This ICCV2013 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

LM method to solve for rotation given n matched feature points, by enforcing coplanar constraints. Also, use eigen perturbation to bound the searching of the lambda (eigen value of the target matrix M)

Direct Optimization of Frame-to-Frame Rotation

Laurent Kneip
Research School of Engineering
Australian National University
laurent.kneip@anu.edu.au

Simon Lynen
Autonomous Systems Lab
ETH Zurich
simon.lynen@mavt.ethz.ch

Abstract

This work makes use of a novel, recently proposed epipolar constraint for computing the relative pose between two calibrated images. By enforcing the coplanarity of epipolar plane normal vectors, it constrains the three degrees of freedom of the relative rotation between two camera views directly—independently of the translation.

The present paper shows how the approach can be extended to n points, and translated into an efficient eigenvalue minimization over the three rotational degrees of freedom. Each iteration in the non-linear optimization has constant execution time, independently of the number of features. Two global optimization approaches are proposed. The first one consists of an efficient Levenberg-Marquardt scheme with randomized initial value, which already leads to stable and accurate results. The second scheme consists of a globally optimal branch-and-bound algorithm based on a bound on the eigenvalue variation derived from symmetric eigenvalue-perturbation theory. Analysis of the cost function reveals insights into the nature of a specific relative pose problem, and outlines the complexity under different conditions. The algorithm shows state-of-the-art performance w.r.t. essential-matrix based solutions, and a frame-to-frame application to a video sequence immediately leads to an alternative, real-time visual odometry solution.

Note: All algorithms in this paper are made available in the OpenGV library. Please visit
http://laurentkneip.github.io/opengv

1. Introduction

The computation of the relative pose between two planar projections of a scene is certainly one of the most studied problems in geometric vision and—more generally—structure from motion. Even in times where powerful motion-model-based egomotion estimation solutions become available, the basic computation of the relative pose between two images remains of utmost importance, especially for the initialization of model-based solutions or pose estimation problems that do not originate from a motion process evolving over time.

Most common solutions to the calibrated relative pose problem make use of the essential matrix parametrization, which can result in a linear solution, depending on the number of employed correspondences. These non-iterative epipolar geometry approaches are generally fast and return accurate results in many situations. However, there remains a number of problems linked to the indirect essential matrix parametrization:

- *Mixing of parameters*: The parameters of the essential matrix do not represent motion-related variables directly, but rather functions of these variables. This complicates the possibility to draw conclusions about the nature of the problem just by analyzing the values.
- *Solution multiplicity*: Minimal solvers naturally return multiple essential matrices. However, even for a single essential matrix the decomposition still leads to multiple solutions. The functional mapping between the motion variables—namely rotation and translation—and the essential matrix is not uniquely invertible. The common approach to disambiguate the solutions consists of triangulating features and checking their reprojection error or location w.r.t. the image plane. The solution hence depends on the computation of structure in parallel to the motion.
- *Structural degeneracy*: The most efficient solution to the problem—the linear solver by Longuet-Higgins [13]—is known to suffer from planar degeneracy. This is due to the fact that the solution for n points is no longer unique in the case of a planar point distribution.
- *Zero translation*: Although most essential-matrix based solutions implicitly solve correctly for rotation in the zero-translation and noise-free situation, the constraint as such deteriorates. $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$ results in a zero matrix, and the constraint is always fulfilled.

Essential-matrix based solutions therefore depend on model selection, decomposition, triangulation, etc., which hinders an intuitive usage in practical situations. The present paper sheds a new light on the relative pose problem by employing a different epipolar constraint not suffering from any of the above mentioned problems. Instead of

solving for all the motion parameters indirectly, we show how the relative pose problem can be turned into an eigenvalue minimization directly over the three parameters of the frame-to-frame rotation. The relative translation results implicitly in form of the corresponding eigenvector. The result is an elegant and simple non-linear optimization over three parameters only, and the cost function typically shows only a small number of geometrically meaningful local minima besides the globally optimal solution. Unlike the essential matrix constraint, the cost function maintains a basin of attraction in the zero-translation situation, and the algorithm can be solved with constant iteration time independently of the number of features. Despite being an iterative approach, the proposed algorithm therefore turns out to be a valid alternative to existing calibrated relative pose algorithms, directly applicable to frame-to-frame motion estimation.

Related work: The first known solution to the relative pose problem dates back to 1913 and has been presented by Kruppa [9]. While this parametrization still leads to a multivariate polynomial equation system with 11 solutions, more recent advances have shown that the actual number of solutions in the minimal case equals to 10. This is for instance the case in the minimal solution of Nistér [16], who derives a tenth-order polynomial that is subsequently solved using Sturm's root-bracketing approach. [18], [10], and [11] represent alternative solutions to the problem using the Gröbner basis, polynomial eigenvalue, or hidden variable resultant technique, respectively. The solutions show minor differences in terms of accuracy and computational complexity, however all make use of the essential matrix parametrization, and thus suffer from the previously mentioned problems. The linear solver by Longuet-Higgins [13]—later on defended by Hartley [4]—, is one of the first essential matrix based solutions. It delivers a unique essential matrix but suffers from the previously mentioned planar degeneracy.

