DGK Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft 785

Martin Reich

Global image orientations from pairwise relative orientations

München 2016

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5197-3

Diese Arbeit ist gleichzeitig veröffentlicht in:

Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover

ISSN 0174-1454, Nr. 328, Hannover 2016

Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 785

Global image orientations from pairwise relative orientations

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation

von

M. Sc. Martin Reich

München 2016

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0065-5325 ISBN 978-3-7696-5197-3

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 328, Hannover 2016

Adresse der der DGK:

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail post@dgk.badw.de ● http://www.dgk.badw.de

Prüfungskommission

Vorsitzender: Prof. Dr.-Ing. habil. Christian Heipke
Korreferenten: Prof. Dr. Cyrill Stachniss
Prof. Dr.-Ing Ingo Neumann

Tag der mündlichen Prüfung: 06.09.2016

ISSN 0065-5325 ISBN 978-3-7696-5197-3

Erklärung

Ich erkläre, dass ich die vorliegende Dissertation selbständig verfasst habe, die benutzten Hilfsmittel vollständig angegeben sind und die Dissertation nicht als Diplomarbeit, Masterarbeit oder andere Prüfungsarbeit verwendet wurde. Weiterhin erkläre ich, dass ich keine anderen Promotionsgesuche eingereicht habe.

Hannover, 6. Juni 2016

Statement

I state that this dissertation has been written entirely by myself. No further sources besides the ones noted in the bibliography have been used and this dissertation has not been submitted as Diploma thesis, Master thesis or any other written examination. Furthermore, I state that I have not applied for any other conferral of a doctorate.

Hannover, 6th June 2016

Abstract

Image orientation, also called pose estimation in disciplines other than classical photogrammetry, has been a key task in photogrammetry and computer vision for quite a time. Recently, autonomous systems, for instance, and the availability and ease of capturing images led to new applications, for which the knowledge about the camera's position and attitude is important. With this diversity of applications, a novel concept for solving the orientation problem up to a final bundle adjustment evolved: *global image orientation*. Approaches that follow this concept aim to improve relevant issues of traditional sequential methods for the computation of initial orientation parameters, mainly the necessity of intermediate bundle adjustment and the consequent inefficiency, as well as the distribution of uncertainties, which is unfavorable, due to the sequential block enlargement.

In global methods, the orientation parameters of all images are estimated simultaneously based on pairwise relative information. This is a prerequisite for a more realistic distribution of uncertainties. The problem is simplified by a partition into three estimation steps, separated by the *type* of parameters: rotations, translations and object coordinates of homologous points. This order is generally fixed. A serious challenge in most practical applications are outliers in both, the image coordinates of homologous points and resultant relative orientations.

This thesis addresses the formulation of a new global image orientation approach. Focusing on the first two steps, the robust estimation of rotations and translations, this approach combines different strategies from related work and provides important extensions regarding an effective outlier elimination. *Firstly*, relative orientations are estimated linearly based on homologous points using a random sampling approach and improved in a reweighted nonlinear Gauss-Helmert-model. *Secondly*, in order to detect inconsistent relative rotations, a graph-based breadth propagation algorithm is developed, which sequentially identifies erroneous relative rotations based on the cycle constraint of rotations. In the *third* stage, global rotations are estimated by a convex semidefinite program (SDP) and a subsequent Lie algebraic averaging. For the SDP, both the objective function and the feasible region of the solution are relaxed. The Lie algebraic averaging uses this solution as initialization for a weighted iterative estimation in the corresponding tangent space. *Finally*, translations are estimated using point tracks, i.e. a set of points, of which in each of the images at least one is visible. These points are selected based on a novel strategy. Together with estimated rotations these points are used to formulate linear constraints for the estimation of translations. Regarding outliers in the image coordinates of the homologous points, two additional conditions are established, one based on the reprojection error of local pairwise reconstructions, the other using a redundant scale estimation from image triplets.

For the analysis and validation of the method, experiments are conducted on synthetic data and real images. Synthetic data is used to evaluate the accuracy and robustness of the approach. Experiments on real images include four different benchmark datasets that are used to assess the proposed method with respect to recent state-of-the-art models. Moreover, large image sets from image-hosting websites are processed in order to demonstrate the scalability of the method. This also includes images with many different camera models, for which only a rough guess about their interior orientation is available. Results show that the proposed approach achieves accurate results, partly more precise than comparable state-of-the-art models, is very robust against outliers and is applicable to various kinds of image data.

Zusammenfassung

Die Orientierung von Bildern, in Bereichen außerhalb der klassischen Photogrammetrie auch Posenschätzung genannt, ist seit langem eine der Hauptaufgaben in der Photogrammetrie und im Bereich der Computer Vision. Seit dem Übergang zur digitalen Fotografie haben unter anderem die hohe Verfügbarkeit und einfache Erfassung von Bildern sowie die Entwicklung autonomer Systeme dazu geführt, dass immer mehr Verfahren auf die Bestimmung der Position und Ausrichtung der Kamera aufbauen. Damit einhergehend entstand ebenfalls ein neues Konzept zur Lösung der Bildorientierung, genau genommen eine Bestimmung von Näherungswerten für eine abschließende Bündelblockausgleichung, die *globale Bildorientierung*. Globale Orientierungsansätze zielen darauf ab, bestimmte Nachteile der sequentiellen Methoden zu überwinden, insbesondere, dass diese Methoden von einer mehrfach durchzuführenden Bündelblockausgleichung abhängen und als Konsequenz relativ ineffizient sind und dass die Unsicherheiten durch den sequentiellen Aufbau des Modells nicht gleich über alle Bilder verteilt sind.

In globalen Methoden werden die Orientierungsparameter aller Bilder gleichzeitig basierend auf paarweiser relativer Information geschätzt. Dies ermöglicht eine günstigere Verteilung der Unsicherheiten. Das Problem wird vereinfacht und auf drei einzelne Schritte verteilt, welche sich unterschiedlichen *Typen* von Parametern widmen: Rotationen, Translationen und Objektkoordinaten der homologen Punkte. Diese Reihenfolge während der Schätzung ist im Allgemeinen fest. Eine wesentliche Herausforderung in den meisten Anwendungen stellt die Detektion und Elimination von Ausreißern dar, sowohl in den Bildkoordinaten der homologen Punkte als auch in den daraus resultierenden relativen Orientierungen.

