat{\boldsymbol{f}}_i^\top \boldsymbol{t}$ in the numerator and therefore tends to infinity for $\boldsymbol{t} \to \boldsymbol{f}_i$. Since the position of the singularity is independent of the translation the eigenvalue-based optimization can be dominated by a single pair of $(\boldsymbol{f}_i, \boldsymbol{f}_i')$.

The Singularity in the Least-Squares Refinement In contrast to the eigenvalue-based optimization, the energy function of the least-squares refinement includes the term $\hat{\boldsymbol{f}}_i^\top \boldsymbol{t}$ in the numerator and the denominator. This can be seen if the energy function is rewritten as

$$e_{P,i}^2(\boldsymbol{R}, \boldsymbol{t}) = \frac{|(\boldsymbol{t} \times \boldsymbol{f}_i)^\top \boldsymbol{R} \boldsymbol{f}_i'|^2}{(\boldsymbol{t} \times \boldsymbol{f}_i)^\top \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top (\boldsymbol{t} \times \boldsymbol{f}_i)} \,. \tag{3.33}$$

The residual is bounded and possesses a finite discontinuity. Fig. 3.7 illustrates the value of the residual near the singularity. We present the directional limit for this finite discontinuity later. The discontinuity is finite, and therefore, the singularity is

less problematic than in the eigenvalue-based optimization. However, it still poses challenges. While the function values of the residual are finite, the derivative is not bounded. The unbounded derivative is problematic for optimization.

Removing the Singularity To address the singularity in both cases, the eigenvalue-based optimization and the least-squares refinement, we employ a simple yet effective regularization scheme. A variance of the form $\sigma_i'^2 = \sigma_i^2 + c$ removes the discontinuity and the singularity. The same type of regularization is also used in the SCF computation of $\boldsymbol{B}_i$ to give it full rank (see Eq. 3.22).

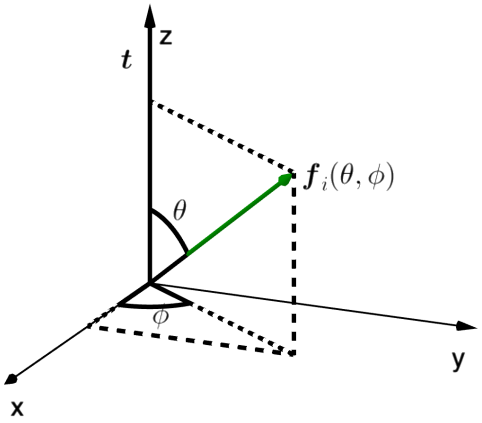


Figure 3.8.: Approaching the singularity on the unit sphere using spherical coordinates. For $\theta \to 0$ the feature vector $\boldsymbol{f}_i \to \boldsymbol{t}$ resulting in a directional limit of the residual. The direction from which the singularity is approached is determined by ϕ.

The Directional Limit The following shows that the singularity of the PNEC residual has no limit. This is shown by deriving the directional limit on the unit sphere. We present the directional limit of $\boldsymbol{f}_i \to \boldsymbol{t}$ instead of the directional limit of $\boldsymbol{t} \to \boldsymbol{f}_i$, since it does not change the limit but more accurately reflects real world scenarios. To derive this directional limit spherical coordinates of the form

$$\boldsymbol{x} = r \begin{pmatrix} \sin\theta \sin\phi \\ -\sin\theta \cos\phi \\ \cos\theta \end{pmatrix} \tag{3.34}$$

are used, with the radius r fixed to 1. Without loss of generality

$$\boldsymbol{t} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.35}$$

is chosen, since every problem can be rotated as a whole in 3D such that this holds for $\boldsymbol{t}$. Approaching $\boldsymbol{t}$ on the unit sphere with the feature vector $\boldsymbol{f}_i$ is now done by writing it in spherical coordinates as

$$\boldsymbol{f}_i(\theta, \phi) = \begin{pmatrix} \sin\theta \sin\phi \\ -\sin\theta \cos\phi \\ \cos\theta \end{pmatrix}, \tag{3.36}$$

where ϕ represents the direction of the approach. Fig. 3.8 illustrates this approach on the unit sphere. Letting $\theta \to 0$ implies $\boldsymbol{f}_i \to \boldsymbol{t}$ such that our directional limit is given by

$$\lim_{\theta \to 0} \frac{|(\boldsymbol{t} \times \boldsymbol{f}_i(\theta))^\top \boldsymbol{R} \boldsymbol{f}_i'|^2}{(\boldsymbol{t} \times \boldsymbol{f}_i(\theta))^\top \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top (\boldsymbol{t} \times \boldsymbol{f}_i(\theta))}. \tag{3.37}$$

The cross product is given by

$$\boldsymbol{t} \times \boldsymbol{f}_i(\theta) = -\sin\theta \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix} = -\sin\theta \boldsymbol{k} \tag{3.38}$$

with $\boldsymbol{k}$ being the unit length vector orthogonal to $\boldsymbol{f}_i$ and $\boldsymbol{t}$. The limit then simplifies to

$$\begin{aligned} &\lim_{\theta \to 0} \frac{|(\boldsymbol{t} \times \boldsymbol{f}_i(\theta))^\top \boldsymbol{R} \boldsymbol{f}_i'|^2}{(\boldsymbol{t} \times \boldsymbol{f}_i(\theta))^\top \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top (\boldsymbol{t} \times \boldsymbol{f}_i(\theta))} \\ &= \lim_{\theta \to 0} \frac{\sin^2\theta}{\sin^2\theta} \frac{|\boldsymbol{k}^\top \boldsymbol{R} \boldsymbol{f}_i'|^2}{\boldsymbol{k}^\top \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \boldsymbol{k}} \\ &= \frac{|\boldsymbol{k}^\top \boldsymbol{R} \boldsymbol{f}_i'|^2}{\boldsymbol{k}^\top \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \boldsymbol{k}}. \end{aligned} \tag{3.39}$$

The above equation shows that the directional limit for $\theta \to 0$ exists and depends on the direction $\boldsymbol{k}$. Consequentially, the limit of the residual does not exist.

3.4.3. Geometric Interpretations of the Probabilistic Normal Epipolar Constraint

This section takes an in-depth look into the derivation of the PNEC energy function as a Mahalanobis distance. We give a geometric reasoning for why the regularization removes the singularity of the PNEC.

For the PNEC the *epipolar normal plane* is described in homogeneous coordinates by

$$\boldsymbol{p} = \begin{pmatrix} \boldsymbol{t} \\ 0 \end{pmatrix} \tag{3.40}$$

and the covariance with its singular value decomposition as

$$\boldsymbol{\Sigma}_{n,i} = \boldsymbol{R}_i^\top \boldsymbol{V}_i \boldsymbol{R}_i, \quad \boldsymbol{V}_i = diag(a^2, b^2, c^2)\,, \tag{3.41}$$

from which the whitening transform is derived. Applying it to the *epipolar normal plane* gives a new transformed plane as

$$\boldsymbol{q} = \begin{pmatrix} \boldsymbol{V}_i^{1/2}\boldsymbol{R}_i & \boldsymbol{0} \\ -\boldsymbol{n}_i^\top & 1 \end{pmatrix} \boldsymbol{p} = \begin{pmatrix} \boldsymbol{V}_i^{1/2}\boldsymbol{R}_i\boldsymbol{t} \\ -\boldsymbol{n}_i^\top \boldsymbol{t} \end{pmatrix}. \tag{3.42}$$

The original Mahalanobis distance

$$d_M = \frac{q_4}{\sqrt{q_1^2 + q_2^2 + q_3^2}} \tag{3.43}$$

is now given by the distance of the origin to the transformed plane, giving the PNEC energy function by squaring it.

$$d_M^2 = \frac{|\boldsymbol{n}_i^\top \boldsymbol{t}|^2}{\boldsymbol{t}^\top \boldsymbol{R}_i^\top \boldsymbol{V}_i^{1/2} \boldsymbol{V}_i^{1/2} \boldsymbol{R}_i \boldsymbol{t}} = \frac{|\boldsymbol{n}_i^\top \boldsymbol{t}|^2}{\boldsymbol{t}^\top \boldsymbol{\Sigma}_{n,i} \boldsymbol{t}}\,. \tag{3.44}$$

Given this geometric interpretation of the PNEC, its singularity can now be explained geometrically. The main characteristic of the distribution of $\boldsymbol{n}_i$ is that its covariance matrix $\boldsymbol{\Sigma}_{n,i} = \hat{\boldsymbol{f}}_i \boldsymbol{R} \boldsymbol{\Sigma}_i \boldsymbol{R}^\top \hat{\boldsymbol{f}}_i^\top$ only has rank 2. Since $\boldsymbol{n}_i = \boldsymbol{f}_i \times \boldsymbol{R}\boldsymbol{f}_i'$ is derived as a cross product its distribution has to be orthogonal to $\boldsymbol{f}_i$. It lies in 3D space on a 2D plane orthogonal to $\boldsymbol{f}_i$ (note that $\boldsymbol{f}_i$ is not random while $\boldsymbol{f}_i'$ is random). A Mahalanobis distance can only be defined for points on this plane. For $\boldsymbol{f}_i \neq \boldsymbol{t}$ this 2D plane and the *epipolar normal plane* intersect in a single line giving a meaningful Mahalanobis distance. This does not hold for $\boldsymbol{f}_i = \boldsymbol{t}$.

The regularization shown in Sec. 3.4.2 resolves this problem by removing the 2D

constraint on the Mahalanobis distance by giving the covariance matrix full rank. This can be seen by the equivalent formulations of the regularization as

$$\begin{aligned} \boldsymbol{t}^\top \left(\boldsymbol{\Sigma}_{n,i} + c\boldsymbol{I}_3\right) \boldsymbol{t} &= \boldsymbol{t}^\top \boldsymbol{\Sigma}_{n,i} \boldsymbol{t} + c\boldsymbol{t}^\top \boldsymbol{I}_3 \boldsymbol{t} \\ &= \boldsymbol{t}^\top \boldsymbol{\Sigma}_{n,i} \boldsymbol{t} + c\,, \end{aligned} \tag{3.45}$$

where $\boldsymbol{\Sigma}_{n,i} + c\boldsymbol{I}_3$ has full rank.

4. Experiments

To evaluate the PNEC, we compare it to the NEC in different experiments. We utilize synthetic data to validate its performance in frame-to-frame rotation estimation. Its performance in a visual odometry (VO) setting is evaluated on real-world data.

On the simulated data, we compare the optimization of the NEC and PNEC for two different camera types over different noise types and noise intensities in Sec. 4.1.2. Furthermore, an evaluation of the SCF for translation estimation is done in Sec. 4.1.4. An ablation study in Sec. 4.1.5 shows the effect of each step in the optimization scheme of Sec. 3.3.

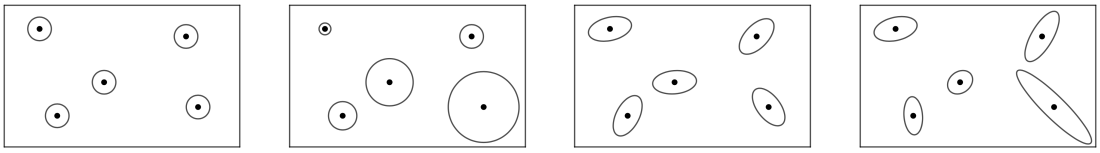
On the real-world data, the NEC and the PNEC are evaluated on the KITTI dataset [20] with regard to frame-to-frame rotation estimation and long-term drift. For the experiments, the PNEC optimization is integrated into the VO algorithm MRO by Chng *et al*. [10]. Furthermore, we replace the ORB features with KLT tracks. The results of this new VO algorithm are compared to the standard MRO as reported in [10], and MRO using the same KLT tracks instead of ORB features (see Sec. 4.2.2). We evaluate the effect of each step in the PNEC optimization scheme in an ablation study in Sec. 4.2.3. An additional experiment is done on the ICL-NUIM dataset to determine the accuracy of the covariances obtained by KLT tracking as stated in Sec. 3.4.1 (see Sec. 4.3).

Appendix A presents an overview of the parameters used in the experiments.

4.1. Simulated Experiments

The experiments on simulated data evaluate the performance of the PNEC in a frame-to-frame setting. The experiments consist of randomly generated two-view problems with known correspondences. The generation of the problems follows the outline proposed by Kneip and Lynen [32] very closely. Nevertheless, the outline is repeated in this work and the differences to the original experiments are emphasized. Since the experiments of Kneip and Lynen are only done for omnidirectional cameras but not for the widely used pinhole cameras, every experiment is repeated for pinhole cameras. The performance is evaluated using

$$\begin{aligned} e_{rot} &:= \angle(\boldsymbol{R}^\top \tilde{\boldsymbol{R}}), \text{ and} \\ e_t &:= \arccos\left(\boldsymbol{t}^\top \tilde{\boldsymbol{t}}\right) \end{aligned} \tag{4.1}$$



(a) isotropic homogeneous (b) isotropic inhomogeneous (c) anisotropic homogeneous (d) anisotropic inhomogeneous

Figure 4.1.: Illustration of different noise types based on Brooks *et al*. [6]. The shape and size between the covariances varies greatly depending on the noise types. Most experiments are repeated for all noise types to show the effects the anisotropy and inhomogeneity have on the PNEC optimization.

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.

4.1.1. Experiment outline

This section gives an overview of the setup used in the following experiments on simulated data. It starts by presenting an overview of the classification of different noise types by Brooks *et al*. [6]. We repeat most experiments for all noise types to capture the effects the different sizes and shapes of the error distributions have on the energy function and its optimization. We then describe the generation of the individual problems for omnidirectional and pinhole cameras where we closely follow the setup by Kneip and Lynen [32]. Each experiment type is repeated for both camera types, although Kneip and Lynen [32] emphasize that their eigenvalue-based solver is particularly well suited for omnidirectional cameras. Since a large portion of available cameras are pinhole cameras, we decided to evaluate the performance of the PNEC on them as well. An overview of the parameters used for the simulated experiments can be found in Appendix A.