Although most algorithms are still able to find the rotation in the noise-free zero-translation situation, the inlier computation still depends on the translation, which becomes unobservable. Torr et al. [20] address this problem by model selection. Lim et al. [12] compute the rotation independently of the translation but depend on a special distribution of the feature correspondences, i.e. antipodal points. Kneip et al. [8] propose an alternative epipolar constraint. Although their minimal solution depends on model selection, they show the general ability to compute the rotation independently of the translation. This ability has been first caught up in [15], with several follow-up works. This line of research serves as a basis for our n-point-solution.

In terms of global optimality, the present paper has analogies with the work of Hartley and Kahl [5], who also devise a branch-and-bound solution in rotation space. Although only minimizing an algebraic error, our solution is more efficient: instead of finding the translation for each rotation via second-order cone programming, we only have to compare the closed-form bound on the eigenvalue variation of a 3×3 matrix. By remaining robust in the zero-translation situation, the cost function used in the algorithm remains applicable to infinitesimal motion too. Disregarding the quadratic terms—which is allowed in the first-order approximation for small angles—shows that the presently employed Cayley parameters become approximately proportional to the rotational velocity in that situation. We thus obtain also a continuous equivalent to [19] and [21], which represent minimal and non-minimal continuous epipolar solutions. The presented work is also related to research conducted around direct optimization on the essential matrix manifold, such as Ma et al. [14] and Helmke et al. [7]. The main difference in our work is the cost function: While all previous works do an optimization over five degrees of freedom, we prove in this work the existence of a different cost function that allows an efficient optimization over three degrees of freedom only. Moreover, none of the advantages is lost: the algorithm can still use prior data compression techniques and the Cayley transformation to end up with simple Gauss-Newton-like optimization in a Euclidean space.

Finally, Nistér et al. [17] presents motion estimation results by continuously applying the five-point algorithm to a video sequence. We compare our results to this approach, and show the benefits of our method compared to essential matrix based parametrizations.

Organization of the paper: Section 2 outlines the formulation of the relative pose problem as an eigenvalue minimization problem. Section 3 presents two approaches to find the solution, an efficient Levenberg-Marquardt scheme as well as a bound on the variation of the eigenvalue allowing the design of a branch-and-bound solution. Section 4 shows comparative results of our algorithm including the possibility to identify the location of local minima in the cost function directly from the geometrical conditioning of the problem. Section 5 finally shows results on real video sequences, and a successful direct frame-to-frame visual odometry solution.

2. Theory

This section outlines the main geometrical concept that allows us to constrain the rotation between different viewpoints independently of the translation, namely the coplanarity of epipolar plane normal vectors. Using this constraint allows us to formulate the relative pose computation as an eigenvalue minimization problem. We furthermore introduce a good parametrization of the rotation matrix for an efficient optimization, and reformulate the problem such that rotations may be validated in constant time, independently of the number of features.

2.1. Preliminaries

We assume to be in the central, calibrated case such that each location in the image plane can be translated into a unique unit bearing vector originating from the camera center. Each pair of bearing vectors $(\mathbf{f}_i, \mathbf{f}'_i)$ denotes a correspondence of bearing vectors pointing at the same 3D world point $\mathbf{p}_i$ from two distinct view-points, where $\mathbf{f}_i$ represents the observation from the first view-point, and $\mathbf{f}'_i$ the one from the second. The relative pose is given by the translation $\mathbf{t}$—expressed in the first frame and denoting the position of the second frame w.r.t. the first one—and the rotation $\mathbf{R}$—transforming vectors from the second into the first frame. These variables are illustrated in Figure 1.

2.2. Relative pose as an eigenvalue problem

The epipolar plane of a correspondence is defined to be the plane that contains the two camera centers as well as the observed 3D point. The set of epipolar planes hence forms a pencil of planes all intersecting in the line of translation. In other words, the normal vectors of the epipolar planes all need to be coplanar. A normal vector in the pure translation situation is easily given by $\mathbf{n}_i = \mathbf{f} \times \mathbf{f}'_i$. In case of having rotation, this formula changes to

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

The work of Kneip [8] proposes to enforce the coplanarity of triplets of normal vectors in order to come up with a minimal solution for translation independent computation of the relative rotation. An interesting result around the novel epipolar constraint is that—when varying the rotation—virtual and notably non-coplanar epipolar plane normal vectors appear even in the zero-translation situation, which renders the constraint robust against vanishing translation magnitudes. However, the Gröbner basis solver presented in [8] still turns out to be unstable in this case, which is related to numerical instabilities in the fixed sequence of s-polynomials when the number of solutions changes. In the following, we will present a more interesting n-point iterative optimization scheme that remains functional for any parallax.