Diese Doktorarbeit addressiert die Formulierung eines neuen globalen Bildorientierungsansatzes, mit dem Fokus auf der robusten Schätzung von Rotationen und Translationen, welcher verschiedene Strategien aus existierenden Ansätzen kombiniert und um wichtige Funktionen zur wirkungsvollen Ausreißerdetektion erweitert. Zunächst werden relative Orientierungen linear mit Hilfe von RANSAC geschätzt und in einem gewichteten Gauß-Helmert-Modell iterativ verbessert (1). Um Ausreißer in den relativen Rotationen zu detektieren, wird danach ein neues, graphenbasiertes Verfahren in Form einer Breiten-Propagierung entwickelt, das sequentiell fehlerhafte relative Rotationen basierend auf ihrer Schleifenbedingung identifiziert (2). Im dritten Teil werden globale Rotationen zunächst in einem konvexen Semidefiniten Optimierungsproblem (SDP) geschätzt und anschließend in einem Mittelungsalgorithmus in der Lie Algebra der Rotationen verbessert (3). Für das SDP werden sowohl die Zielfunktion als auch die zulässige Lösungsmenge relaxiert; die Lösung dient dann der Initialisierung der nachfolgenden iterativen und gewichteten Schätzung im Tangentialraum der Rotationen, der Lie Algebra. Schließlich werden die Translationen mit Hilfe von den geschätzten Rotationen und spezieller Punkte geschätzt, von denen in jedem der Bilder mindestens einer zu sehen ist (4). Für die Auswahl dieser Punkte wird eine neue Strategie vorgestellt. Um die Ausreißer in den Bildkoordinaten der homologen Punkten zu eliminieren, werden zwei zusätzliche Bedingungen aufgestellt, die eine basierend auf dem Rückprojektionsfehler in einer lokalen paarweisen Rekonstruktion, die andere unter Verwendung einer redundanten Maßstabsschätzung für Bildtripel.

Zur Analyse und Validierung der Methode werden Experimente basierend auf synthetischen Daten und realen Bildern durchgeführt. Synthetische Daten dienen der Evaluierung der Genauigkeit und Robustheit des Ansatzes. Die Experimente auf den realen Bildern beinhalten vier unterschiedliche Benchmark-Datensätze, mit Hilfe derer die entwickelte Methode mit derzeitigen State-of-the-art-

Modellen verglichen wird. Weiterhin werden große Bilddatensätze von speziellen Bild-Webseiten prozessiert, um die Skalierbarkeit der Methode zu demonstrieren. Dabei spielen Daten, bei denen verschiedene Kameramodelle verwendet wurden, deren innere Orientierung nur sehr grob bekannt ist, ebenfalls eine Rolle. Die Ergebnisse zeigen, dass der vorgestellte Ansatz im Vergleich zu State-of-the-Art-Methoden zum Teil präzisere Orientierungen liefert, sehr robust gegenüber Ausreißern ist und auf verschiedende Arten von Bilddaten anwendbar ist.

Notation and table of symbols

General notation

a, b, α, β, X, Y	Scalars
a, **b**, **d**, **u**, **x**, **y**	Vectors
A, **B**, **C**, **X**, **Y**, **Z**	Matrices
$\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{T}$, $\mathcal{W}$	Sets
A, B, C, D, T, W	Items of sets
$\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$, $\mathbb{N}$, $\mathbb{R}$, $\mathbb{Z}$	Spaces

Image orientation

$\mathcal{I}$	set of images
I_i	image i
$\mathbf{R}_i$, $\mathbf{t}_i$	global rotation and translation of I_i
$\mathbf{R}_{ij}$, $\mathbf{t}_{ij}$	relative rotation and translation between images $\{\mathsf{I}_i, \mathsf{I}_j\}$
P	set of points
$\mathbf{P}^l$	point l in object coordinates
$\mathbf{p}_i^l$	image coordinates of point l in I_i
$\mathcal{R}$	set of global rotations estimated in breadth-propagation
$\mathcal{R}^\star$	set of global rotations estimated in SDP
$\mathbf{R}_i^\star$	global rotation of I_i estimated in SDP
$\mathcal{R}^{\star\star}$	set of global rotations estimated in SDP and Lie algebraic averaging
$\mathbf{R}_i^{\star\star}$	global rotation of I_i estimated in SDP and Lie algebraic averaging
$\mathcal{R}^{BA}$	set of global rotations after bundle adjustment
$\mathbf{R}_i^{BA}$	global rotation of I_i after bundle adjustment
$\mathcal{T}^{\star\star}$	set of estimated global translations
$\mathbf{t}_i^{\star\star}$	estimated global translation of I_i
$\mathcal{T}^{BA}$	set of global translations after bundle adjustment
$\mathbf{t}_i^{BA}$	global translation of I_i after bundle adjustment
$\mathbf{R}_{ij}^0$, $\mathbf{t}_{ij}^0$	initial relative rotation and translation between images $\{\mathsf{I}_i, \mathsf{I}_j\}$
$\mathbf{R}_{ij}^\star$, $\mathbf{t}_{ij}^\star$	relative rotation and translation between images $\{\mathsf{I}_i, \mathsf{I}_j\}$ after M-estimation
$\mathbf{\Sigma}_{\mathbf{R}_{ij}^\star}$, $\mathbf{\Sigma}_{\mathbf{t}_{ij}^\star}$	covariance matrices of $\mathbf{R}_{ij}^\star$ and $\mathbf{t}_{ij}^\star$
$\mathbf{p}_{ij}$, $\mathbf{p}_{ij}^\star$	image coordinates of homologous points of image pair $\{\mathsf{I}_i, \mathsf{I}_j\}$ before and after M-estimation
$\tau_{\lvert\mathbf{p}\rvert}$	threshold for number of correspondences
$d_{chordal}$	chordal distance between two rotations
$d_{quaternion}$	quaternion distance between two rotations
d_α	angular distance between two rotations

Groups and manifolds

$O(3)$	orthogonal group of dimension 3
$SO(3)$	special orthogonal group of dimension 3
$\mathfrak{so}(3)$	Lie algebra to $SO(3)$
Q	quaternion sphere

Convex optimization

$f_0(\mathbf{x})$	objective function
$f_i(\mathbf{x})$, $g_j(\mathbf{x})$	inequality and equality constraints
n, m, p	dimension of parameter space, number of inequality and equality constraints
$\mathbf{dom}\, f$	domain of f; the subset of points, for which f is defined
$\nabla f(\mathbf{x_1})$	first order derivative of $f(\mathbf{x_1})$
$\mathbf{conv}(\mathcal{D})$	convex hull of set $\mathcal{D}$
$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$	Lagrange function
$g(\boldsymbol{\lambda}, \boldsymbol{\nu})$	Lagrange dual function