Noise Types Each experiment is repeated for different noise types to capture their effect on the performance of the PNEC. This work uses the classification by Brooks *et al*. [6] for different noise types. Fig. 4.1 illustrates the four different noise types used in the experiments. The following gives an overview of how to generate the covariance matrices for each noise type.

The covariance matrices are generated using the following parameterization

$$\boldsymbol{\Sigma}_{2\mathrm{D}} = s\boldsymbol{R}_\alpha \begin{pmatrix} \beta & 0 \\ 0 & 1-\beta \end{pmatrix} \boldsymbol{R}_\alpha^\top \tag{4.2}$$

with a scaling factor s, an anisotropy term β, and a 2D rotation matrix

$$\boldsymbol{R}_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} . \tag{4.3}$$

The parameters for *isotropic homogeneous* noise are $s = 1$, $\beta = 0.5$, and $\alpha = 0$. For *isotropic inhomogeneous* noise they are $\beta = 0.5$, $\alpha = 0$, and s is sampled uniformly between 0.5 and 1.5 for each covariance. For *anisotropic homogeneous* noise $s = 1$, β is sampled uniformly between 0.5 and 1 once for each experiment, and α is sampled uniformly between 0 and π for each covariance. For *anisotropic inhomogeneous* noise all parameters are uniformly sampled for each covariance, s between 0.5 and 1.5, β between 0.5 and 1, and α between 0 and π.

Omnidirectional Cameras This experiment recreates the one by Kneip and Lynen [32] with some alterations. The following describes the outline and emphasizes deviations from the original.

A single two-view problem is generated randomly by fixing the position and orientation of the first camera frame to the origin and identity rotation. The second camera frame is generated relative to the first one. The translation offset is chosen from a uniform distribution of a random direction with a maximum length of 2. The orientation offset is generated with randomly generated Euler angles with a maximum magnitude of 0.5 radian. For each experiment, random points are generated around the origin. Their Euclidean distance to the origin is chosen uniformly between 4 and 8. These points are transformed into both camera frames and projected onto a sphere assuming an omnidirectional camera with a focal length of 800 pixel, giving bearing vectors for each feature. In contrast to the original experiment by Kneip and Lynen [32], we only add noise on the bearing vector of the second camera frame instead of both. To compensate for the lack of noise in the first camera frame, we scale the noise offset in the second camera frame by a factor of 2. The noise is added in the tangential plane of the bearing vector. It is sampled using the covariance matrices later used in the PNEC optimization. Unlike the original setup, this work repeats the experiment for each noise type classified by Brooks *et al.* [6] using the parameterization presented in Eq. 4.2.

Pinhole Cameras Since the experimental setup for pinhole cameras is similar to omnidirectional cameras, we only present their differences and refer the reader to the previous paragraph for more details.

The camera positions of the two-view problem are generated like for omnidirectional cameras. The points are generated such that they are in view of the pinhole cameras. This is done by generating points in the image plane of the first camera with an image

width of 1200 pixel and an image height of 800 pixel. These points are then unprojected using a random depth between 2.0 and 5.0. The unprojected points in 3D are projected back into both camera frames with a pinhole camera model with a focal length of 800 pixel. Noise is added in the image plane.

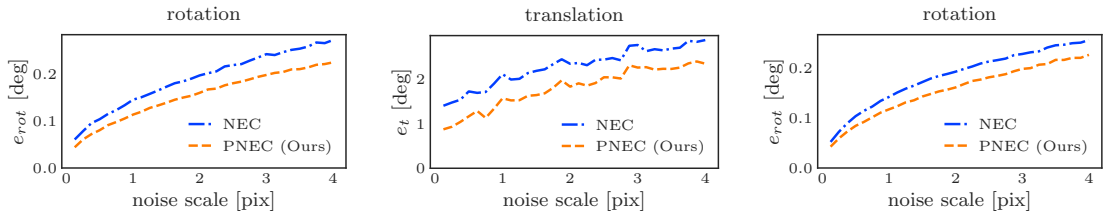
4.1.2. Noise Levels

Motivation The energy function of the PNEC Eq. 3.16 accounts for the individual shapes and sizes between the error distributions of different feature correspondences. The overall noise level does not influence the relative weighting. This experiment investigates both methods similar to the experiment by Kneip and Lynen [32, Sec. 4.4]. It evaluates whether the noise levels have a similar influence on the PNEC as on the NEC. To show the performance of the PNEC for purely rotational motion, the whole experiment repeated with the translation of the second camera fixed to the origin.

Experiment To investigate the influence of noise levels, we generate 10 000 problems for each different intensity of noise. We look at noise levels ranging from smaller than a pixel on average to more than a few pixels (we exclude no noise since a zero covariance matrix does not work with the PNEC). Both the NEC and the PNEC use the same 10 points generated for each individual problem, and the PNEC uses the covariance matrix of the error distribution used to generate the offset. The starting point for both algorithms is the same, generated randomly around the ground truth rotation. As in [33], a maximum derivation of 0.01 radian is chosen so that both methods spot the global minimum. To compare both approaches, we average the results for each noise level and present the rotational and translational error as given by Eq. 4.1. Due to its relevance for real-world data (see Fig. 3.6), we focus on the experiments with anisotropic and inhomogeneous noise. However, the results for the other noise types are also presented.

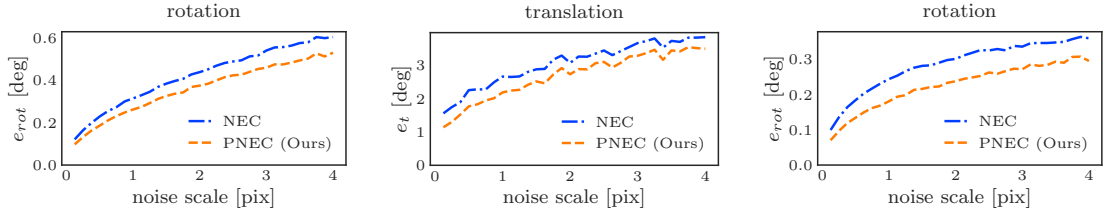
Results Fig. 4.2 present the results for the omnidirectional camera experiment for anisotropic inhomogeneous noise. It shows the average errors over 10 000 random problems for each noise level. Fig. 4.2a and Fig. 4.2b depict the rotational and translational error for the experiment as described in the outline. Fig. 4.2c shows the rotational error for the pure rotation experiment. No translational error is presented for pure rotation since the chosen cosine error metric is not defined for such a case. The PNEC achieves lower error in all cases consistently, outperforming the NEC.

Fig. 4.3 shows the results for the NEC and PNEC over different noise levels for anisotropic inhomogeneous noise for pinhole cameras. Similar to the omnidirectional



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

Figure 4.2.: Frame-to-frame rotation estimation experiments for omnidirectional cameras over different noise intensities. Results are averaged over 10,000 randomly generated problems for anisotropic inhomogeneous noise for each noise intensities. The PNEC consistently outperforms the NEC [32] for all noise levels. This holds for in the general case (Fig. 4.2a and Fig. 4.2b), respectively, as well as for pure rotation (Fig. 4.2c).



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

Figure 4.3.: Experiments for pinhole cameras repeated for different noise levels. For more details see Fig. 4.2. The PNEC consistently outperforms the NEC [32] for all noise levels. This holds for in the general case (Fig. 4.3a and Fig. 4.3b), respectively, as well as for pure rotation (Fig. 4.3c).

camera experiment, the PNEC outperforms the NEC consistently. However, the overall error is higher for pinhole cameras. A more detailed analysis reveals that the relative difference in the performance between both methods is slightly smaller for pinhole cameras for experiments with translation. It is larger for experiments without translation.

Fig. 4.4 and Fig. 4.5 depict the results for the other noise types for omnidirectional and pinhole cameras, respectively. As for anisotropic inhomogeneous noise, the PNEC consistently achieves better results than the NEC.

This experiment shows the effectiveness of the PNEC for rotation estimation. It consistently outperforms the NEC. The experiments over different noise types show the effects the directionality and the inhomogeneity of the covariances have on the optimization. Having anisotropic inhomogeneous noise is the most beneficial for PNEC

optimization. However, even for isotropic homogeneous noise, the PNEC still achieves better results than the NEC. This shows that the geometry of the problem also has an influence on the weighting of the feature correspondences. While the covariance matrices $\mathbf{\Sigma}_i$ of the error distributions are the same for each correspondence, the variance of each residual σ_i^2 is still different. Additionally, the results of this experiment support the finding of Kneip and Lynen [32] that the NEC (and to the same extent the PNEC) is constrained better for omnidirectional cameras. Another important finding of the presented experiments is that the PNEC consistently achieves excellent results for purely rotational motion, although it models a non-zero translational problem.

4.1.3. Energy-Error Correlation

Motivation Next to the results in the error metrics, we investigate the ability of the energy function to model the rotation estimation problem. This experiment evaluates how strong the correlation between the energy function and the rotational error is. Because the optimization scheme for the PNEC does not guarantee a globally optimal solution, we investigate how good the minima that the PNEC reaches are in comparison to the NEC.

Experiment For this experiment, we use the results of the previous one (see Sec. 4.1.2), but we focus on the relation between the energy function and the rotational error metrics. First, the median energy function values over the different noise levels for the NEC and the PNEC are presented. The scale of the energy function ranges over several magnitudes, making the average susceptible to single problems dominating the whole experiment. Therefore, we choose to show the median energy function values rather than the average values. We investigate the correlation between the energy function and the rotational error in two ways. First, by counting how often a lower energy function also leads to a lower rotational error. Second, by calculating the correlation coefficient for each noise type. To eliminate the scale of the energy function from the correlation coefficient we look at the correlation between $e_{rot,pnec} - e_{rot,nec}$ and the quotient of the energy functions $E_{P,pnec}/E_{P,nec}$.

Results We first present the results for omnidirectional cameras as a whole and then for pinhole cameras.

Fig. 4.6 shows the energy function values for all experiments for omnidirectional cameras. The results show that the optimization of Sec. 3.3 is effective in minimizing the energy function, outperforming the NEC constantly. Furthermore, the results show that the PNEC energy function is independent of the overall noise scale of the problem.

Noise Type	w t		w/o t	
isotropic homogeneous	61.13	38.83	53.61	45.85
	0.01	0.03	0.27	0.27
isotropic inhomogeneous	62.34	37.62	54.54	44.94
	0.01	0.03	0.27	0.26
anisotropic homogeneous	64.07	35.91	58.84	40.80
	0.00	0.02	0.19	0.17
anisotropic inhomogeneous	64.65	35.33	59.32	40.37
	0.00	0.02	0.17	0.15

Table 4.1.: Comparison of energy and error. Each cell gives the percentage of how often the PNEC has: lower energy and error (**upper left**); lower energy and higher error (**upper right**); higher energy and lower error (**lower left**); higher energy and higher error (**lower right**). The percentage is calculated over all 320 000 experiments (32 different noise scales with 10 000 experiments).

Noise Type	w t	w/o t
isotropic homogeneous	0.35	0.20
isotropic inhomogeneous	0.34	0.20
anisotropic homogeneous	0.31	0.25
anisotropic inhomogeneous	0.29	0.24

Table 4.2.: Correlation coefficient between $e_{rot,pnec} - e_{rot,nec}$ and $E_{P,pnec}/E_{P,nec}$ for omnidirectional cameras over different noise types over all 320 000 experiments (32 different noise scales with 10 000 experiments).

Tab. 4.1 gives an overview of how often the rotational error and the energy function of the PNEC are better than for the NEC. The PNEC achieves a lower energy function value than the NEC in over 99% of the problems for all experiment types. This lower energy function also leads to a lower rotational error in a majority of these cases. However, a lower energy value results in a higher rotational error in a large percentage of the problems.

Tab. 4.2 gives the average correlation coefficient of the correlation matrix between the difference in the rotational error $e_{rot,pnec} - e_{rot,nec}$ and the quotient of the energy functions $E_{P,pnec}/E_{P,nec}$. The results show a consistent correlation between the energy function and the rotational error. The correlation is higher for experiments with translation than for experiments with pure rotation for each experiment type.

The results for pinhole cameras in Fig. 4.7, Tab. 4.3, and Tab. 4.4 are similar to the results for omnidirectional cameras. The PNEC consistently achieves lower median

Noise Type	w t		w/o t	
isotropic homogeneous	59.33	40.56	57.62	42.00
	0.02	0.09	0.23	0.15
isotropic inhomogeneous	60.28	39.62	58.39	41.24
	0.01	0.08	0.21	0.16
anisotropic homogeneous	61.89	38.04	62.41	37.29
	0.01	0.06	0.20	0.10
anisotropic inhomogeneous	62.55	37.37	62.23	37.46
	0.01	0.06	0.19	0.11

Table 4.3.: Comparison of energy and error. Each cell gives the percentage of how often the PNEC has: lower energy and error (**upper left**); lower energy and higher error (**upper right**); higher energy and lower error (**lower left**); higher energy and higher error (**lower right**). The percentage is calculated over all 320 000 experiments (32 different noise scales with 10 000 experiments).

Noise Type	w t	w/o t
isotropic homogeneous	0.19	0.16
isotropic inhomogeneous	0.19	0.17
anisotropic homogeneous	0.21	0.18
anisotropic inhomogeneous	0.23	0.18

Table 4.4.: Correlation coefficient between $e_{rot,pnec} - e_{rot,nec}$ and $E_{P,pnec}/E_{P,nec}$ for pinhole cameras over different noise types over all 320 000 experiments (32 different noise scales with 10 000 experiments).

energy function values than the NEC. The count of how many problems have a lower energy value, but a greater rotational error is higher for pinhole cameras. Furthermore, the correlation coefficient is lower.