The basic intuition to enforce the coplanarity of n epipolar plane normals consists of treating the set of normal vectors as a point cloud, and canceling the second moment or dilatation in one direction. This is easily achieved by stacking all normal vectors into a 3-by-n matrix $\mathbf{N} = \begin{bmatrix}\mathbf{n}_1 & \dots & \mathbf{n}_n\end{bmatrix}$, and minimizing the smallest eigenvalue of $\mathbf{N}\mathbf{N}^T$. Let $\mathbf{N}\mathbf{N}^T = \sum_{i=1}^{n} \mathbf{n}_i\mathbf{n}_i^T = \mathbf{M}$. $\mathbf{M}$ is a real symmetric positive-definite 3-by-3 matrix and a function of $\mathbf{R}$. If $\lambda_{\mathbf{M},min}$ denotes the smallest eigenvalue of $\mathbf{M}$, the final problem parametrization becomes

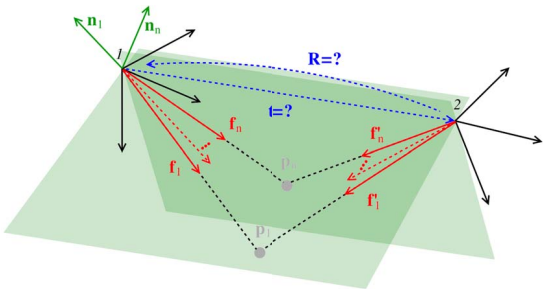


Figure 1. Synopsis of the n-point relative pose problem. Bearing vectors (in red) from two view-points are the known variables, and the relative pose (in blue) is the searched unknown.

$$\begin{aligned} \mathbf{R} &= \operatorname{argmin}_{\mathbf{R}} \lambda_{\mathbf{M},min} \\ \text{with } \mathbf{M} &= \sum_{i=1}^{n} (\mathbf{f}_i \times \mathbf{R}\mathbf{f}'_i)(\mathbf{f}_i \times \mathbf{R}\mathbf{f}'_i)^T. \end{aligned} \tag{2}$$

Note that $\mathbf{M}$ is a real symmetric and positive-definite matrix with rank at most 2. 2 is therefore equivalent to iterative rank minimization. Also note that—at the end of the optimization—the direction of translation is automatically given by the eigenvector that corresponds to the smallest eigenvalue. The size of the remaining eigenvalues incorporates the condition of the problem, meaning how close we are to a zero-translation situation. The rotation having three degrees of freedom, this represents at best a non-linear optimization over three parameters only—depending on the rotation parametrization.

2.3. Minimal parametrization of the rotation matrix

There exists a large number of possible rotation matrix parametrizations. However, the avoidance of additional constraints and the efficiency of the optimization are good reasons to choose a minimal parametrization. It is also a good idea to select a symmetric parametrization, which is why the algorithm uses the Cayley transformation [1]. It represents a good choice because—in practice—the frame-to-frame rotation does barely exceed $\frac{\pi}{2}$ about any of the basis' axes. If the Cayley parameters are denoted with $\mathbf{v} = (x\ y\ z)^T$, the parametrization involves a scale factor $(1 + x^2 + y^2 + z^2)^{-1}$. An interesting observation is that this scale factor is always positive and can be factorized from (1). It therefore only affects the magnitude of the normal vector, and can be omitted in the coplanarity maximization (2). We therefore chose the following, quadratic parametrization

$$\mathbf{R} = 2(\mathbf{v}\mathbf{v}^T - [\mathbf{v}]_\times) + (1 - \mathbf{v}^T\mathbf{v})\mathbf{I}. \tag{3}$$

2.4. Rendering the iteration time constant

An important observation is that the rotation matrix can be factorized inside the expression $(\mathbf{f}_i \times \mathbf{R}\mathbf{f}'_i)(\mathbf{f}_i \times \mathbf{R}\mathbf{f}'_i)^T$. On the other hand, the summation over i in $\mathbf{M}$ can be pulled

into the different elements of the matrix. If $\mathbf{r}_i$ denotes row i of $\mathbf{R}$ and m_{ij} an element of $\mathbf{M}$, we obtain

$$m_{11} = \mathbf{r}_3(\sum_{i=1}^{n} f_{yi}^2 \mathbf{F}'_i)\mathbf{r}_3^T - \mathbf{r}_3(2\sum_{i=1}^{n} f_{yi}f_{zi}\mathbf{F}'_i)\mathbf{r}_2^T + \mathbf{r}_2(\sum_{i=1}^{n} f_{zi}^2 \mathbf{F}'_i)\mathbf{r}_2^T$$

$$m_{22} = \mathbf{r}_1(\sum_{i=1}^{n} f_{zi}^2 \mathbf{F}'_i)\mathbf{r}_1^T - \mathbf{r}_1(2\sum_{i=1}^{n} f_{xi}f_{zi}\mathbf{F}'_i)\mathbf{r}_3^T + \mathbf{r}_3(\sum_{i=1}^{n} f_{xi}^2 \mathbf{F}'_i)\mathbf{r}_3^T$$

$$m_{33} = \mathbf{r}_2(\sum_{i=1}^{n} f_{xi}^2 \mathbf{F}'_i)\mathbf{r}_2^T - \mathbf{r}_1(2\sum_{i=1}^{n} f_{xi}f_{yi}\mathbf{F}'_i)\mathbf{r}_2^T + \mathbf{r}_1(\sum_{i=1}^{n} f_{yi}^2 \mathbf{F}'_i)\mathbf{r}_1^T$$