M-estimation relative orientations

$\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$	condition equation for M-estimation of relative orientation
$\mathbf{B}$	partial derivatives with respect to observations
$\mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$	Jacobian matrix for M-estimation of relative orientation
$\mathbf{v}$	vector of residuals
$\mathbf{x}$	vector of parameters $[t_{ij,x}, t_{ij,y}, t_{ij,z}, \omega_{ij}, \varphi_{ij}, \kappa_{ij}]$
$\mathbf{W}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$	weight matrix for M-estimation of relative orientation
$\boldsymbol{\Sigma}_{\mathbf{p}_{ij}}$	covariance matrix of image coordinates of the homologous points in images (i, j)
$\tau_{\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}$	threshold for convergence

View-graph

$\mathcal{G}$	view graph
$\mathcal{V}$	set of vertices
$\mathcal{E}$	set of edges
$\mathcal{E}^+$, $\mathcal{E}^-$	set of consistent and inconsistent edges
V_i	vertex corresponding to image i
V^s	starting vertex sth sequence
E_{ij}	edge between vertices (i, j)
$\mathcal{E}_i$	set of edges incident to V_i
$\mathcal{E}_i^s$	set of examined edges in sequence s incident to V_i
$\mathbf{d}^1$	vector of distances to starting vertex
$\mathbf{D}$	distance matrix
$\mathbf{A}$	adjacency matrix
$\mathcal{R}_j$	set of rotation estimates for V_j
$\mathcal{R}_j^+$, $\mathcal{R}_j^-$	set of consistent and inconsistent rotation estimates for V_j
τ_α	threshold for angular distance
τ_c	threshold for the ratio of consistent and inconsistent rotation estimates

SDP estimation

$\mathbf{M}$	Gramian matrix that includes a pairwise multiplication of rotation matrices
$\mathbf{M}^0$	matrix that includes the relative rotations
$\mathbf{M}_{ij}$	ijth 3×3 submatrix of $\mathbf{M}$
$d_{\mathbf{M}_{ij}}$	residual of $\mathbf{M}^0_{ij}$ / convex cost function in $\mathbf{M}_{ij}$
$w_{\mathbf{M}_{ij}}$	weights for each $\mathbf{M}_{ij}$

Lie algebraic averaging

$\Delta\mathbf{r}$	vector of residuals
$\mathbf{J}_{\Delta\mathbf{r}_{ij}}$	Jacobian matrix for Lie algebraic averaging
$\Delta\mathbf{r}^0$	vector of reduced relative rotations
$\Delta\mathbf{r}^{\star\star}$	vector of parameter offsets, corresponding to $\mathcal{R}^{\star\star}$
$\mathbf{W}_{\Delta\mathbf{r}}$	weight matrix for Lie algebraic averaging
λ, ν	parameters for weighting function

Translation estimation

$\tilde{\mathbf{t}}^{\star}_{ij}$	rotated relative translations, corresponding to $\mathbf{t}^{\star}_{ij}$
$\mathbf{P}^l_{ij}$	object coordinates of $\mathbf{P}^l$ intersected from $\{\mathsf{I}_i, \mathsf{I}_j\}$
s^l_{ij}	estimated scale number for the vector between the projection center of I_i and $\mathbf{P}^l_{ij}$ in the pair $\{\mathsf{I}_i, \mathsf{I}_j\}$
$\mathbf{R}_{\gamma^l_{ij}}$	matrix to rotate the relative translation $\mathbf{t}_{ij}$ to the vector between the projection center of I_i and $\mathbf{P}^l_{ij}$ by angle γ^l_{ij}
$\bar{\mathbf{p}}^{l,i}_{ij}$	projection of $\mathbf{P}^l_{ij}$ in image I_i
$\Delta\bar{\mathbf{p}}^l_{ij}$	average reprojection error of image pair $\{\mathsf{I}_i, \mathsf{I}_j\}$
τ_r	threshold for the reprojection error
$s^l_{\mathbf{t}_{ijk}}$	ratio between $\lVert\mathbf{t}_j - \mathbf{t}_i\rVert$ and $\lVert\mathbf{t}_k - \mathbf{t}_i\rVert$ using point $\mathbf{P}^l$
$\sigma_{\mathbf{s}^l_{\mathbf{t}_{ijk}}}$	standard deviation of $s^l_{\mathbf{t}_{ijk}}$ using point $\mathbf{P}^l$
$\mathbf{s}^l_{\mathbf{t}}$	vector including all $\sigma_{\mathbf{s}^l_{\mathbf{t}_{ijk}}}$ using point $\mathbf{P}^l$
$\mathbf{C}$	constraint matrix
$\mathbf{w}_{\mathbf{t}}$	vector of weights

Contents

1. Introduction

The determination of three-dimensional information from two-dimensional images that are overlapping in the pictured scene, has been a fundamental task in photogrammetry and computer vision for many decades. With the imaging process, depth information of the environment is lost. This depth information can only be restored with further information, e.g. in form of a picture of the same scene from a different perspective. Essential for this three-dimensional reconstruction is an at least implicit knowledge about the camera location and the viewing direction, i.e. the orientation or pose of the images that are involved. The goal of reconstruction from images dates back to at least the mid-nineteenth century when photography was invented. In the analogue and analytical eras of photogrammetry, airborne and terrestrial images were processed using e.g. comparators and analytical plotters in order to generate maps or object models. Since digital image sensors emerged by the end of the last century and with the increasing computational power in recent years, the field of photogrammetric applications widened. The digital revolution led to a plethora of images. With almost every modern cellphone having one (or even more) cameras, images can be seen as the most popular instance for mapping our environment. Naturally, most of these images are not primarily taken for the purpose of gathering three-dimensional information but taken together they have the ability for large-scale reconstructions as is shown exemplarily for Notre Dame de Paris in Figure 1.1. Also for measuring or navigation tasks, images are in most cases a valuable source of information because of various reasons: they are easy to acquire and easy to access and cameras are relatively lightweight and cheap, which for many applications are important criteria. Nowadays, the range of image based applications is large and growing: Cars use cameras for the driver's assistance, moving robots process captured image sequences for orientation and navigation and unmanned aerial systems are able to acquire images from almost all thinkable perspectives for an exhaustive mapping and modelling - only to name a few examples.