This experiment shows the effectiveness of the optimization scheme of Sec. 3.3 to minimize the PNEC energy function. Furthermore, it shows that the minimization of the PNEC energy function often leads to improved results compared to the NEC. However, it does not necessarily lead to better results. Further investigations to determine if the optimization found the global minimum are needed. However, they are out of the scope of this thesis. Nevertheless, a correlation between the energy function and the rotational error is present, although it is smaller for pinhole cameras and in cases of zero translation. Since the NEC and the PNEC are similar, this supports the findings of Kneip and Lynen [32] that the NEC is better constrained for omnidirectional cameras. The purely rotational experiments have a smaller correlation than their counterparts with translation. A difference between both experiment types is the two degrees of

freedom in the translation optimization of the energy function. The PNEC models a two-view motion estimation that is assumed to have a translational component, and therefore we optimize over the unit-sphere. The purely rotational experiments do not reflect this assumption. This gives the optimization two degrees of freedom that allow further minimization.

4.1.4. Self-Consistent-Field vs. Eigenvector

Motivation As Zhang and Chang [79] have shown, the optimization of the sum of GRQs is not trivial and proposed the SCF method for an efficient yet not optimal solution. This experiment investigates if the SCF method for optimization over the translation is beneficial for the PNEC. We compare it to a simpler eigenvalue-based solution, inspired by Kneip and Lynen [32], for the translation where the solution is given by the eigenvector to the smallest eigenvalue of the matrix $\boldsymbol{M}_P$ (see Eq. 3.24).

Experiment For this experiment, we restrict ourselves to the experiments of Sec. 4.1.2 for anisotropic inhomogeneous noise with a noise level of 1 pixel. We only look at the experiments with translation. All experiments use the ground truth rotation $\boldsymbol{R}$ but generate a random starting translation around the ground truth $\boldsymbol{t}$. The direction and the angular offset of the translation are uniformly sampled. The SCF is compared to the eigenvalue-based solution of [32] adapted to the PNEC over different levels of the translational offset in degrees. We compare the mean and median translational error of all 10 000 problems.

Results Fig. 4.8 shows the results of both methods over different starting points for omnidirectional cameras. The SCF achieves results independent of the maximum angular offset, while the performance of the eigenvalue-based method deteriorates with a larger starting position error. The SCF outperforms the eigenvalue-based method overall.

Fig. 4.9 shows that the SCF outperforms the eigenvalue-based solution for translation estimation for pinhole cameras. The gap between both methods is smaller than for omnidirectional cameras. The eigenvalue-based solution only achieves better results for starting points very close to the ground truth. Overall the SCF algorithm achieves consistent results.

This experiment shows the dependency of the eigenvalue-based solution for the translation on the starting point. In contrast, the SCF method is independent of the starting position as a consequence of its sampling strategy on the unit sphere leading to consistent results. Overall the SCF outperforms the simpler eigenvalue-based solution.

Noise level [px]	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
NEC	0.103	0.144	0.172	0.196	0.218	0.241	0.252	0.271
NEC-LS	0.092	0.130	0.158	0.185	0.208	0.231	0.244	0.260
PNEC only LS	0.076	0.108	0.132	0.154	0.173	0.191	0.204	0.217
PNEC w/o LS	0.086	0.120	0.145	0.163	0.186	0.204	0.214	0.231
PNEC	0.080	0.113	0.138	0.159	0.179	0.197	0.209	0.223

Table 4.5.: Ablation study for omnidirectional cameras. Comparison between the NEC, the NEC with a least-squares refinement of Eq. 3.6, just the least-squares refinement of the PNEC, just the iterative optimization of the PNEC and the full PNEC.

4.1.5. Ablation

Motivation This experiment aims to investigate the influence of each component of the optimization scheme of Sec. 3.3 for the rotation estimation on the simulated data. Additionally, the optimization is compared to the NEC and a least-squares refinement of the NEC energy function Eq. 3.6 to evaluate whether a similar two-step approach for the NEC yields better results.

Experiment This experiment compares: the NEC optimization; only the iterative PNEC optimizer (PNEC w/o LS); only the least-squares refinement (PNEC only LS); the full PNEC optimization. Additionally, a least-squares refinement of the NEC (NEC-LS) energy function Eq. 3.6 (with the NEC result as a starting value) is evaluated to compare it to the full PNEC optimization. We restrict this experiment to the anisotropic inhomogeneous noise type over selected values of the noise level. Each method starts with the same initialization, chosen as in Sec. 4.1.2. We compare the average rotational error over all 10 000 problems for each noise level.

Results Tab. 4.5 shows the results for omnidirectional cameras. The PNEC only LS implementation consistently has the best results while the full PNEC implementation is only slightly worse. The PNEC w/o least-squares refinement outperforms the NEC. The full PNEC rotation estimation outperforms the NEC with least-squares refinement.

The ablation study for pinhole cameras in Tab. 4.6 shows the same ordering as for omnidirectional cameras. The least-squares only approach achieves the best results on the simulated experiments.

The ablation study shows that the full optimization approach for the PNEC is not the best for the simulated experiment setup. It is consistently outperformed by the least-squares only approach. Given the experimental setup with an initialization near the ground truth, this is of little surprise since the least-squares refinement, as proposed

Noise level [px]	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
NEC	0.229	0.314	0.384	0.438	0.488	0.540	0.575	0.603
NEC-LS	0.218	0.309	0.377	0.437	0.491	0.540	0.578	0.617
PNEC only LS	0.178	0.252	0.313	0.362	0.406	0.439	0.472	0.505
PNEC w/o LS	0.201	0.273	0.337	0.382	0.434	0.472	0.505	0.539
PNEC	0.186	0.262	0.324	0.374	0.423	0.457	0.493	0.527

Table 4.6.: Ablation study for pinhole cameras. Comparison between the NEC, the NEC with a least-squares refinement of Eq. 3.6, just the least-squares refinement of the PNEC, just the iterative optimization of the PNEC and the full PNEC.

in Sec. 3.3.4, is highly dependent on the starting values. Additionally, the ablation study shows that the eigenvalue-based optimization of the PNEC outperforms the NEC optimization and the least-squares refinement of the NEC has little impact.

4.2. Odometry Datasets

Besides the simulated experiments, the PNEC optimization of Sec. 3.3 is also evaluated on real-world data, namely the highly popular KITTI odometry dataset [20], in a visual odometry (VO) setting. The PNEC approach is compared to the MRO algorithm by Chng *et al*. [10], which uses the NEC optimization [32] for its baseline. To have a fair comparison, we compare the NEC and the PNEC in the baseline version. Both include neither rotation averaging nor loop closure. Since the results of MRO reported in [10] could not be replicated reliably, with the algorithm often terminating early due to a lack of feature matches, this work uses the results reported in [10]. An overview of the parameters used for the experiments on KITTI can be found in Appendix A.

4.2.1. The KITTI Odometry Dataset

The highly popular KITTI odometry dataset [20] consists of 11 sequences of varying lengths. The sequences were recorded in and around the German city of Karlsruhe with a car. Therefore, the setting of the dataset is an urban to a rural environment, and forward movement is dominant. The results presented in this work are generated from monocular camera images and compared to the ground truth. The trajectories presented in Fig. 4.10 and Fig. 4.11 are generated with the estimated frame-to-frame rotations and the ground truth frame-to-frame translations.

We use the definition of the rotation-only version of the *Relative Pose Error* (*RPE*) for n camera poses as proposed in [10] to compare the methods on the KITTI dataset. The

RPE evaluates the root-mean-square error ($RMSE$)

$$RMSE(\Delta) := \left(\frac{1}{m} \sum_{i=1}^{m} E_i^2 \right)^{\frac{1}{2}} \tag{4.4}$$

over $m := n - \Delta$ residuals for frame pairs that are a "time-step" Δ apart. For the rotation-only version of the RPE the residual

$$E_i := \angle((\boldsymbol{R}_i^\top \boldsymbol{R}_{i+\Delta})^\top (\tilde{\boldsymbol{R}}_i^\top \tilde{\boldsymbol{R}}_{i+\Delta})) \tag{4.5}$$

is given by the angular error between the ground truth ($\boldsymbol{R}_i^\top \boldsymbol{R}_{i+\Delta}$) and the estimated ($\tilde{\boldsymbol{R}}_i^\top \tilde{\boldsymbol{R}}_{i+\Delta}$) relative rotations of a frame pair. To capture the local frame-to-frame rotation error, the RPE_1

$$RPE_1 := RMSE(1) \tag{4.6}$$

is evaluated. However, the RPE_1 cannot measure long-term drift important to VO systems. To measure these long-term drifts the RPE_n

$$RPE_n := \frac{1}{n} \sum_{\Delta=1}^{n} RMSE(\Delta) \tag{4.7}$$

is also evaluated.

4.2.2. The Baseline Method

Motivation We evaluate the performance of the PNEC for frame-to-frame rotation estimation in a visual odometry setting. This experiment compares the PNEC to the NEC on all KITTI sequences. To have a fair comparison of the NEC and the PNEC we test the PNEC in the same visual odometry framework as the NEC. We choose the MRO algorithm [10] since it already uses the eigenvalue-based NEC optimization for its frame-to-frame rotation estimation.

Experiment We incorporate the PNEC optimization of Sec. 3.3 into MRO. For this integration, we have to change two things.

First, the ORB feature extraction and matching used in MRO is replaced by the KLT-based tracking implementation of [71]. The KLT tracks give the covariance information used in the PNEC. The extraction of the uncertainty is detailed in Sec. 3.4.1. Second, the NEC optimization of [32] is replaced by the PNEC optimization of Sec. 3.3. To address the influences the KLT tracks have on the rotation estimation performance, this experiment compares: MRO, NEC optimization with ORB features; KLT-NEC,

NEC optimization with KLT tracks; KLT-PNEC, PNEC optimization with KLT tracks. Furthermore, the RANSAC [16] routine of MRO is used to filter failed KLT tracks that result in outliers. The same tracks are used for KLT-NEC and KLT-PNEC. The rotation averaging and the loop closure proposed in [10] are deactivated to focus on rotation estimation. Since we were not able to reliably reproduce the results reported in [10] we present their results for MRO. However, Fig. 4.10 shows a qualitative trajectory for seq. 08 of the KITTI dataset with results obtained by us that has similar metrics as reported in the paper. The trajectory from the PNEC optimization is presented as a comparison.

We compare all methods on the RPE_1 and the RPE_n metric on a single run of all sequences.

Results Tab. 4.7 presents the results for all KITTI sequences. Except for seq. 01 the KLT-based methods outperform ORB-based MRO significantly. A closer investigation of this sequence finds that the KLT implementation of [71] fails and produces many wrong tracks due to self-similar structure. Since neither tracks nor covariances are correct, we omit this sequence in the following evaluation and the ablation study. Comparing KLT-NEC and KLT-PNEC shows that optimizing the PNEC leads to better results in both metrics on almost all sequences. Especially the RPE_n metric for long-term drift is improved significantly.

The results show the benefits of using the PNEC instead of the NEC for real-world tasks of VO systems. The PNEC reduces the frame-to-frame error noticeably on most sequences. Furthermore, it leads to significantly reduced drift. This low drift is important in VO system and allows for accurate positional tracking over long periods of time, even without additional techniques like loop closure.

4.2.3. Ablation Study

Motivation As for the simulated experiments, an ablation study on real-world data gives additional insight into the optimization scheme of Sec. 3.3. This experiment aims to investigate the influence of each component of the optimization for the rotation estimation.

Experiment The ablation study is done on all sequences of the KITTI dataset, except on seq. 01 due to poor quality of the KLT tracks. As for the ablation study on the simulated data (see Sec. 4.1.5) we compare: the NEC optimization; PNEC w/o LS; PNEC only LS; PNEC. Additionally, the least-squares refinement of the NEC (NEC-LS) energy function Eq. 3.6 is evaluated to compare it to the full PNEC optimization. For a more detailed explanation of the methods we refer the reader to Sec. 4.1.5. All methods

Seq.	MRO [10] RPE_1	MRO [10] RPE_n	KLT-NEC RPE_1	KLT-NEC RPE_n	KLT-PNEC RPE_1	KLT-PNEC RPE_n
00	0.36	8.67	0.127	4.935	**0.121**	**4.706**
01*	**0.29**	**16.03**	0.692	25.548	0.853	27.783
02	0.29	16.03	**0.087**	**5.876**	0.101	6.010
03	0.28	5.47	**0.056**	2.453	0.060	**1.410**
04	0.04	1.08	0.042	0.792	**0.038**	**0.531**
05	0.25	11.36	0.085	4.641	**0.056**	**2.746**
06	0.18	4.72	0.144	4.443	**0.081**	**2.967**
07	0.28	7.49	0.074	5.207	**0.070**	**2.149**
08	0.27	9.21	0.063	5.593	**0.056**	**2.909**
09	0.28	9.85	0.104	**3.526**	**0.081**	3.866
10	0.38	13.25	0.086	5.094	**0.071**	**4.012**

Table 4.7.: Quantitative comparison for KITTI on the RPE_1 and RPE_n metrics. Best results are bold, second-best results are underlined. Changing from ORB features to KLT tracks improves the results of the NEC as the significant gap between MRO and KLT-NEC shows. The results using KLT tracks are further improved by applying the PNEC optimization. The PNEC has the best results on almost all sequences. (* In seq. 01 the KLT implementation of [71] fails and produces many wrong tracks due to self-similar structure. Since neither tracks nor covariances are correct, this sequence is omitted in the ablation study.)

use the same KLT tracks and are initialized with the relative pose obtained by the previous frame-to-frame rotation estimation.

Results Tab. 4.8 shows the results on each sequence as well as the average. Similar to the simulated experiments PNEC only LS almost always has the best performance of the methods. However, due to poor results on Seq. 05, it has the worst average. Without the PNEC only LS method the full PNEC optimization performs the best on most sequences. Overall the full PNEC optimization leads to the best performance on average.