$$\begin{aligned} m_{12} = m_{21} \quad = \quad & \mathbf{r}_1(\sum_{i=1}^{n} f_{yi}f_{zi}\mathbf{F}'_i)\mathbf{r}_3^T - \mathbf{r}_1(\sum_{i=1}^{n} f_{zi}^2 \mathbf{F}'_i)\mathbf{r}_2^T \\ & -\mathbf{r}_3(\sum_{i=1}^{n} f_{xi}f_{yi}\mathbf{F}'_i)\mathbf{r}_3^T + \mathbf{r}_3(\sum_{i=1}^{n} f_{xi}f_{zi}\mathbf{F}'_i)\mathbf{r}_2^T \\ m_{13} = m_{31} \quad = \quad & \mathbf{r}_2(\sum_{i=1}^{n} f_{xi}f_{yi}\mathbf{F}'_i)\mathbf{r}_3^T - \mathbf{r}_2(\sum_{i=1}^{n} f_{xi}f_{zi}\mathbf{F}'_i)\mathbf{r}_2^T \\ & -\mathbf{r}_1(\sum_{i=1}^{n} f_{yi}^2 \mathbf{F}'_i)\mathbf{r}_3^T + \mathbf{r}_1(\sum_{i=1}^{n} f_{yi}f_{zi}\mathbf{F}'_i)\mathbf{r}_2^T \\ m_{23} = m_{32} \quad = \quad & \mathbf{r}_1(\sum_{i=1}^{n} f_{xi}f_{zi}\mathbf{F}'_i)\mathbf{r}_2^T - \mathbf{r}_1(\sum_{i=1}^{n} f_{zi}f_{yi}\mathbf{F}'_i)\mathbf{r}_1^T \\ & -\mathbf{r}_3(\sum_{i=1}^{n} f_{xi}^2 \mathbf{F}'_i)\mathbf{r}_2^T + \mathbf{r}_3(\sum_{i=1}^{n} f_{xi}f_{yi}\mathbf{F}'_i)\mathbf{r}_1^T \end{aligned}$$

with $\mathbf{F}'_i = \mathbf{f}'_i \cdot \mathbf{f}'^T_i$. A notable fact in this expression is that all the summation terms have to be computed only once over the original feature correspondences (linear complexity), and can then be reused constantly throughout the entire rank minimization (i.e., the rotation optimization). Similar to Helmke et al. [7], this data compression technique results in constant iteration time.

3. Solving the problem

In this paragraph, we first show how the non-linear problem (2) can be solved efficiently in a Levenberg-Marquardt scheme. Second, we use eigenvalue perturbation theory in order to derive a bound on the variation of the smallest eigenvalue, finally enabling a globally optimal branch-and-bound optimization of the relative rotation.

3.1. Levenberg-Marquardt scheme

A straightforward solution to the minimization problem (2) is given by applying the gradient descent algorithm. Notably, as shown in the supplemental material, $\lambda_{\mathbf{M},min}$ can be retrieved in closed form. In order to find the minimum more efficiently, we apply the following trick. The Levenberg-Marquardt scheme is a Gauss-Newton-like procedure for finding the zero of a system of multiple, multivariate, and notably nonlinear constraints. The number of constraints needs to be at least equal to the number of unknowns. To obtain such constraints and apply Levenberg-Marquardt, we simply enforce the first-order partial derivatives of $\lambda_{\mathbf{M},min}$ w.r.t. the Cayley parameters to be zero, which also defines a minimum.

$$\frac{\partial \lambda_{\mathbf{M},min}}{\partial x} = 0, \quad \frac{\partial \lambda_{\mathbf{M},min}}{\partial y} = 0, \quad \frac{\partial \lambda_{\mathbf{M},min}}{\partial z} = 0. \qquad (4)$$

The derivation of these constraints in closed-form can again be found in the supplemental material. The online Levenberg-Marquardt solver derives the Jacobians of these constraints (i.e. the second order derivatives of $\lambda_{\mathbf{M},min}$) numerically. Local minima are avoided by a random variation of the starting point.

3.2. Branch and bound the variation of $\lambda_{\mathbf{M},\min}$

It is almost trivial to derive an absolute bound on the variation of the rotation matrix $\mathbf{R}$ based on an absolute bound ϵ on the variation of x, y, and z. Using this bound, we can derive an absolute bound on the variation of the elements of $\mathbf{M}$. Let $\mathcal{M}$ be the real symmetric 3-by-3 matrix formed by these bounds. $\mathcal{M}$ can be regarded as a bound on a Hermitian perturbation of $\mathbf{M}$, and hence the eigensystem. A formal derivation of these bounds can be found in the supplemental material.