The increasing computational power allows for a more and more effective processing of the images. The line of approach in image orientation or reconstruction is the following: On the one hand, the scale of the reconstructed scene will become larger, which implies more images to be oriented and, on the other hand, reconstruction will become more accurate, for which, next to the number of images, the geometrical resolution is responsible. In either way, the computational demand is growing, which implies besides increasing resources a demand for more efficient solutions.

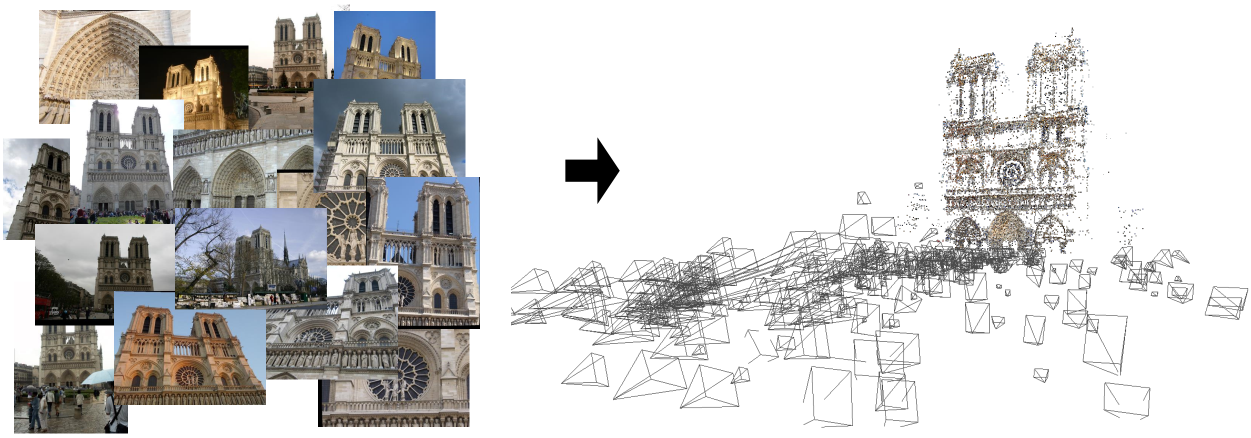


Figure 1.1.: Collection of images of Notre Dame de Paris [Snavely et al., 2006] and photogrammetric reconstruction using the method proposed in this thesis.

The functional relationship between observations in the image, image orientation parameters and object points is given by the nonlinear collinearity equations. The best linear unbiased estimation, i.e. a maximum-likelihood solution, is derived using bundle adjustment. Considering bundle adjustment as the gold standard for estimation, the image orientation task is reduced to finding proper initial values for this nonlinear and thus iterative optimization. Because bundle adjustment is nonlinear and nonconvex, the quality of the estimated parameters depends on the quality of these initial values. Thus, a pivotal precept is to determine initial values accurate enough for the bundle adjustment to converge to the global minimum.

A common strategy for the estimation of initial orientation parameters is a sequential block enlargement. Starting with a pair of images, initial values for object points and image orientation parameters are computed by successive spatial intersection and resection. To obtain a grip on accumulating errors and drift effects, intermediate bundle adjustments are necessary. Another way of computing an initialization is a hierarchical approach. This hierarchy concerns the number of images, thus, starting with individual subsets of images, a joint orientation is then derived by successive merging. The third and more recent possibility are global methods. These methods solve the image orientation jointly for all images and divide the estimation in the space of orientation parameters instead. In general, rotations are estimated first, followed by translations and object coordinates of homologous points. A division in this manner leads to a simplified solution space, concerning rotations, translations and points separately. This simplification of the solution space in the individual computation steps often allows for a convex estimation of the unknown parameters. In this way, a global optimal solution for the initial orientation is derived.

In this thesis, a novel global image orientation approach is proposed by a combination of several convex and nonconvex optimizations extended by a rigorous treatment of outliers. In the following section, this approach is motivated, its basic characteristics are outlined and the objective of this thesis is formulated.

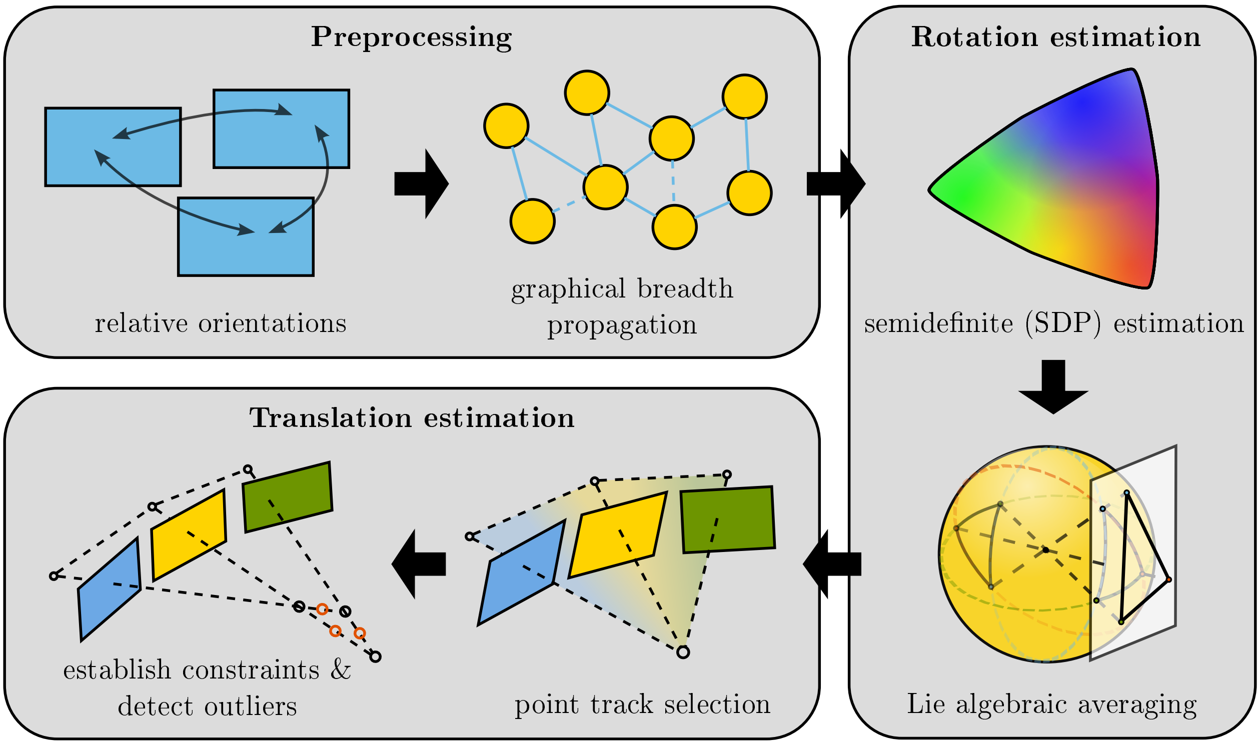


Figure 1.2.: Schematic illustration of the workflow.