This experiment shows the need for the full PNEC optimization scheme. While the least-squares only approach often performs slightly better, it is prone to large errors. Fig. 4.11 gives a qualitative trajectory of PNEC only LS on seq. 05 where it has the worst performance by a wide margin. PNEC only LS achieves good performance for large parts of the sequence. However, the sensitivity of the least-squares refinement leads to large errors in the first curve. This is of interest regarding the simulated experiments, where the least-squares only approach consistently achieves the best results.

Seq.	NEC RPE_1	NEC RPE_n	NEC-LS RPE_1	NEC-LS RPE_n	PNEC w/o LS RPE_1	PNEC w/o LS RPE_n	PNEC only LS RPE_1	PNEC only LS RPE_n	PNEC RPE_1	PNEC RPE_n
00	0.127	4.935	0.122	5.948	0.119	9.155	**0.117**	**3.706**	0.121	4.706
02	0.087	5.876	0.078	5.530	0.080	7.013	**0.077**	**4.034**	0.101	6.010
03	0.056	2.453	0.071	2.703	0.066	1.930	**0.054**	1.449	0.060	**1.410**
04	0.042	0.792	0.042	0.560	**0.038**	0.933	**0.038**	**0.423**	**0.038**	0.531
05	0.085	4.641	0.058	4.046	0.061	3.949	0.806	29.520	**0.056**	**2.746**
06	0.144	4.443	0.083	2.753	0.062	**1.559**	**0.058**	3.220	0.081	2.967
07	0.074	5.207	0.080	5.520	0.091	4.069	**0.068**	3.988	0.070	**2.149**
08	0.063	5.593	0.060	5.202	0.063	5.287	**0.056**	**2.889**	**0.056**	2.909
09	0.104	3.526	0.099	5.508	0.079	2.770	**0.058**	**2.062**	0.081	3.866
10	0.086	5.094	0.067	**3.512**	0.083	4.211	**0.063**	3.941	0.071	4.012
Average	0.087	4.128	0.076	4.256	0.074	4.088	0.140	5.523	**0.073**	**3.131**

Table 4.8.: Ablation study. RPE_1 and RPE_n for all KITTI sequences and the average. The best results are bold, second-best results are underlined. Results reveal that often the least-squares only PNEC optimization achieves the best results. However, on seq. 05 PNEC only LS performs drastically worse than every other method. Fig. 4.11 shows its qualitative trajectory illustrating the bad performance. On average, the full PNEC optimization has the best results while the iterative eigenvalue-based part has the second-best. This is in contrast to simulated experiments, where the PNEC only LS optimization has the best results. For KITTI, it performs the worst out of all methods.

4.2.4. Runtime

Motivation VO systems are often employed in real-world applications and need to run in real-time. Therefore, we investigate not only the rotation estimation performance of the PNEC optimization but also its runtime. This experiment breaks down the runtime of the different steps for the rotation estimation and compares it to the NEC optimization.

Experiment To evaluate the runtime MRO, KLT-NEC, and KLT-PNEC (as explained in Sec. 4.2.2) are timed on every sequence of the KITTI dataset (except seq. 01). The average runtime for the frame-to-frame rotation estimation is calculated. We break down the different steps of the rotation estimation. The *feature creation* includes the ORB feature extraction for MRO and for the KLT-based methods the track generation in the host-frame and tracking into the target frame. MRO also times the *feature matching* that is not needed for KLT. Finally, the optimization is timed. For MRO the same configuration as for their demo is used.

	MRO [10]	KLT-NEC	KLT-PNEC
feature creation	36	23	23
matching	120		
optimization	5	33	54
total time	161	56	77

Table 4.9.: Average frame processing time in milliseconds. MRO needs additional feature matching, which takes the largest amount of time. KLT-PNEC is slightly slower than KLT-NEC, but achieves real-time performance on KITTI.

Results Tab. 4.9 shows the breakdown of the runtime as performed on a laptop with a 2.4 GHz Quad-Core Intel Core i5 processor and 8 GB of memory. Results show the significant benefit of using KLT tracks instead of ORB features. Both KLT-based methods show a lower feature creation time. Additionally, no time is used for matching, further reducing the runtime. For the optimization, the ORB-based MRO is the fastest, whereas KLT-PNEC is the slowest. Overall both KLT-NEC and KLT-PNEC achieve real-time performance on the KITTI dataset.

As expected the PNEC optimization, which includes more optimization steps, is slower than the NEC if both use KLT tracks. However, the PNEC optimization is fast enough such that it runs on real-time on the KITTI dataset. Additionally, the usage of KLT tracks is beneficial to the runtime, especially with regard to additional techniques like rotation averaging. The addition of further connections into the covisibility graph does not require additional ORB feature matching.

4.3. Covariance Extraction

Motivation This experiment aims to validate whether the covariances extracted from the KLT tracks accurately reflect the error distribution exhibited by the feature correspondences. Sec. 3.4.1 details how we extract the uncertainty information in form of covariance matrices from the KLT tracking energy function. We perform this experiment on the ICL-NUIM dataset [23] because it allows for accurate determination of the ground truth feature positions. The ICL-NUIM dataset is a synthetic dataset that allows the accurate determination of feature positions on the image and in 3D. In order to use a more realistic setting for the KLT tracker, we use the ICL-NUIM RGB images with noise, but the noise-free depth images since we are interested in a very accurate determination of the ground truth. An overview of the parameters used for this experiment can be found in Appendix A.

Experiment In this experiment, we do frame-to-frame tracking with KLT tracks and compare the tracked position to the ground truth with regard to the extracted covariance matrix. We use the same KLT tracking implementation as for the KITTI odometry dataset. The ground truth is determined using the following projection from the host frame into the target frame [13]

$$\boldsymbol{p}_{gt} = \Pi_c(\boldsymbol{R}_{gt}^{-1}(\Pi_c^{-1}(\boldsymbol{p}, d_p) - \boldsymbol{t}_{gt})) \tag{4.8}$$

where Π_c is the camera projection function, Π_c^{-1} the unprojection function using the ground truth depth from the depth image, and $\boldsymbol{R}_{gt}, \boldsymbol{t}_{gt}$ describe the ground truth relative pose between the two images. To validate the ground truth position in the target frame, the image points are back-projected into the host frame. We discard points that do not land in the vicinity of their starting point. This validation removes errors due to occlusion or edges.

In this experiment we look at the error distribution of the KLT tracks. In order to evaluate the covariance matrices extracted from the KLT formulation we look at the unaltered error

$$\boldsymbol{e}_i = \boldsymbol{p}_{i,gt} - \boldsymbol{p}_{i,klt}\,, \tag{4.9}$$

between the ground truth position $\boldsymbol{p}_{i,gt}$ and the tracked position $\boldsymbol{p}_{i,klt}$, and the error

$$\boldsymbol{e}_{i,w} = \boldsymbol{C}_i(\boldsymbol{p}_{i,gt} - \boldsymbol{p}_{i,klt}) \tag{4.10}$$

after applying a whitening transformation to it, where $\boldsymbol{C}_i\boldsymbol{C}_i^\top = \boldsymbol{\Sigma}_i^{-1}$ is obtained with the Cholesky decomposition.

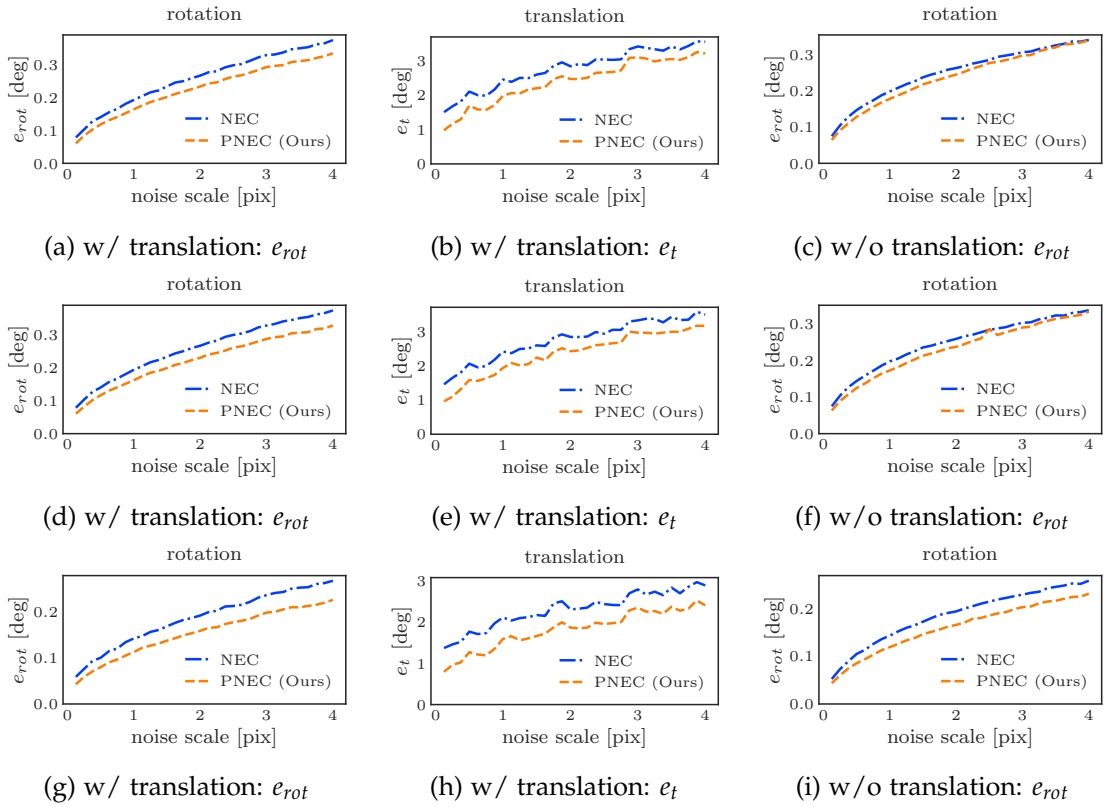
We compare both error distributions to the distribution of 2 independent Gaussians. If the covariance correctly accounts for the error distribution, $\boldsymbol{e}_w$ should follow the Gaussian distribution very closely. To validate this we look at the squared sum distribution $\|\boldsymbol{e}_w\|^2$ and the angular error distribution $\angle\boldsymbol{e}_w$. For the 2 independent Gaussians this results in the χ_2^2 distribution [60], given by

$$\chi_2^2 \sim \frac{1}{2}e^{-x/2}\,, \tag{4.11}$$

for the squared sum and a uniform angular distribution. Since the covariance matrices do not account for the overall scale of the error distribution, we scale both error distributions such that the median of $\|\boldsymbol{e}\|^2$ and $\|\boldsymbol{e}_w\|^2$ is equivalent to the median of the χ_2^2 distribution.

Results The distributions shown in Fig. 4.12 and Fig. 4.13 are generated from the living room 3 and the office room 0 sequence of the ICL-NUIM dataset, respectively. The distributions for the other sequences are depicted in Appendix B. Fig. 4.12a and Fig. 4.13a show the squared sum distributions before and after applying the whitening transformation as well as the χ^2_2 distribution on a logarithmic and a linear axis. $\boldsymbol{e}$ and $\boldsymbol{e}_w$ exhibit disproportionately more errors with a small norm compared to the χ^2_2 distribution. Consequentially, they have disproportionately few errors that are large. Applying the whitening transformation has little impact on the distribution of the squared sum. Fig. 4.12b and Fig. 4.13b show the angular distribution. While the angular error of $\boldsymbol{e}$ on the living room 3 sequence shows some anisotropy, it is significantly larger the office room 0 sequence. Applying the whitening transformation results in little anisotropy for $\boldsymbol{e}_w$ on both sequences. On the office room 0 sequence, a significant amount of anisotropy is removed.

This experiment evaluates the modeling of the tracking error distribution by the extracted covariance matrices. Comparing the squared sum of $\boldsymbol{e}$ and $\boldsymbol{e}_w$ shows that the covariances do not capture the scale of the error correctly. Fig. 4.14 and Fig. 4.15 show examples for the tracking of two sequences. They show that, although the tracking error in low-textured areas is small, the uncertainty is estimated to be large. Therefore, the error is often overestimated, resulting in a disproportionate distribution in the squared sum. While the covariances do not correctly account for the scale, they account for the direction of the error distribution. The dataset has many edges that are horizontally or vertically placed in the image. The features located on these edges are expected to have an error mostly along these edges, resulting in the large anisotropy for the unaltered error (see Fig. 4.14b). The covariances account for this bias in the direction of the edge. The results on the remaining sequence show similar behavior, and therefore they are not discussed in detail. They are presented in Appendix B. Overall, this experiment shows that not all errors follow the same distribution. The covariances capture the anisotropy of the error distributions for the ICL-NUIM dataset.