Let $\mathcal{M}^*$ be a concrete perturbation of the eigensystem[1]. The perturbation follows $\mathbf{M} \rightarrow \mathbf{M} + \mathcal{M}^*$. An important result from the eigenvalue perturbation theory—the *Weyl-theorem* as presented in [2]—tells us that the relative perturbation of the eigenvalues of $\mathbf{M}$ is then bounded by

$$\frac{|\lambda_{i,perturbed} - \lambda_i|}{|\lambda_i|} \leq \|\mathbf{M}^{-1/2}\mathcal{M}^*\mathbf{M}^{-1/2}\|_2. \qquad (5)$$

In an aim to bound the spectral norm, we first of all note that the computation of the eigenvalues of a real symmetric positive-definite 3-by-3 matrix can be done very efficiently and in closed-form. We obtain $\mathbf{M} = \mathbf{V}\text{diag}(\lambda_1, \lambda_2, \lambda_3)\mathbf{V}^T$, with positive eigenvalues only. This means that $\mathbf{M}^{-1/2}$ is easily given by $\mathbf{M}^{-1/2} = \mathbf{V}\text{diag}(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \lambda_3^{-1/2})\mathbf{V}^T$, and remains a real symmetric positive-definite matrix. $\mathbf{M}^{-1/2}\mathcal{M}^*\mathbf{M}^{-1/2}$ hence remains a real symmetric matrix (not necessarily positive definite). Another important property from the *spectral theorem* then tells us that the spectral norm of a real symmetric matrix is given by the absolute value of its largest eigenvalue. The final step hence consists of deriving the characteristic polynomial

$$\det(\mathbf{M}^{-1/2}\mathcal{M}^*\mathbf{M}^{-1/2} - \mu\mathbf{I}_3) = -\mu^3 + a_2\mu^2 + a_1\mu + a_0, \qquad (6)$$

and bounding the absolute value of its roots. This implicitly bounds the largest eigenvalue of $\mathbf{M}^{-1/2}\mathcal{M}^*\mathbf{M}^{-1/2}$

[1] $\mathcal{M}^*$ is a real symmetric matrix where the absolute value of each entry is smaller or equal to the corresponding entry in $\mathcal{M}$. $\mathcal{M}^*$ is not necessarily positive-definite anymore.

and hence the spectral norm. A good bound on the absolute value of the roots of (6) is for example given by the Lagrangian bound $|a_2| + |a_1|^{1/2} + |a_0|^{1/3}$ (conservative approximation). With α_2, α_1, and α_0 representing bounds on $|a_2|$, $|a_1|$, and $|a_0|$ as a function of the bound on $\mathcal{M}^*$—notably $\mathcal{M}$—and the absolute value of the entries of $\mathbf{M}^{-1/2}$, we finally obtain

$$\begin{aligned} \frac{|\lambda_{i,perturbed} - \lambda_i|}{\lambda_i} &\leq \alpha_2 + \alpha_1^{1/2} + \alpha_0^{1/3} \\ \Rightarrow |\lambda_{i,perturbed} - \lambda_i| &\leq \lambda_i(\alpha_2 + \alpha_1^{1/2} + \alpha_0^{1/3}). \quad (7) \end{aligned}$$

The derivation of α_2, α_1, and α_2 can again be found in the supplemental material. As expected, the bound on the eigenvalue variation converges to zero along with $\mathcal{M}$.

4. Simulation results

The branch-and-bound algorithm is essentially able to find the globally optimal solution with desired resolution. It is however not very efficient due to the fact that the bound on the eigenvalue variation turns out to be fairly conservative. The approach is also not taking into account any outliers. While leaving it as the first step towards an interesting future alternative to [5], the remainder of this paper focusses on an evaluation of the real-time Levenberg-Marquardt scheme. We first start with analyzing the behavior of the cost function, and thus the complexity of avoiding local minima. We then proceed to a comparison of computational efficiency and noise resilience to different state-of-the-art algorithms. The most critical part—namely the performance in presence of outliers and convergence in abscence of any prior knowledge—is finally analyzed by embedding and comparing all algorithms within a RANSAC framework.

4.1. Experiment outline

We generate random problems by first fixing the position of the first frame to the origin and its orientation to identity. The translational offset of the second frame is chosen with uniformly distributed random direction and maximum magnitude of 2. The orientation of the second frame is generated with random Euler angles bounded to 0.5 rad in absolute value. This generates random relative poses as they would appear in practical situations. Bearing vector correspondences result from uniformly distributed random points around the origin with a distance varying between 4 and 8, transforming those points into both frames, and normalizing. Noise is added by assuming a spherical camera with a focal length of 800 pix, extracting the tangential plane of each bearing vector, and adding a uniformly distributed random offset expressed in pixels inside this plane. We execute 1000 problems for each noise level/outlier percentage. We compare the mean and median error of our solution (eig) to all commonly applied algorithms, namely [18] (stew), [16] (nist), [8] (kneip), [6] (7pt), [4] (8pt), as well as two-view bundle-adjustment (nonlin). The translational accuracy is expressed by the angular error of the normalized direction, and the error in rotation is expressed by the norm of the difference between both the estimated and the ground truth rotation vectors.