1.1. Motivation and objective of the thesis

"There are several strategies that may be used to obtain the initial reconstruction, though this area is still to some extend a black-art."([Hartley & Zisserman, 2003], p. 435).

Given the fact that this quote is more than ten years old, one might hesitate and ask *is there still no gold standard method to compute the initial image orientation?* The basic problem of image orientation has been a central part in photogrammetric and computer vision research for a long time and many well working approaches evolved since then, competing with various kinds of image data. However, research in image orientation is still an ongoing matter. Encouraged by the large interest in research and the growing field of convex optimization, *in this thesis a novel robust global image orientation approach is developed, in which a series of convex and nonconvex optimizations is combined in a new way, delivering a set of accurate initial values for various types of image data.* Individual concepts from related work are captured, combined and extended in order to develop a versatile model for the estimation of image orientation parameters. The proposed approach is founded on a set of redundant relative orientations and can be structured into three major stages, a *preprocessing* step (1), improving the relative orientations, the *estimation of rotations* (2) and the *estimation of translations* (3). A schematic overview is outlined in Figure 1.2 and explained in the following.

During *preprocessing*, relative orientations are computed from image coordinates of homologous points, using the concept of the essential matrix and then by a subsequent iterative M-estimation in a Gauss-Helmert-model. This leads to a maximum likelihood solution and allows a rigorous propagation of variance information that is used in the subsequent stages to incorporate prior knowledge about the quality of relative orientations in form of individual weights. A major issue in related research is dealing with erroneous and inconsistent relative orientations (e.g. [Zach et al., 2010; Enqvist et al., 2011b; Hartley et al., 2011; Chatterjee & Govindu, 2013]). The authors distinguish between using a robust cost function like the L_1 norm and eliminating outliers before the estimation by applying heuristic constraints. A robust cost function limits the influence of outliers on the result, hence, there is no need for a two-step procedure like carrying out the elimination of outliers and the estimation separately. On the other hand, outliers remain in the set of observations and may distort the results. Considering a high accuracy, it is best to eliminate outliers in advance and apply an unbiased estimation afterwards.

In this thesis, the focus of *preprocessing* lies in the detection and elimination of inconsistent relative orientations. This requires an effective constraint to distinguish between inliers and outliers, which is formulated by a propagation of relative rotations along cycles in form of a new breadth-propagation algorithm in a graphical structure called view-graph. For this algorithm, convergence is proven and its effectiveness is evaluated and assessed.

The *estimation of rotations* addresses the computation of rotations for all images in a joint coordinate system from redundant preprocessed relative rotations. The original optimization problem is nonlinear and requires initialization. It is therefore common to relax the problem so that it becomes linear and directly solvable (e.g. [Govindu, 2001; Martinec & Pajdla, 2007]). Regarding the characteristics of rotations, the solution should be required to be inside the rotation manifold $SO(3)$. This manifold is a Lie group, which involves nonlinear constraints like a determinant of the rotation matrix being equal to 1. In order to apply this constraint to a linear optimization problem, it has to be relaxed as well (e.g. [Arie-Nachimson et al., 2012; Horowitz et al., 2014]). In general, the solution of the simplified problem does not meet the optimal solution of the original problem. Nonlinear iterative methods, like the one proposed by Govindu [2004], find a maximum likelihood solution in the Lie algebra, the tangent space of $SO(3)$, and are often more accurate.

The approach presented in this thesis combines the two concepts. Initial rotations are computed with the convex semidefinite program (SDP) developed in Saunderson et al. [2014], requiring the solution to be in the convex hull of $SO(3)$, and, subsequently, improved by an iterative Lie algebraic averaging similar to Govindu [2004]. Both optimizations are assisted by a covariance based weighting of the involved relative rotations.

For the *estimation of translations*, the estimated rotations and relative translations are considered. Several approaches address this problem, they comprise quasiconvex L_∞ estimation (e.g. [Hartley & Schaffalitzky, 2004; Kahl, 2005]) and linear optimization from pairwise or tripletwise constraints (e.g. [Arie-Nachimson et al., 2012; Cui et al., 2015]). All these methods have certain characteristics, which have to be taken into account: The drawback of L_∞-optimization is its sensitivity to outliers. Moreover, the maximum residual is minimized, which often does not lead to the desired outcome.

Using pairwise constraints may lead to problems for collinear images (images taken along a linear path), whereas models with triplet constraints require a higher overlap between the images.

In this thesis, the linear model of Cui et al. [2015] is used to estimate translations, in which constraints for image triplets based on pairwise local reconstructions of points are formulated. It produces accurate results, while not being demanding regarding the overlap between images. The constraints themselves are sensitive to outliers in the homologous points and relative translations. Thus, the model is extended by a new outlier detection scheme using pairwise and tripletwise constraints and a new algorithm to select suitable points taking also the distribution in image space into account.

In summary, this thesis pursues the following objective: *Solve the image orientation problem by a novel global image orientation approach that achieves accurate orientation parameters, is robust against outliers in relative orientations and image coordinates of the homologous points and is applicable to various types of image data.* A focus hereby lies on a convex formulation of the involved optimizations. The task of this thesis is to present an extensive study of this method and to evaluate the proposed approach in order to show that all characteristics are accomplished.