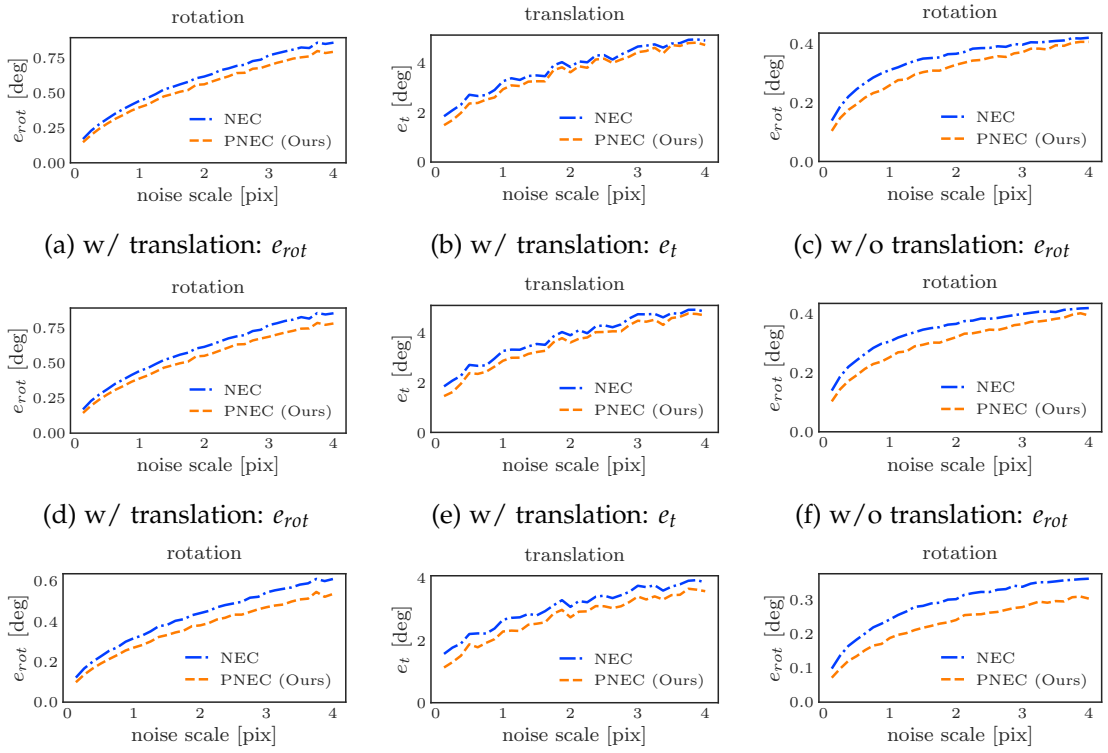


(a) w/ translation: e_{rot} (b) w/ translation: e_t (c) w/o translation: e_{rot}

(d) w/ translation: e_{rot} (e) w/ translation: e_t (f) w/o translation: e_{rot}

(g) w/ translation: e_{rot} (h) w/ translation: e_t (i) w/o translation: e_{rot}

Figure 4.4.: Experiments for omnidirectional cameras over different noise types and noise intensities. The results are shown for *isotropic homogeneous*, *isotropic inhomogeneous*, and *anisotropic homogeneous* noise in descending order. Both the general case and the purely rotational case are shown for each noise type. As for anisotropic inhomogeneous noise the results are averaged over 10 000 random problems. The PNEC performs better for all noise types. The gap size between the PNEC and NEC is dependent on the noise type showing the effects the directionality and the inhomogeneity of the covariances have.



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

(d) w/ translation: e_{rot} (e) w/ translation: e_t (f) w/o translation: e_{rot}

(g) w/ translation: e_{rot} (h) w/ translation: e_t (i) w/o translation: e_{rot}

Figure 4.5.: Experiments for pinhole cameras over different noise types and noise intensities. The results are shown for *isotropic homogeneous*, *isotropic inhomogeneous*, and *anisotropic homogeneous* noise in descending order. Both the general case and the purely rotational case are shown for each noise type. The results show a similar effect of the directionality and the inhomogeneity of the covariances as for omnidirectional cameras (see Fig. 4.4 for more details).

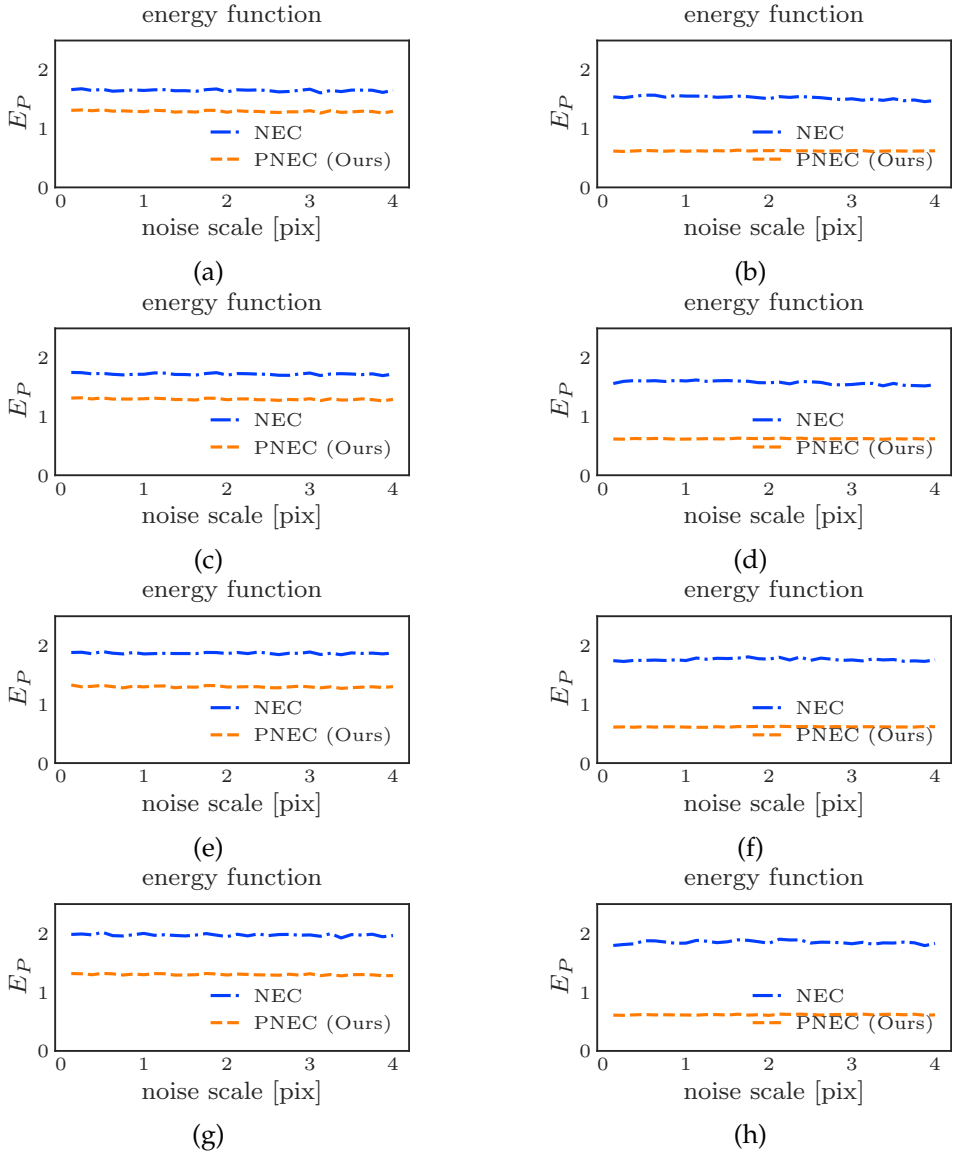


Figure 4.6.: Energy function values for all simulated experiments for omnidirectional cameras. Due to the volatility of the energy function the median instead of the mean is presented. The first column shows the experiments with translation, the second column the pure rotation experiments. The rows show *isotropic homogeneous*, *isotropic inhomogeneous*, *anisotropic homogeneous*, and *anisotropic inhomogeneous* in descending order. The PNEC is effective in achieving lower energy values for all experiments. Because of the inclusion of the covariance matrices the order of magnitude of the energy function is independent of the noise intensity.

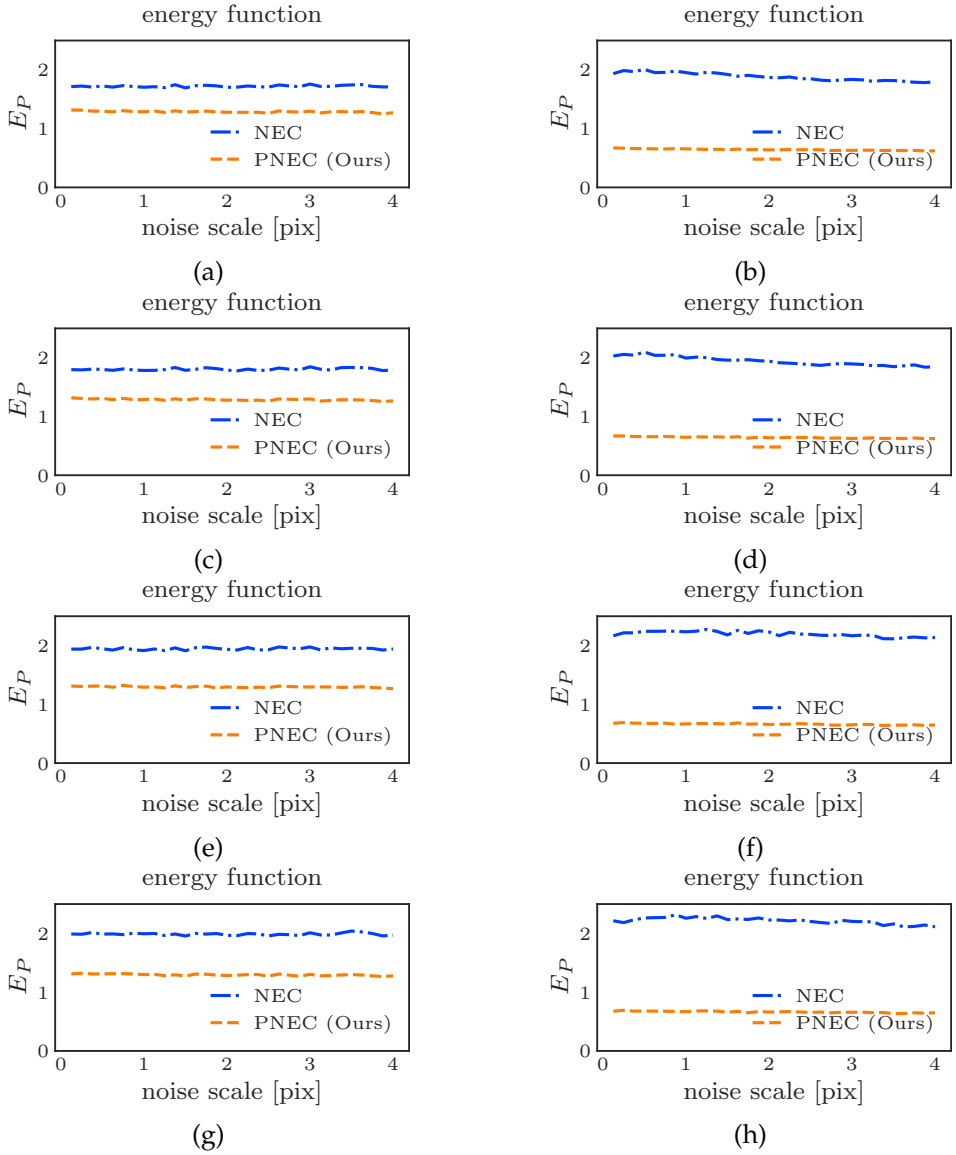
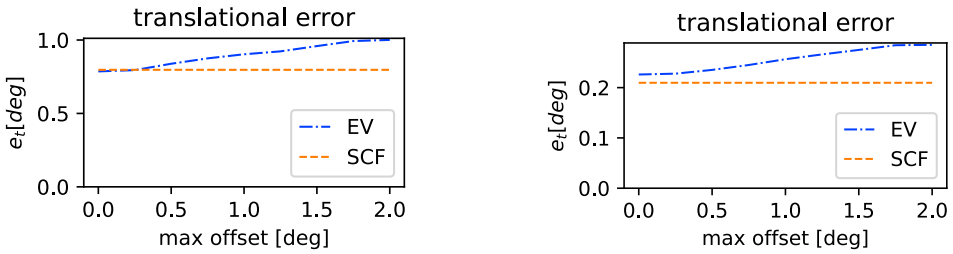
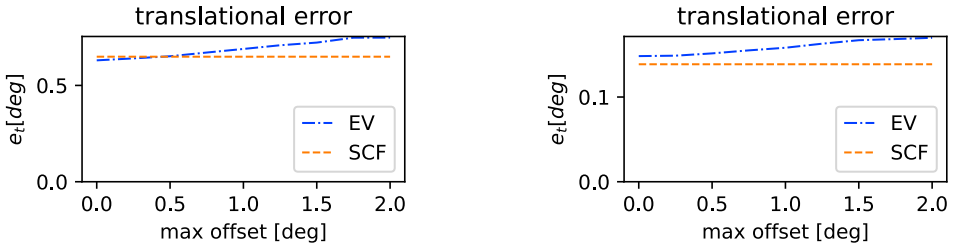


Figure 4.7.: Energy function values for all simulated experiments for pinhole cameras. Due to the volatility of the energy function, the median instead of the mean is presented. The first column shows the experiments with translation, the second column the pure rotation experiments. The rows show *isotropic homogeneous*, *isotropic inhomogeneous*, *anisotropic homogeneous*, and *anisotropic inhomogeneous* in descending order. The PNEC is effective in achieving lower energy values. Because of the inclusion of the covariance matrices, the order of magnitude of the energy function is independent of the noise intensity.



(a) Mean translational error (b) Median translational error

Figure 4.8.: The mean and median translational errors of the SCF algorithm and the eigenvalue-based (EV) translation estimation over different intensities of starting value error for omnidirectional cameras. The SCF algorithm has a consistent error over the offset while the EV estimation performance decreases with a higher starting error. The SCF outperforms the EV method consistently.



(a) Mean translational error (b) Median translational error

Figure 4.9.: The mean and median translational errors of the SCF algorithm and the eigenvalue-based (EV) translation estimation over different intensities of starting value error for pinhole cameras. The SCF algorithm has a consistent error over the offset while the EV estimation performance decreases with a higher starting error. The SCF outperforms the EV method consistently.