4.2. Behavior of the cost function

Direct iterative optimization of the relative pose necessarily requires a comparison to standard two-view bundle adjustment, which depends on proper initialization. The biggest difference lies in the dimensionality of the problem: While standard two-view bundle adjustment reduces the geometric error over $5 + 3n$ degrees of freedom (n being the number of correspondences), our approach minimizes an algebraic error function over 3 variables only. As indicated in Figure 2, the smooth cost function typically shows only a small number of geometrically meaningful local minima in the neighbourhood of the globally optimal rotation. By including simple mechanisms to avoid local minima (e.g. random variation of the starting point), a direct optimization hence becomes possible under the practically valid assumption that the rotation between the view-points is bounded.

Local minima are rotations that cause a disparity similar to the one generated by the true camera displacement. As illustrated in Figure 3, a purely translational displacement parallel to the image plane can cause a similar disparity than a pure rotation around an orthogonal axis in the image plane, and vice-versa. Small translations w.r.t. the depth of the points can cause a local minimum to be very close to the global one. These statements are of course only valid for pinhole-like cameras. We obtain a large basin of attraction around the global minimum in the case of omni-directional bearing vector distributions, which is why our algorithm is particularly well suited for omni-directional cameras.

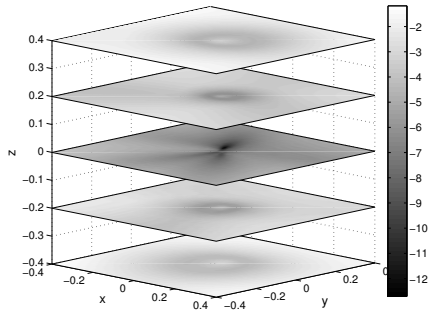


Figure 2. Behavior of the cost function $\lambda_{\mathbf{M},min}$ around the global minimum $(0, 0, 0)$ for a random experiment. For better viewing convenience, the figure plots the log-value of $\lambda_{\mathbf{M},min}$.

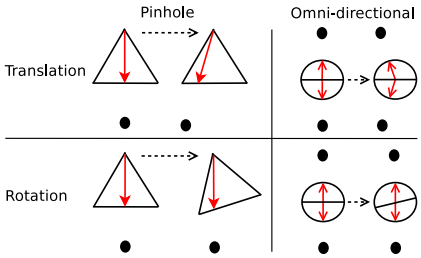


Figure 3. Similarity and dissimilartiy of disparities caused by different motion and different camera types.

4.3. Efficiency and noise resilience

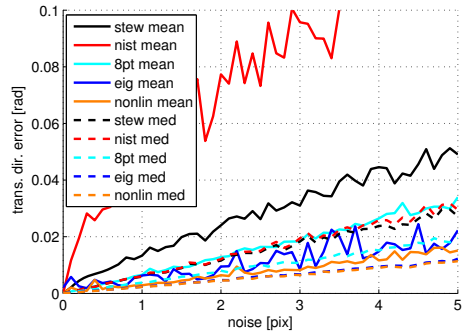
In this experiment, we use 10 random points without outliers and add different levels of noise. We select 10 points for a better conditioning of the cost function. In order to still have a fair comparison, each algorithm is using all 10 points (except (kneip)). The starting value for our iterative eigenvalue-based solver as well as non-linear optimization is set by a uniform variation of the true rotation [2]. Figure 4 shows that our parametrization induces lower errors, and clearly outperforms all other solutions for all tested noise levels. Moreover, note that—for all reference algorithms—the best essential matrix from multiple solutions as well as the correct rotation and translation from the essential matrix decomposition is each time selected based on a comparison to ground truth, whereas our algorithm simply delivers a unique solution.

The median execution times per iteration are 105 μs (stew), 248 μs (nist), 943 μs (kneip), 26 μs (7pt), 26 μs (8pt), and 83 μs (eig), respectively. Even though the execution time of our iterative solver depends on the number of iterations in the optimization, we at least note that it is real-time compliant. On the other hand, we were not able to reproduce the advantage in terms of computational efficiency of [16] commonly reported in the literature, which at least points out the complexity of a proper implementation of [16].

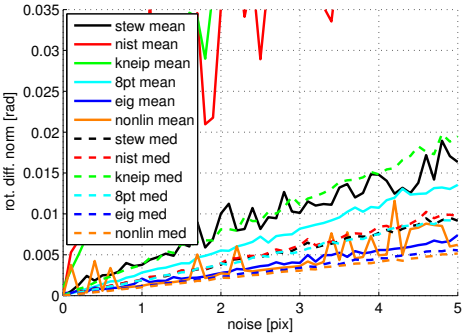
4.4. Convergence and resilience to outliers

In this experiment, we use 100 random points and introduce up to 20% outliers into the measurements. The noise is fixed to 0.5 pixels. All algorithms are embedded into a RANSAC [3] scheme using the same outlier threshold and inlier criterium. In order to exploit the commonly reported advantage of minimal solvers, we use the minimum number of points for all reference algorithms. For the essential-matrix based solutions, a couple of additional points are used in each iteration in order to automatically disambiguate the multiple solutions. We leave the sample size for our iterative solver at 10. We try to avoid local minima by starting at identity rotation, and adding a random variation in each RANSAC iteration.