1.2. Reader's guide

This thesis is structured as follows. Chapter 2 is dedicated to give some basic information about three major topics, important for this thesis: image orientation parameters, mathematical groups and manifolds, and convex optimization. In the first part, important terminology and different parameterizations of image orientation parameters are introduced. These are put into context to Lie groups in the second part. The third part familiarizes the reader with fundamental aspects of convex optimization like the convexity of functions and sets and strategies for solving such problems. In Chapter 3, state-of-the-art approaches for the estimation of image orientation parameters are reviewed. A brief review of the estimation of relative orientations is followed by a study of existing approaches for sequential, hierarchical and global image orientation. Chapter 4 addresses a detailed presentation of the new method. After a brief introduction including a comparison of the objective of this thesis to the state-of-the-art, the three fundamental steps of preprocessing, estimation of rotations and estimation of translations are explained. The last part of this chapter comprised information about the bundle adjustment used for a final optimization of the orientation parameters. An exhaustive evaluation of the proposed method is given in Chapter 5 using synthetic data and real images. Results for the accuracy, robustness and applicability, as well as limitations are shown and discussed. Finally, Chapter 6 draws conclusions and prospects for future research.

2. Basics

In this chapter, basic concepts about the *image orientation parameters* (2.1), *groups and manifolds* (2.2) and *convex optimization* techniques (2.3) are explained. The first section gives an insight into the mathematical description of the image orientation problem, which is the main focus of this dissertation. The reader is introduced to the concept of *relative* and *global* image orientation. Different parametric representations of the orientation parameters and transitions between these are discussed. The second section introduces algebraic groups, in particular the *rotation group*, i.e. the *special orthogonal group* $SO(3)$. Using this concept, the algebraic group properties are united with the geometric properties of a differentiable manifold, which plays a key role for the estimation of rotations. In the third part, basic information is given about *convex optimization problems*, how they can be formulated and solved.

2.1. Image orientation parameters

In this section, most important parameters related to image orientation are explained. The term *image* in this work is defined as the perspective transformation of a scene in the 3D object space to the 2D image space captured simultaneously through a lens with a unique projection center. Thus, all projection rays intersect in a single point, which is a simplification of the real imaging process. This definition includes still images as well as frames from image sequences.

This work deals with the estimation of image orientation parameters, which shall be defined as well. In general, one distinguishes between two different types of orientation, *interior* and *exterior* orientation, while both combined describe the geometry of the imaging process. *Interior orientation* parameters, also known as intrinsic parameters, model the geometry of the camera. A simple parameterization includes the *calibrated focal length* c, also called camera constant, i.e. the distance between the projection center and the image plane, and the intersection of the optical axis with the image plane, called the *principal point*, $\mathbf{h} = (h_x, h_y)$. This set can be enlarged by various parameters that describe the deviation from the perspective mapping, e.g. distortion coefficients of the lens. In the remainder of this work, all interior orientation parameters are assumed to be known and constant if not stated otherwise. More details on interior orientation and lens distortion can be found in Kraus [1993] (chapter 3.1) or in Brown [1971]..

The focus of this work lies on the *exterior orientation*, which is also known as pose or extrinsic parameters. The exterior orientation describes the location of the projection center and the viewing direction with respect to an (arbitrary) object coordinate system and comprises six parameters, three translations and three rotations, respectively. In the following, the terms *translation* and *location* will be used synonymously to describe the translation parameters between the origin of the object coordinate system and the location of the projection center.

2.1.1. Single image geometry

It is assumed that for an image I_i the translation is given by $\mathbf{t}_i \in \mathbb{R}^3$ and the rotation is described by the matrix $\mathbf{R}_i \in SO(3) \subseteq \mathbb{R}^{3\times 3}$ (the term $SO(3)$ will be introduced in Section 2.2). Further, let $\mathcal{V} = \{\mathsf{V}_1 \dots \mathsf{V}_n\}$ with $\mathsf{V}_i = \{\mathbf{t}_i, \mathbf{R}_i\}$ be the set of exterior orientations for a set of n images. In every image I_i, a set of object points is observed with $\mathbf{p}_i^l = \left[p_{i,x}^l, p_{i,y}^l\right]^T$ being the 2-dimensional observation of point $\mathbf{P}^l = \left[P_x^l, P_y^l, P_z^l\right]^T$ in the image coordinate system of image I_i. The mapping $H_o^i : \mathbf{P}^l \to \mathbf{p}_i^l$ maps $\mathbf{P}^l$ from the object coordinate system to the image coordinate system of image I_i and is given by the *collinearity equations*[1]

$$\begin{aligned} p_{i,x}^l &= h_x - c\frac{R_{i,11}(P_x^l - t_{i,x}) + R_{i,21}(P_y^l - t_{i,y}) + R_{i,31}(P_z^l - t_{i,z})}{R_{i,13}(P_x^l - t_{i,x}) + R_{i,23}(P_y^l - t_{i,y}) + R_{i,33}(P_z^l - t_{i,z})} \\ p_{i,y}^l &= h_y - c\frac{R_{i,12}(P_x^l - t_{i,x}) + R_{i,22}(P_y^l - t_{i,y}) + R_{i,32}(P_z^l - t_{i,z})}{R_{i,13}(P_x^l - t_{i,x}) + R_{i,23}(P_y^l - t_{i,y}) + R_{i,33}(P_z^l - t_{i,z})} \quad . \end{aligned} \tag{2.1}$$

A simpler form of these equations is derived using homogeneous coordinates, i.e. $\mathbf{p}_i^l = \left[p_{i,x}^l, p_{i,y}^l, 1\right]^T$ and $\mathbf{P}^l = \left[P_x^l, P_y^l, P_z^l, 1\right]^T$

$$\mathbf{p}_i^l = \mathbf{K}\mathbf{R}_i^T \left[\mathbf{I}_{3\times 3}| - \mathbf{t}_i\right] \mathbf{P}^l \quad , \tag{2.2}$$

with $\mathbf{K}$ being the *calibration matrix* of the form

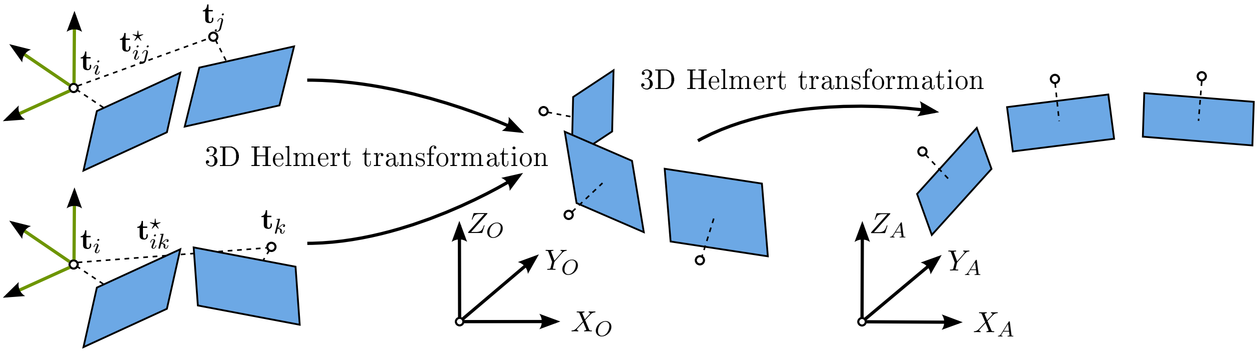
$$\mathbf{K} = \begin{bmatrix} c & 0 & h_x \\ 0 & c & h_y \\ 0 & 0 & 1 \end{bmatrix} \quad . \tag{2.3}$$

The collinearity equations (Equations (2.1) and (2.2)) include a perspective projection from 3D to 2D space, which is not bijective due to the loss of explicit depth information of the projected points.