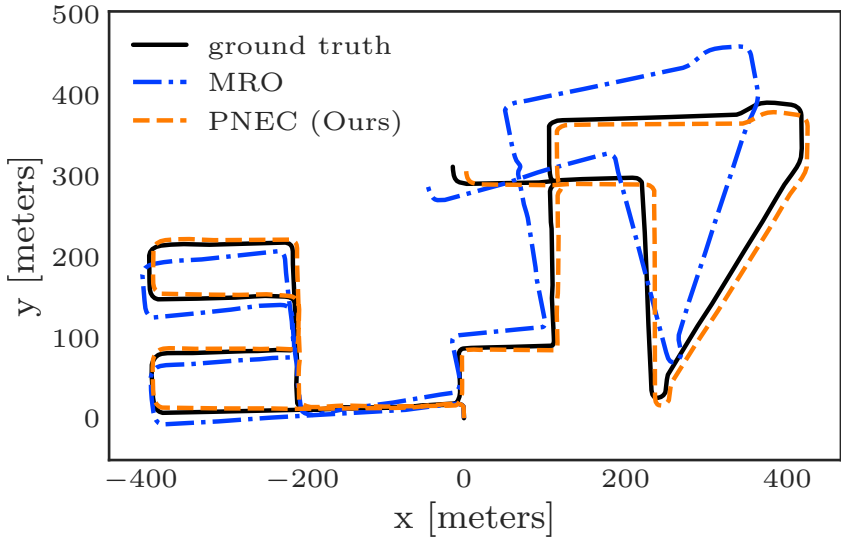


Figure 4.10.: Qualitative trajectory comparison for KITTI seq. 08. The trajectory is generated with the estimated rotations of MRO [10] and the PNEC optimization with KLT tracks, respectively, and are combined with the ground truth translations for visualization purposes. Computing relative rotations with the PNEC leads to a significantly reduced drift.

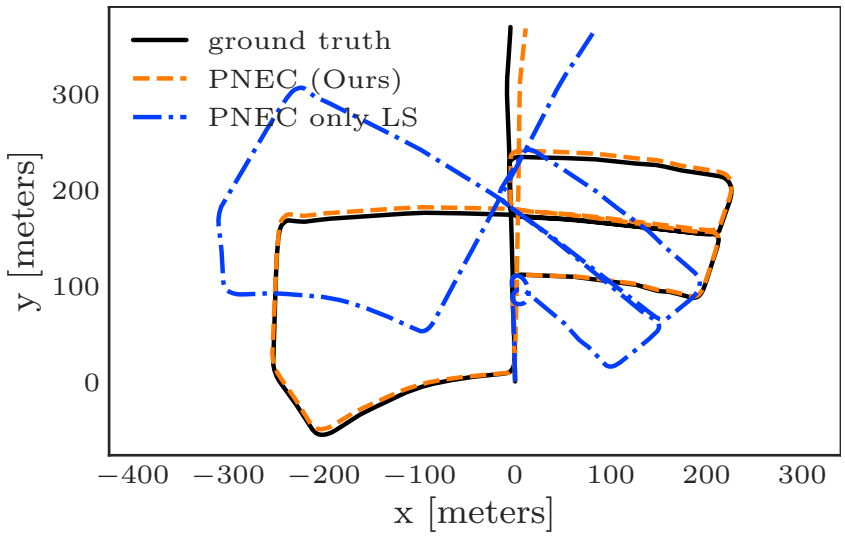
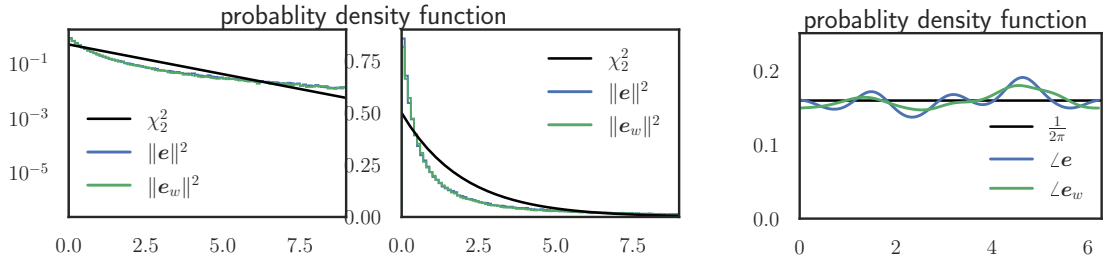


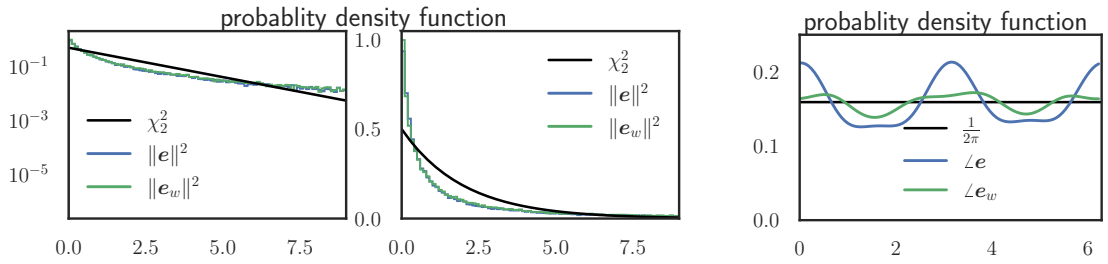
Figure 4.11.: Qualitative trajectory of the full PNEC optimization and only the least sqaures refinement optimization (PNEC only LS) for KITTI seq. 05. The trajectory shows that the rotation estimation using only least-squares performs excellent on large parts of the trajectory. However, wrong estimates near the first corner result in a large drift overall. This illustrates the necessity of the full PNEC optimization.



(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis

(b) smoothed $\angle\cdot$ distribution

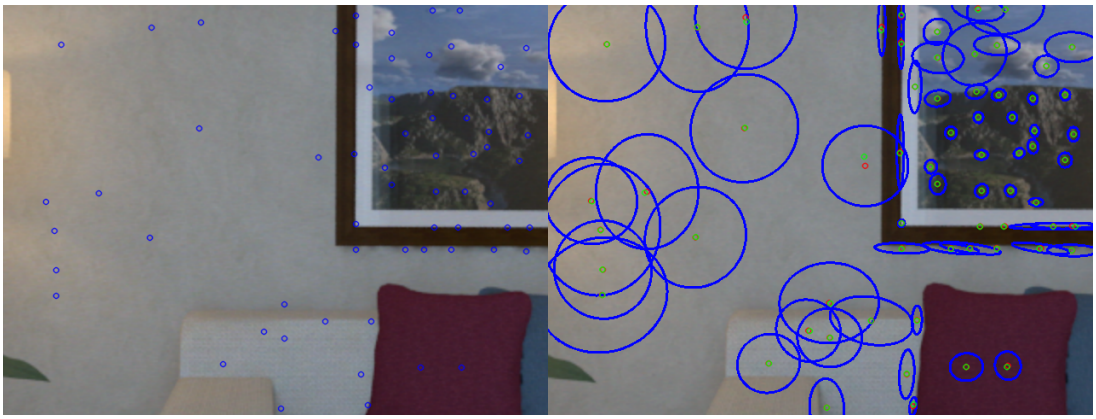
Figure 4.12.: Error distribution of the KLT tracking on the ICL-NUIM sequence living room 3. The error distributions before (e) and after (e_w) applying the whitening transformation with the extracted covariance matrices are shown. Fig. 4.12a shows the squared sum of the error distributions and the χ^2_2 distribution with a logarithmic and linear y-axis. Fig. 4.12b shows the angular error distributions. Applying the whitening transform changes the squared sum distribution only negligible. Both distributions have a disproportionately higher amount of small errors compared to the χ^2_2 distribution. e and e_w exhibit anisotropy in the angular distribution. The anisotropy is slightly smaller after the whitening transformation.



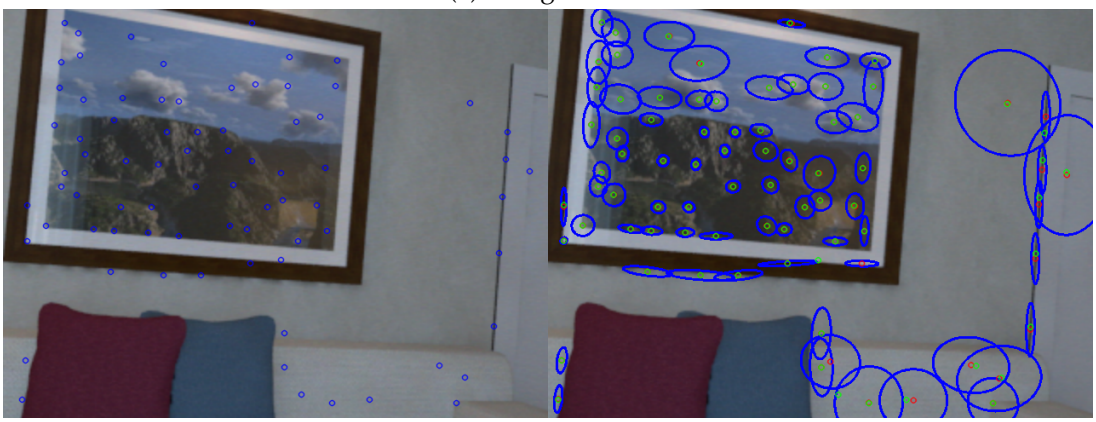
(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis

(b) smoothed $\angle\cdot$ distribution

Figure 4.13.: Error distribution of the KLT tracking on the ICL-NUIM sequence office room 0. For more details, see Fig. 4.12. Applying the whitening transform changes the squared sum distribution only negligible. Both distributions have a disproportionately higher amount of small errors compared to the χ^2_2 distribution. $\boldsymbol{e}_w$ shows significantly less anisotropy in the angular distribution than $\boldsymbol{e}$. The covariances correctly account for the direction of the error distribution.

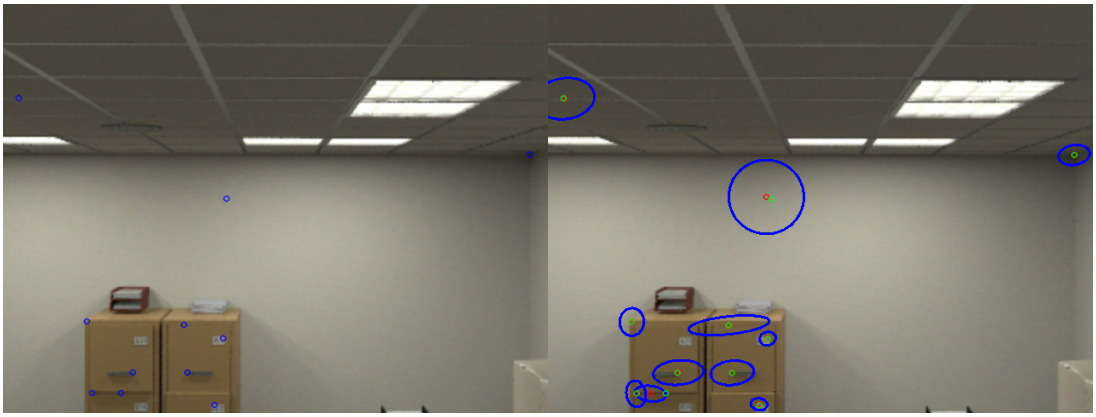


(a) living room 3



(b) living room 3

Figure 4.14.: Example tracks on the sequence living room 3. For visualization purposes, only a section of the images is shown. On the left: positions of the tracks in the host frame (blue). On the right: tracked positions (red), ground truth (green), and the extract covariance (blue, enlarged for visualization purposes). The KLT tracker result in good tracks even in low textured areas leading to an overestimation of the error there. Examples in the second image Fig. 4.14b show that the anisotropy of the error on the door is accurately represented as the error tends to be parallel to the edge. Accounting for the bias removes anisotropy of the error distribution.



(a) office room 0

(b) office room 0

Figure 4.15.: Example tracks on the sequence office room 0. For visualization purposes, only a section of the images is shown. On the left: positions of the tracks in the host frame (blue). On the right: tracked positions (red), ground truth (green), and the extract covariance (blue, enlarged for visualization purposes). The KLT tracker results in good tracks even in low textured areas leading to an overestimation of the error there.

5. Discussion and Future Work

This work introduces the novel probabilistic normal epipolar constraint (PNEC) and energy function for estimating rotation between two frames, even for purely rotational motion, and investigates its properties. The focus of this thesis is the derivation of the PNEC and its optimization for rotation estimation. The experiments of Ch. 4 show the effectiveness of the PNEC to estimate the rotation on simulated data and in combination with KLT-based tracking on real-world data. Based on the results *future work* can be divided into two directions. The first direction focuses on the optimization of the PNEC energy function with the main focus of avoiding suboptimal minima. The second direction addresses the distribution of the feature positions and how to derive accurate covariance matrices.

Optimizing the probabilistic normal epipolar constraint. While the optimization scheme of Sec. 3.3 is effective in minimizing the energy function Eq. 3.16, its two-step approach is complex. Furthermore, the ablation study on the simulated data shows, the least-squares only refinement is superior. The experiments on real-world data, however, show the necessity for a good initialization of the least-squares only approach. We provide this initialization with the eigenvalue-based optimization. The proposed optimization is effective for rotation estimation, but it is not suited to reliably find the global optimum of the PNEC energy function. Furthermore, the investigations into the energy function on the simulated experiments show that the optimization is prone to local minima that seldomly lead to worse results.

In contrast to the NEC, the PNEC is not capable of decoupling the rotation estimation from the translation leading to a more involved optimization strategy. While the rotation optimization of Sec. 3.3.3 keeps the elegance of eigenvalue-based optimization of the NEC [32], alternating it with the SCF algorithm for optimizing over the translation leads to a computationally more expensive iterative optimization. While the runtime study (see Tab. 4.9) shows that the PNEC is real-time capable, its optimization is still slower than for the NEC.

Simplifying the PNEC optimization is a promising direction for future work. However, the optimization over the translation as a sum of GRQs on its own is an actively studied problem for which no simple solution is known. Therefore, finding an elegant optimization strategy for Eq. 3.16 could be difficult. One approach to streamline the

optimization is to approximate the variances of Eq. 3.15 to remove their dependence on the translation and/or the rotation. If a successful approximation is found, a solver similar to the eigenvalue-based optimization of the NEC [32] can be employed as an initialization scheme for the least-squares refinement. A non-iterative initialization approach would speed up the rotation estimation.