[2] A maximum deviation of 0.01, which ensures that we spot the global minimum for a proper evaluation of noise resilience



(a)



(b)

Figure 4. Translation (a) and rotation (b) error for different noise levels and algorithms, each time averaged over 1000 random problems. Each algorithm uses the same 10 points in each iteration, resulting in lowest errors for our iterative solver. (eig) has best performance with an error similar to two-view bundle adjustment. Note that (kneip) computes only the rotation, and that (7pt) falls back to (8pt) for non-minimal sets of correspondences.

As outlined in Figure 5, the best hypothesis of our iterative solver is again outperforming alternative solutions, this time even clearer since using the biggest sample size. This also proves the ability of our iterative random variation scheme to find the global minimum. As shown in Figure 6(a), the number of iterations is not affected by our random variation scheme. This is, on one hand, explained by the fact that there are only few local minima—meaning low probability to miss the global minimum—and, on the other hand, by the bigger sample size[3]. The best sample size is not always minimal, but results from a trade-off between outlier fraction and noise. Our solver requires the smallest number of iterations below 5% outliers. Using more

[3] For 10 inlier correspondences and differing random displacements for both ground truth and initial transformation, our approach finds the global minimum in roughly 40% of all cases, compared to 20% for two-view bundle adjustment. If chosing the best out of 5 random trials for each experiment, our approach finds the global minimum in 86% of all cases, compared to 68% for two-view bundle adjustment.

features cancels noise, and thus increases the probability to find the correct inliers. This is underlined by Figure 6(b), showing that our iterative solution essentially finds the correct amount of inliers. Figure 6(c) shows that, again, the entire sample consensus scheme has state-of-the-art computational efficiency.

5. Direct frame-to-frame visual odometry

In order to demonstrate the performance of our algorithm on real images, we finally applied it to a video sequence captured with a global shutter WVGA camera. We compare the relative rotation accuracy of our approach and [16] against ground truth data delivered by a Vicon motion capture system. Both algorithms are applied on a frame-to-frame basis, and we employ fast corner extraction and patch matching in order to establish feature correspondences. Bearing vectors are created by applying a standard pinhole-camera model with radial distortion parameters. The final result of our approach is refined by solving (2) over all inliers, and using the best RANSAC hypothesis as a starting point for the optimization. Moreover, the summation terms in Section 2.4 are weighted by the corresponding feature matching scores. For the approach in [16], the final result is refined by running standard two-view bundle adjustment over all inliers.

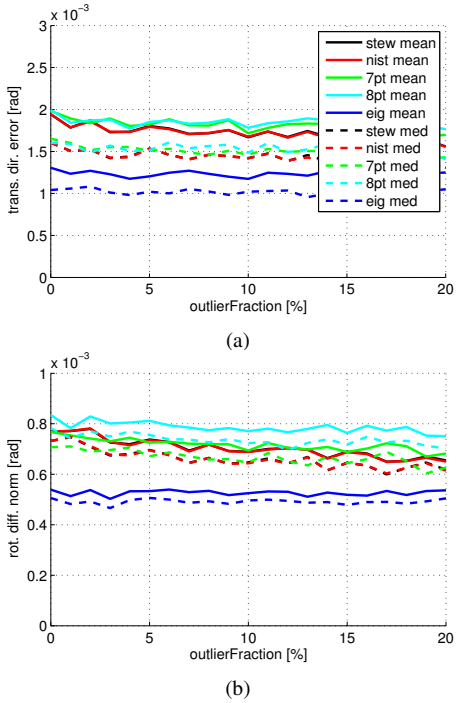


Figure 5. Accuracy of the different algorithms embedded into a RANSAC scheme and with different outlier percentages. Each value is averaged over 1000 random experiments. All algorithms use the same 100 points in each iteration. The error of the iterative solver remains lowest.

The dataset starts with moderate translational displacements, then contains rotations about all three camera axes, and finally combined rotation and translation. Errors are mainly caused by inhomogeneous bearing vector distributions in the pinhole-camera case, resulting in the previously mentioned ambiguities between rotational and translational motion estimation. As indicated in Figure 7, the relative rotation accuracy of our approach outperforms the approach in [16], with an error in rotation staying below 1° per pair of frames. This at the same time proves that our random variation scheme effectively avoids local minima. More exhaustive motion estimation results of our real-time pipeline can be found in our supplementary video file.

6. Conclusion

The present paper introduced a novel paradigm for solving the relative pose between two calibrated images, which consists of iterative enforcement of the coplanarity of epipolar plane normal vectors. In contrast to many approaches in the literature, the optimization is carried out directly over the frame-to-frame rotation, meaning three parameters only. The cost function shows only a small number of geometrically meaningful local minima besides the globally optimal solution. This leads to an elegant direct optimization without the requirement for any sophisticated initialization procedures. The paper demonstrates that this simple approach can do equally well if not better than all commonly used