2.1.2. Two-view geometry

The exterior orientation of two images $\{\mathsf{I}_i, \mathsf{I}_j\}$ is determined by twelve independent parameters $\{\mathsf{V}_i, \mathsf{V}_j\}$. In the line of argument, it is necessary to distinguish *global* and *relative* orientation. The *global orientation* is determined up to a seven parameter *3D Helmert* transformation, which

[1] In general, observations in digital images are given in the sensor coordinate system. The transformation between image coordinates and sensor coordinates is bijective and can e.g. be found in Mugnier et al. [2004], pp. 217.



(a) relative orientation (b) global orientation (c) absolute orientation

Figure 2.1.: Relations between different types of image orientation. Each relative orientation (2.1a) can be transformed to a global orientation (2.1b) by a 3D Helmert transformation. Likewise, the relation between the global orientation and the absolute orientation (2.1c) is a 3D Helmert transformation.

allows a mapping into every other 3-dimensional Euclidean coordinate system. Thus, the *relative orientation*, describing the orientation of one image with respect to the other, is parameterized by five independent parameters. Using the term *global* instead of *absolute*, which is the common terminology in the photogrammetric literature, clarifies that the orientation is computed in an arbitrary object coordinate system that is not necessarily a superordinate coordinate system. The relation between the different coordinate systems is depicted in Figure 2.1.

The *relative rotation* $\mathbf{R}_{ij}$ is defined as

$$\mathbf{R}_{ij} = \mathbf{R}_i^T \mathbf{R}_j. \tag{2.4}$$

Because only directions are measured in images, the scale cannot be estimated, assuming the absence of further information, which is why the *relative translation* between images I_i and I_j, $\mathbf{t}_{ij} = \mathbf{t}_j - \mathbf{t}_i$ is given only up to scale and comprises just a direction.

In the literature, one can find two different parameterizations of the relative orientation (e.g. [Mugnier et al., 2004] pp. 252), independent models and dependent-images. In the remainder of this work, the *dependent-images* parameterization is used. It is assumed that the local coordinate system is situated in the projection center of image I_i, hence $\mathbf{t}_i = \mathbf{0}$ and $\mathbf{R}_i = \mathbf{I}_{3\times 3}$ and that the base has unit length, hence $\|\mathbf{t}_{ij}\| = 1$.

The parameter set $\{t_{ij,x}, t_{ij,y}, t_{ij,z}, \omega_{ij}, \varphi_{ij}, \kappa_{ij} : \|\mathbf{t}_{ij}\| = 1\}$ can be determined from at least five corresponding observations using the *coplanarity constraint*[2], derived from Equation (2.2):

$$\mathbf{R}_j\,{}^n\mathbf{p}^l_j \cdot \left(\mathbf{t}_{ij} \times \left(\mathbf{R}_i\,{}^n\mathbf{p}^l_i\right)\right) = 0, \tag{2.5}$$

with the normalized observation ${}^n\mathbf{p}^l_i = \mathbf{K}^{-1}\mathbf{p}^l_i$. This constraint requires the base vector $\mathbf{t}_{ij}$ and

[2]Here, the Euler angles representation is used (see Section 2.1.3), any other representation is also possible.

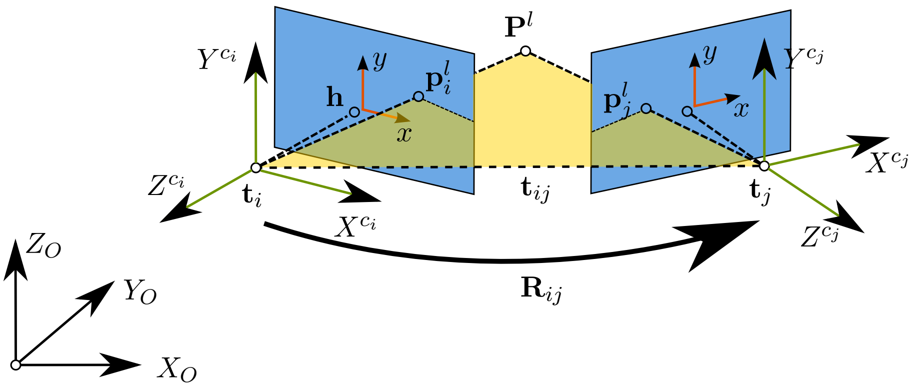


Figure 2.2.: Geometry of exposure: The camera coordinate system $[X^c, Y^c, Z^c]$ (green) is located in the projection centers of images I_i and I_j. The relation between the camera coordinate systems and the object coordinate system $[X_O, Y_O, Z_O]$ (black) is given by the global orientations $\{\mathbf{t}_i, \mathbf{R}_i\}$ and $\{\mathbf{t}_j \mathbf{R}_j\}$. The image plane lies at a distance of the camera constant c in the negative Z^c-direction and the optical axis intersects the image plane at the principal point $\mathbf{h}$. Object point $\mathbf{P}^l$ is measured at $\mathbf{p}_i^l$ and $\mathbf{p}_j^l$ in the image coordinate system $[x, y]$ (red). The relative orientation $\{\mathbf{t}_{ij}, \mathbf{R}_{ij}\}$ describes the transformation between $[X^{c_i}, Y^{c_i}, Z^{c_i}]$ and $[X^{c_j}, Y^{c_j}, Z^{c_j}]$. The epipolar plane $\{\mathbf{P}^l, \mathbf{t}_i, \mathbf{t}_j\}$ is depicted in yellow.

the two vectors pointing from each projection center of the two images to the object point ($\mathbf{R}_i\, {}^n\mathbf{p}_i^l$ and $\mathbf{R}_j\, {}^n\mathbf{p}_j^l$) to lie in the same plane, the so-called *epipolar plane*. The geometry of exposure of the two-view case is pictured in Figure 2.2.