Uncertainty of feature positions. While this work gives an example of how to extract the needed covariance information, we do no in-depth analysis on how to estimate good covariances for the PNEC. The presented covariance extraction, together with the PNEC, results in substantial improvements on the KITTI dataset. However, the experiments in Sec. 4.3 show this uncertainty extraction strategy to be suboptimal. Covariances that more closely resemble the error distribution should lead to further improvements in the results. To this end, deep learning could be employed to boost the performance of the PNEC. Recent works have shown the benefits of deep learning algorithms in visual odometry [73, 74, 22]. To a similar extent, deep learning algorithms could be trained to estimate the uncertainty in a feature position.

This work only addresses the probabilistic normal epipolar constraint that follows a tracking-based approach, having a perfectly localized feature in the host-frame and uncertainty in the target frame. A similar version of the PNEC can be derived analogously for feature position uncertainty in the host frame. However, neither of these approaches is suited for feature extraction and matching like ORB. In this scenario, one needs to consider uncertainty information in the host and the target-frame leading to a new probabilistic normal epipolar constraint. While the derived energy function would be more complex than the one presented in this thesis, this would expand the utilization of the PNEC to most feature correspondence finding algorithms.

6. Conclusions

This thesis presents a novel constraint on the rotation between two camera views. The presented probabilistic normal epipolar constraint (PNEC) allows the integration of 2D feature position uncertainty into the rotation estimation independent of the translation, improving upon the normal epipolar constraint (NEC). We present a derivation of the PNEC energy function together with a two-step optimization scheme that achieves real-time performance on real-world data. Furthermore, we investigate the properties of the PNEC with regard to singularities and present a regularization that removes them. We evaluate the performance of the proposed optimization scheme on simulated data and the real-world KITTI odometry dataset.

The experiments on simulated data show the effectiveness of the PNEC to estimate frame-to-frame rotation compared to the NEC. While the PNEC consistently improves upon the NEC, the benefit depends on the anisotropy and inhomogeneity of the positional uncertainty. Experiments without a translational offset demonstrate that, like the NEC, the PNEC does not suffer for purely rotational motion. Investigations with the self-consistent-field algorithm show its benefits for optimizing the PNEC.

We integrate the PNEC together with a KLT tracker into a state-of-the-art VO system. This achieves better results than the NEC on the real-world KITTI dataset. An ablation study on the simulated data and the KITTI dataset demonstrate the necessity of the proposed two-step optimization for consistently excellent results with the PNEC.

Investigations into the extracted covariances of the KLT tracker show them to be suitable to estimate the shape, but not the scale, of the positional uncertainty of the features on the ICL-NUIM dataset. Nevertheless, this covariance extraction, together with the PNEC, leads to improvements on the KITTI dataset.

This work presents a novel way to integrate uncertainty into relative pose estimation. The PNEC correctly accounts for this uncertainty information, leading to more accurate relative pose estimations between a pair of images. In this thesis, we present work that may lead to further improvements in the topic of 3D vision, particularly relative rotation estimation in the case of pure rotation.

Acronyms

VO visual odometry 1

NEC normal epipolar constraint iv

PNEC probabilistic normal epipolar constraint iv

KLT Kanade-Lucas-Tomasi 5

GRQ generalized Rayleigh quotient 2

SCF self-consistent-field 2

IRLS iteratively reweighted least squares 20

List of Figures

List of Tables

Bibliography

[1] S. Baker and I. Matthews. "Equivalence and efficiency of image alignment algorithms." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2001 (cit. on p. 5).

[2] S. Baker and I. A. Matthews. "Lucas-Kanade 20 Years On: A Unifying Framework." In: *International Journal of Computer Vision (IJCV)* 56 (2004) (cit. on pp. 5, 23, 24).

[3] H. Bay, T. Tuytelaars, and L. Van Gool. "Surf: Speeded up robust features." In: *European Conference on Computer Vision (ECCV)*. 2006 (cit. on p. 5).

[4] A. A. Binbuhaer. "On Optimizing the Sum of Rayleigh Quotients on the Unit Sphere." PhD thesis. The University of Texas at Arlington, 2019 (cit. on pp. 9, 18).

[5] C. M. Bishop. *Pattern recognition and machine learning, 5th Edition*. Springer, 2007 (cit. on p. 22).

[6] M. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel. "What value covariance information in estimating vision parameters?" In: *IEEE International Conference on Computer Vision (ICCV)*. 2001 (cit. on pp. 1, 5, 6, 31, 32).

[7] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age." In: *IEEE Transactions on robotics* 32 (2016) (cit. on p. 3).

[8] L. Carlone, G. C. Calafiore, C. Tommolillo, and F. Dellaert. "Planar Pose Graph Optimization: Duality, Optimal Solutions, and Verification." In: *IEEE Transactions on Robotics* (2016) (cit. on p. 7).

[9] A. Chatterjee and V. M. Govindu. "Efficient and Robust Large-Scale Rotation Averaging." In: *IEEE International Conference on Computer Vision (ICCV)*. 2013 (cit. on pp. 7, 8).

[10] C.-K. Chng, Á. Parra, T.-J. Chin, and Y. Latif. "Monocular Rotational Odometry with Incremental Rotation Averaging and Loop Closure." In: *Digital Image Computing: Techniques and Applications (DICTA)* (2020) (cit. on pp. 4, 7, 8, 30, 40–43, 45, 53).

[11] F. Dellaert, D. M. Rosen, J. Wu, R. Mahony, and L. Carlone. *Shonan Rotation Averaging: Global Optimality by Surfing $SO(p)^n$*. 2020. arXiv: 2008.02737 [cs.CV] (cit. on p. 7).

[12] L. B. Dorini and S. K. Goldenstein. "Unscented feature tracking." In: *Computer Vision and Image Understanding* 115 (2011) (cit. on p. 6).

[13] J. Engel, V. Koltun, and D. Cremers. "Direct Sparse Odometry." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2018) (cit. on pp. 3, 6–8, 46).

[14] J. Engel, T. Schöps, and D. Cremers. "LSD-SLAM: Large-Scale Direct Monocular SLAM." In: *European Conference on Computer Vision (ECCV)*. 2014 (cit. on pp. 3, 6, 8).

[15] K. Fathian, J. P. Ramirez-Paredes, E. A. Doucette, J. W. Curtis, and N. R. Gans. "Quest: A quaternion-based approach for camera motion estimation from minimal feature points." In: *IEEE Robotics and Automation Letters (RAL)* 3 (2018) (cit. on p. 4).

[16] M. A. Fischler and R. C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." In: *Communications of the ACM* 24 (1981) (cit. on p. 42).

[17] W. Förstner and E. 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 (cit. on p. 5).

[18] X. Gao, R. Wang, N. Demmel, and D. Cremers. "LDSO: Direct sparse odometry with loop closure." In: *IEEE International Conference on Intelligent Robots and Systems (IROS)*. 2018 (cit. on pp. 7, 8).

[19] S. Gauglitz, C. Sweeney, J. Ventura, M. Turk, and T. Höllerer. "Live tracking and mapping from both general and rotation-only camera motion." In: *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2012, pp. 13–22. DOI: 10.1109/ISMAR.2012.6402532 (cit. on p. 8).

[20] A. Geiger, P. Lenz, and R. Urtasun. "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012 (cit. on pp. 30, 40).

[21] Á. González. "Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices." In: *Mathematical Geosciences* 42 (2010) (cit. on p. 19).

[22] W. N. Greene and N. Roy. "Metrically-Scaled Monocular SLAM using Learned Scale Factors." In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2020 (cit. on p. 59).

[23] A. Handa, T. Whelan, J. McDonald, and A. Davison. "A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM." In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2014 (cit. on p. 45).

[24] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, 2004 (cit. on pp. 3, 16).

[25] R. I. Hartley. "In defense of the eight-point algorithm." In: *IEEE Transactions on pattern analysis and machine intelligence* 19 (1997) (cit. on p. 4).

[26] D. 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 (cit. on p. 18).

[27] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder. "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds." In: *Inf. Fusion* 14 (2013) (cit. on p. 21).

[28] K. Kanatani. "For geometric inference from images, what kind of statistical model is necessary?" In: *Systems and Computers in Japan* 35 (2004) (cit. on p. 6).

[29] K. Kanatani. "Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis." In: *International Journal of Computer Vision (IJCV)* 80 (2008) (cit. on p. 6).

[30] Y. Kanazawa and K. Kanatani. "Do we really have to consider covariance matrices for image features?" In: *IEEE International Conference on Computer Vision (ICCV)*. 2001 (cit. on p. 6).

[31] G. Klein and D. Murray. "Parallel Tracking and Mapping for Small AR Workspaces." In: *IEEE and ACM International Symposium on Mixed and Augmented Reality*. 2007 (cit. on pp. 3, 5, 6).

[32] L. Kneip and S. Lynen. "Direct Optimization of Frame-to-Frame Rotation." In: *IEEE International Conference on Computer Vision (ICCV)*. 2013 (cit. on pp. 1, 4, 8, 10, 16, 17, 20, 30–35, 37, 38, 40, 41, 58, 59).

[33] L. Kneip, R. Siegwart, and M. Pollefeys. "Finding the Exact Rotation between Two Images Independently of the Translation." In: *European Conference on Computer Vision (ECCV)*. 2012 (cit. on pp. 1, 4, 8, 10, 11, 33).

[34] E. Kruppa. *Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung*. Hölder, 1913 (cit. on p. 3).

[35] Z. Kukelova, M. Bujnak, and T. Pajdla. "Polynomial Eigenvalue Solutions to the 5-pt and 6-pt Relative Pose Problems." In: *British Machine Vision Conference (BMVC)*. 2008 (cit. on p. 3).

[36] S. H. Lee and J. Civera. "Geometric Interpretations of the Normalized Epipolar Error." In: *ArXiv* abs/2008.01254 (2020) (cit. on pp. 4, 10).

[37] S. H. Lee and J. Civera. "Rotation-Only Bundle Adjustment." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021 (cit. on pp. 8, 10, 12, 17).

[38] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. "Keyframe-based visual–inertial odometry using nonlinear optimization." In: *The International Journal of Robotics Research* (2015) (cit. on pp. 4, 7, 8).

[39] K. Levenberg. "A method for the solution of certain non-linear problems in least squares." In: *Quarterly of applied mathematics* 2 (1944) (cit. on p. 17).

[40] H. Li and R. Hartley. "Five-point motion estimation made easy." In: *IEEE International Conference on Pattern Recognition (ICPR)*. 2006 (cit. on pp. 1, 3).

[41] J. Lim, N. Barnes, and H. Li. "Estimating relative camera motion from the antipodal-epipolar constraint." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 32 (2010) (cit. on p. 4).

[42] H. Longuet-Higgins. "Readings in computer vision: issues, problems, principles, and paradigms." In: *A computer algorithm for reconstructing a scene from two projections* (1987) (cit. on pp. 1, 3, 4).

[43] D. G. Lowe. "Distinctive image features from scale-invariant keypoints." In: *International Journal of Computer Vision (IJCV)* 60 (2004) (cit. on p. 5).

[44] B. D. Lucas and T. Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision." In: *International Joint Conference on Artificial Intelligence (IJCAI)*. 1981 (cit. on pp. 5, 24).

[45] D. W. Marquardt. "An algorithm for least-squares estimation of nonlinear parameters." In: *Journal of the society for Industrial and Applied Mathematics* 11 (1963) (cit. on p. 17).

[46] J. Meidow, C. Beder, and W. Förstner. "Reasoning with uncertain points, straight lines, and straight line segments in 2D." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (2009) (cit. on p. 6).

[47] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System." In: *IEEE Transactions on Robotics* 31 (2015) (cit. on pp. 3, 5–8).

[48] R. Mur-Artal and J. D. Tardós. "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras." In: *IEEE Transactions on Robotics* 33 (2017) (cit. on pp. 3, 5, 6).

[49] R. Mur-Artal and J. D. Tardós. "Visual-Inertial Monocular SLAM With Map Reuse." In: *IEEE Robotics and Automation Letters* (2017) (cit. on pp. 4, 8).

[50] D. Nister. "An efficient solution to the five-point relative pose problem." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2003 (cit. on pp. 1, 3, 4).

[51] D. Nister and H. Stewenius. "Scalable Recognition with a Vocabulary Tree." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2006 (cit. on p. 7).

[52] M. Nixon and A. Aguado. *Feature extraction and image processing for computer vision*. 2019 (cit. on p. 5).

[53] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999 (cit. on p. 21).

[54] U. Orguner and F. Gustafsson. "Statistical Characteristics of Harris Corner Detector." In: *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*. 2007 (cit. on p. 5).

[55] C. Pirchheim, D. Schmalstieg, and G. Reitmayr. "Handling pure camera rotation in keyframe-based SLAM." In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2013, pp. 229–238. DOI: 10.1109/ISMAR.2013.6671783 (cit. on p. 8).

[56] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. "ORB: An efficient alternative to SIFT or SURF." In: *IEEE International Conference on Computer Vision (ICCV)*. 2011 (cit. on p. 5).

[57] S. Sheorey, S. Keshavamurthy, H. Yu, H. Nguyen, and C. N. Taylor. "Uncertainty estimation for KLT tracking." In: *Asian Conference on Computer Vision*. 2014 (cit. on p. 5).

[58] J. Shi and C. Tomasi. "Good features to track." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1994 (cit. on p. 5).

[59] G. Sibley, L. Matthies, and G. Sukhatme. "A sliding window filter for incremental SLAM." In: *Unifying perspectives in computational and robot vision*. 2008 (cit. on pp. 6, 7).

[60] A. F. Siegel. "The noncentral chi-squared distribution with zero degrees of freedom and testing for uniformity." In: *Biometrika* (1979) (cit. on p. 46).