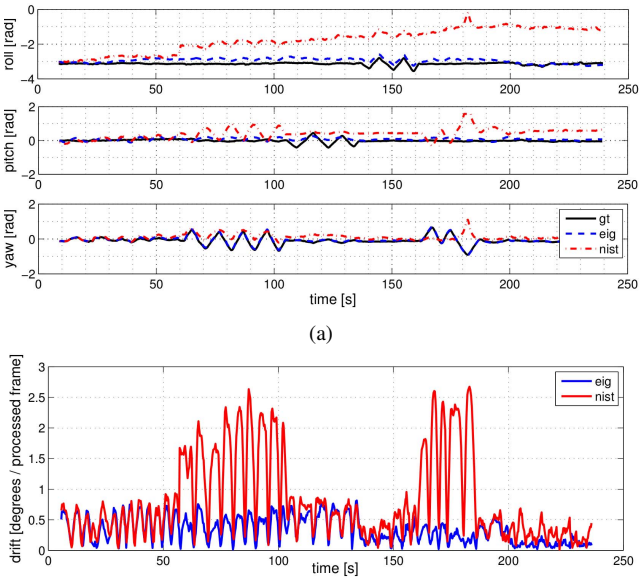
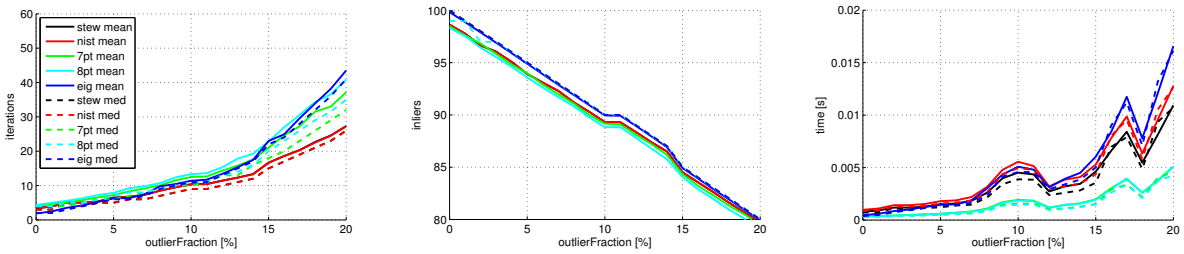


Figure 7. Rotation over time (a) and relative rotation accuracy (b) of our solution (eig) and [16] (nist) over a real image sequence. The errors are computed with respect to ground truth data (gt) obtained from a Vicon motion capture system.



(a) (b) (c)

Figure 6. Number of iterations (a), number of found inliers (b), and required execution time (c) for all tested algorithms.

essential-matrix-based approaches, which all depend on additional heuristics in order to come up with a unique solution. High computational efficiency is achieved by keeping the iteration time constant, independently of the number of features.

ACKNOWLEDGMENT

The research leading to these results has received funding from ARC grants DP120103896 and DP130104567, and the EU project FP7-269916 (*Vcharge*).

References

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

[2] F. Dopico, J. Moro, and J. Molera. Weyl-type relative perturbation bounds for eigensystems of hermitian matrices. *Linear Algebra and its Applications*, 309(1):3–18, 2000.

[3] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] R. Hartley. In Defense of the Eight-Point Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(6):580–593, 1997.

[5] R. Hartley and F. Kahl. Global Optimization through Rotation Space Search. *International Journal of Computer Vision (IJCV)*, 84(1):64–79, 2009.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, second edition, 2004.

[7] U. Helmke, K. Hüper, P. Y. Lee, and J. Moore. Essential Matrix Estimation Using Gauss-Newton Iterations on a Manifold. *International Journal of Computer Vision (IJCV)*, 74(2):117–136, 2007.

[8] L. Kneip, R. Siegwart, and M. Pollefeys. Finding the exact rotation between two images independently of the translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Florence, Italy, 2012.

[9] E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitzgsber. Akad. Wien, Math. Naturw. Abt., IIa.*, 122:1939–1948, 1913.

[10] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial Eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Proceedings of the British Machine Vision Conference (BMVC)*, Leeds, UK, 2008.

[11] H. Li and R. Hartley. Five-point motion estimation made easy. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 630–633, 2006.

[12] J. Lim, N. Barnes, and H. Li. Estimating relative camera motion from the antipodal-epipolar constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(10):1907–1914, 2010.

[13] H. Longuet-Higgins. *Readings in computer vision: issues, problems, principles, and paradigms*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[14] Y. Ma, J. Košecká, and S. Sastry. Optimization Criteria and Geometric Algorithms for Motion and Structure Estimation. *International Journal of Computer Vision (IJCV)*, 44(3):219–249, 2001.

[15] H.-H. Nagel. On the derivation of 3-d rigid point configurations from image sequences. In *Proceedings of the IEEE Conference on Pattern Recognition and Image Processing*, Dallas, USA, 1981.

[16] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26(6):756–777, 2004.

[17] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–659, Washington, DC, USA, 2004.

[18] H. D. Stewénius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006.

[19] H. D. Stewénius, C. Engels, and D. Nistér. An Efficient Minimal Solution for Infinitesimal Camera Motion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, USA, 2007.

[20] P. Torr, A. Fitzgibbon, and A. Zisserman. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 485–491, Bombay, India, 1998.

[21] X. Zhuang, T. S. Huang, A. N., and R. M. Haralick. A simplified linear optic flow-motion algorithm. *Computer Vision, Graphics, and Image Processing*, 42(3):334–344, 1988.