[61] R. Steele and C. Jaynes. "Feature uncertainty arising from covariant image noise." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2005 (cit. on p. 5).

[62] H. Stewenius, C. Engels, and D. Nistér. "Recent developments on direct relative orientation." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (2006) (cit. on p. 3).

[63] H. Strasdat, J. Montiel, and A. J. Davison. "Scale drift-aware large scale monocular SLAM." In: *Robotics: Science and Systems VI* (2010) (cit. on p. 8).

[64] L. von Stumberg, V. Usenko, and D. Cremers. "Direct Sparse Visual-Inertial Odometry using Dynamic Marginalization." In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2018 (cit. on pp. 4, 8).

[65] L. Stumberg, P. Wenzel, N. Yang, and D. Cremers. "LM-Reloc: Levenberg-Marquardt Based Direct Visual Relocalization." In: *2020 International Conference on 3D Vision (3DV)* (2020) (cit. on p. 3).

[66] Y. Sun, L. Zhao, S. Huang, L. Yan, and G. Dissanayake. "L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry." In: *ISPRS Journal of Photogrammetry and Remote Sensing* (2014) (cit. on p. 5).

[67] R. Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010 (cit. on p. 3).

[68] C. Tomasi and T. Kanade. "Detection and Tracking of Point Features." In: *International Journal of Computer Vision (IJCV)* 9 (1991) (cit. on pp. 5, 24).

[69] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. "Bundle adjustment—a modern synthesis." In: *International workshop on vision algorithms*. 1999 (cit. on pp. 3, 7).

[70] J. K. Uhlmann. "Dynamic map building and localization: New theoretical foundations." PhD thesis. University of Oxford Oxford, 1995 (cit. on p. 13).

[71] V. Usenko, N. Demmel, D. Schubert, J. Stückler, and D. Cremers. "Visual-Inertial Mapping With Non-Linear Factor Recovery." In: *IEEE Robotics and Automation Letters (RAL)* 5 (2020) (cit. on pp. 3, 5, 6, 22, 23, 41–43, 71).

[72] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós. "A comparison of loop closing techniques in monocular SLAM." In: *Robotics and Autonomous Systems* (2009) (cit. on p. 8).

[73] N. Yang, L. von Stumberg, R. Wang, and D. Cremers. "D3VO: Deep Depth, Deep Pose and Deep Uncertainty for Monocular Visual Odometry." In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020 (cit. on p. 59).

[74] N. Yang, R. Wang, J. Stückler, and D. Cremers. "Deep Virtual Stereo Odometry: Leveraging Deep Depth Prediction for Monocular Direct Sparse Odometry." In: *European Conference on Computer Vision (ECCV)*. 2018 (cit. on p. 59).

[75] X.-M. Yang, W. Wu, L.-B. Qing, H. Hua, and X.-h. He. "Image feature extraction and matching technology [J]." In: *Optics and Precision Engineering* (2009) (cit. on p. 5).

[76] B. Zeisl, P. Georgel, F. Schweiger, E. Steinbach, and N. Navab. "Estimation of Location Uncertainty for Scale Invariant Feature Points." In: *British Machine Vision Conference (BMVC)*. 2009 (cit. on pp. 5, 6).

[77] H. Zhang, D. Grießbach, J. Wohlfeil, and A. Börner. "Uncertainty model for template feature matching." In: *Pacific-Rim Symposium on Image and Video Technology*. Springer. 2017, pp. 406–420 (cit. on p. 5).

[78] L.-H. Zhang. "On optimizing the sum of the Rayleigh quotient and the generalized Rayleigh quotient on the unit sphere." In: *Computational Optimization and Applications* 54 (2013) (cit. on pp. 9, 18).

[79] L.-H. Zhang and R. Chang. "A Nonlinear Eigenvalue Problem Associated with the Sum-of-Rayleigh-Quotients Maximization." In: *Transactions on Applied Mathematics* 2 (2021) (cit. on pp. 9, 18, 38).

A. Hyperparameter

Hyperparameter	Simulated	KITTI	ICL-NUIM
EV iterations	20	10	
SCF itertations	10	10	
Fibonacci lattice points	500	500	
regularization	0	10^{-13}	
KLT parameters			
pattern size		52	52
grid size		30	30
pyramid-levels		4	4
optical flow iterations		40	40
optical flow max recovered distance		0.04	0.04

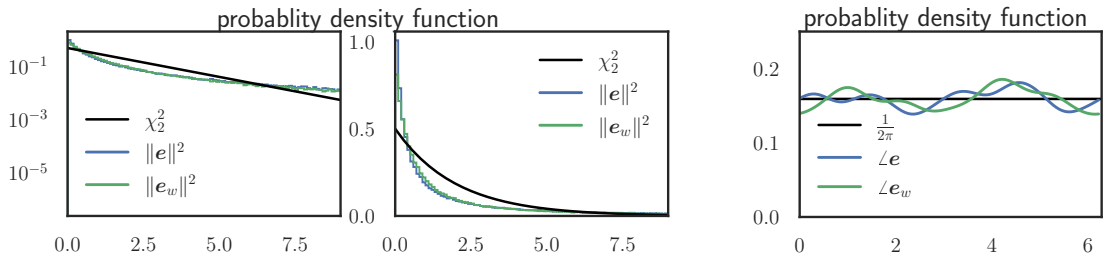
Table A.1.: Parameters used for the experiments.

Tab. A.1 gives an overview and a short explanation of the parameters used in the experiments.

For the PNEC we use: **EV iterations** is the number of iterations in Alg. 3 before the least squares refinement; **SCF iterations** is the number of iterations the SCF method (see Alg. 1) is run; **Fibonacci lattice points** is the number of points we sample on the unit sphere using the Fibonacci lattice; **regularization** is the regularization constant proposed in Sec. 3.4.2 for the PNEC.

The KLT parameters are: **pattern size** the pattern layout of the KLT tracker, see `include/basalt/optical_flow/patterns.h` in the implementation of [71] for more details; **grid size** is the length of each square in the image for which a track is extracted; **pyramid-levels** is the number of pyramid levels over which the KLT tracker tracks, where the scale factor between each pyramid level is 2; **optical flow iterations** is the number of iterations of tracking on each pyramid level, respectively; **optical flow max recovered distance** is the maximum distance between its original position and the forward-backward tracking position, otherwise it is discarded.

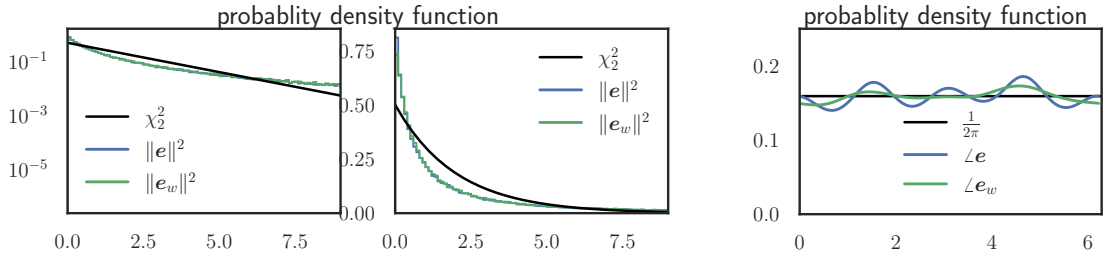
B. Covariance Extraction



(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis

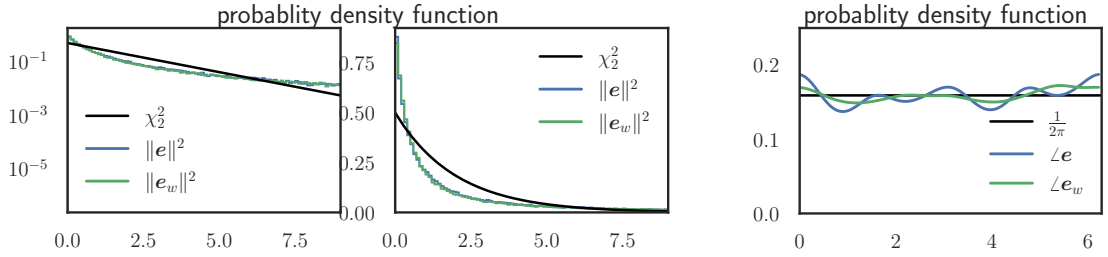
(b) smoothed $\angle\cdot$ distribution

Figure B.1.: Error distribution on sequence living room 0. The error distributions before (e) and after (e_w) applying the whitening transformation with the extracted covariance matrices are shown. Fig. B.1a shows the squared sum of the error distributions and the χ^2_2 distribution with a logarithmic and linear y-axis. Applying the whitening transformation results in a very similar distribution. Both exhibit more small errors than the χ^2_2 distribution. Fig. B.1b shows the angular error distributions. The anisotropy of the angular distribution in both cases is very similar.



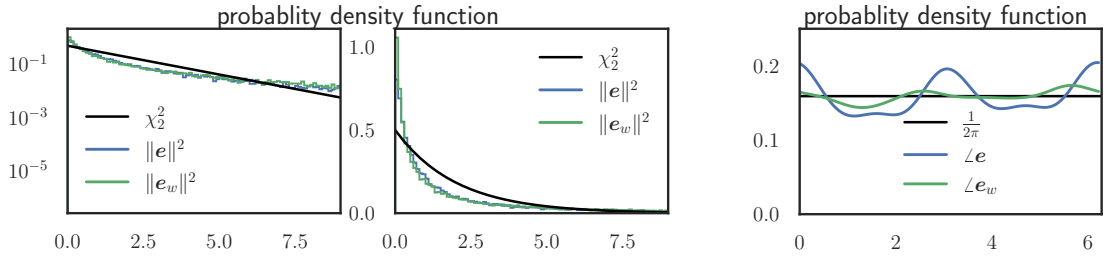
(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis (b) smoothed $\angle\cdot$ distribution

Figure B.2.: Error distribution on sequence living room 1. For more details, see Fig. B.1. Applying the whitening transformation results in a very similar distribution of the squared sum. Both exhibit more small errors than the χ^2_2 distribution. The anisotropy of the angular distribution is significantly smoothed after applying the whitening transformation.



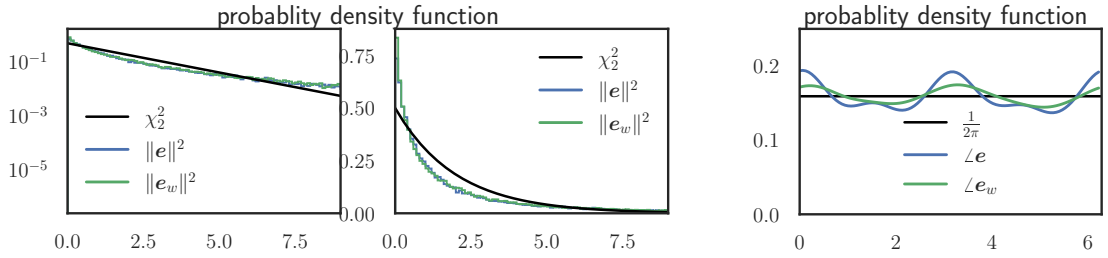
(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis (b) smoothed $\angle\cdot$ distribution

Figure B.3.: Error distribution on sequence living room 2. For more details, see Fig. B.1. Applying the whitening transformation results in a very similar distribution of the squared sum. Both exhibit more small errors than the χ^2_2 distribution. The anisotropy of the angular distribution is smoothed after applying the whitening transformation.



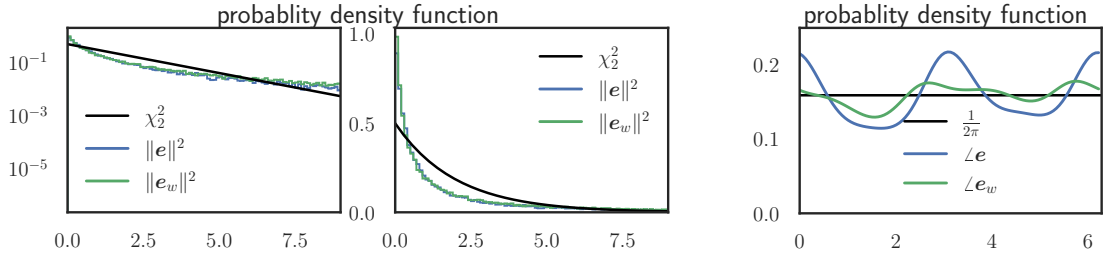
(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis (b) smoothed $\angle\cdot$ distribution

Figure B.4.: Error distribution on sequence office room 1. For more details, see Fig. B.1. Applying the whitening transformation results in a very similar distribution of the squared sum. Both exhibit more small errors than the χ^2_2 distribution. The anisotropy of the angular distribution is greatly reduced after applying the whitening transformation.



(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis (b) smoothed $\angle\cdot$ distribution

Figure B.5.: Error distribution on sequence office room 2. For more details, see Fig. B.1. Applying the whitening transformation results in a very similar distribution of the squared sum. Both exhibit more small errors than the χ^2_2 distribution. The anisotropy of the angular distribution is significantly reduced after applying the whitening transformation.



(a) $\|\cdot\|^2$ distribution with logarithmic and linear y-axis

(b) smoothed $\angle\cdot$ distribution

Figure B.6.: Error distribution on sequence office room 3. For more details, see Fig. B.1. Applying the whitening transformation results in a very similar distribution of the squared sum. Both exhibit more small errors than the χ^2_2 distribution. The anisotropy of the angular distribution is greatly reduced after applying the whitening transformation.

C. Contributions

The following contributions to this thesis were made by or are based on the direct work by Lukas Koestler:

Sec. 3.4.1 is based on the work by Lukas Koestler.

Fig. 3.5 was made by Lukas Koestler.

Fig. 3.7 was made by Lukas Koestler.

Alg. 1, Alg. 2, Alg. 3 are taken from the paper submission. They were written down in this form by Lukas Koestler.