Using the skew symmetric matrix $[\mathbf{t}_{ij}]_\times$, the coplanarity constraint can be written as

$$ {}^n\mathbf{p}_j^{l,T} \mathbf{R}_j^T \left[\mathbf{t}_{ij}\right]_\times \mathbf{R}_i\, {}^n\mathbf{p}_i^l = 0. \tag{2.6}$$

By means of the dependent-images parameterization, the coplanarity constraint (2.6) is reduced to

$$ {}^n\mathbf{p}_j^{l,T} \mathbf{R}_{ij}^T \left[\mathbf{t}_{ij}\right]_\times {}^n\mathbf{p}_i^l = 0, \tag{2.7}$$

with $[\mathbf{t}_{ij}]_\times$ being composed of the base vector $\mathbf{t}_{ij}$ that points to $\mathbf{t}_j$ and is constrained to be of length one due to the scale ambiguity. The central term in equation (2.7) is commonly referred to as the *essential matrix* $\mathbf{E}$, i.e. $\mathbf{E} = \mathbf{R}_{ij}^T [\mathbf{t}_{ij}]_\times$, a rank 2 matrix with zero determinant. This matrix can be estimated by at minimum of five homologous points with the 5-point algorithm [Nistér, 2004; Stewenius et al., 2006].

Given a valid essential matrix, the parameters of the relative orientation are typically derived via factorization (Hartley & Zisserman [2003] pp. 239). The orientation is retrieved only up to a four-fold ambiguity. A geometrically valid solution is found by spatial intersection of a homologous point that is required to lie in front of both cameras, thereby fulfilling the *cheirality constraint*.

2.1.3. Rotations - various representations

In the previous section, relative and global rotations have been introduced as matrices in $\mathbb{R}^{3\times3}$. These matrices are not arbitrary but obey certain characteristics, called the *orthonormality constraints*

1. The column vectors of a rotation matrix are pairwise orthogonal.
2. The column vectors are of length one, which implies that the determinant of a rotation matrix is equal to one.

Every matrix satisfying these constraints is a rotation matrix. Each of these two constraints pose three conditions, thus a 3×3 rotation matrix with its nine elements is parameterized by three independent variables. In the following, some important rotation representations are outlined, which are used in the proposed method.

Rotations - Euler angles

One representation is given by the three *Euler angles* $\{\omega, \varphi, \kappa\}$. Each of these angles comprises the rotation around one of the three coordinate axes. This representation is not redundant but a rotation composed by Euler angles is not unique, because the overall rotation depends on the order of the three individual rotations. A common order in photogrammetry is $\omega_X \to \phi_Y \to \kappa_Z$ and is also used in this work:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & -\sin\omega \\ 0 & \sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{bmatrix} \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.8}$$

Even if the order is fixed, the mapping from $\mathbf{R}$ to $\{\omega, \varphi, \kappa\}$ is ambiguous due to the periodicity of the trigonometric functions. Moreover, it has to be defined whether the axes stay tight or are rotated after every individual rotation. Certain configurations lead to singularities (gimbal lock) so that two angles cannot be distinguished anymore.

Rotations - axis-angle representation

Every rotation can be described as a rotation by some angle $\alpha \in [0, \pi]$ around an arbitrary axis $\bar{\mathbf{r}}, \|\bar{\mathbf{r}}\| = 1$. This representation is called the *axis-angle representation* and is given by $\mathbf{r} = \alpha\bar{\mathbf{r}}$.

A closed 3-ball $B_3 \subset \mathbb{R}^3$ with radius π is a common representation of the set of all possible rotations. Every point inside B_3 represents a rotation by an angle equivalent to the distance to the origin. Points on the boundary of this ball (i.e. rotations by $\alpha = \pi$) are not unique because diametrically opposite points represent equivalent rotations, i.e. $\mathbf{r} \mathrel{\widehat{=}} -\mathbf{r} \Leftrightarrow \|\mathbf{r}\| = \pi$. Thus, the mapping from $\mathbf{r}$ to $\mathbf{R}$ is *surjective* (i.e. every rotation $\mathbf{R}$ is mapped by at least one $\mathbf{r}$) but not *injective* (one-to-one) for rotations by π. The mapping is given by the *exponential map* and can be computed using *Rodrigues' formula* (see proof in Appendix C.1 or e.g. Faugeras [1993], p. 268):

$$\mathbf{R} = \exp\left([\mathbf{r}]_\times\right) = \mathbf{I}_{3\times3} + \sin(\alpha)\,[\bar{\mathbf{r}}]_\times + (1-\cos(\alpha))\,[\bar{\mathbf{r}}]_\times^2\,. \tag{2.9}$$

The inverse mapping is given by the *logarithm map* (see proof in Appendix C.2):

$$[\mathbf{r}]_\times = \log(\mathbf{R}) = \begin{cases} \mathbf{0}_{3\times3} & \text{if} \quad \mathbf{R} = \mathbf{I}_{3\times3} \\ \arcsin\|\mathbf{W}\|_2 \frac{\mathbf{W}}{\|\mathbf{W}\|_2} & \text{else,} \end{cases} \tag{2.10}$$
$$\text{with} \;\; \mathbf{W} = \frac{\mathbf{R}-\mathbf{R}^T}{2}.$$

It is important to note that, because the exponential map is not injective, the logarithm map is not unique for rotations by π (see an alternative computation in Appendix C.3). This topic will be further studied in Section 2.2.2.

Rotations - Quaternions

Quaternions can be seen as an extension of complex numbers and consist of real 4-vectors in $\mathbb{R}^4$, forming the so called *quaternion sphere* Q. They are constructed by a scalar s (the real part) and a vector $\mathbf{v}$ (the imaginary part), $\mathbf{q} = (s, \mathbf{v})$, and have certain rules for computation (see e.g. [Förstner & Wrobel, 2004], pp. 47). Rotations are represented by unit quaternions, hence quaternions that fulfill $\|\mathbf{q}\| = 1$. There is a simple connection to the axis-angle representation, which is $\mathbf{q} = (\cos(\alpha/2), \sin(\alpha/2)\,\bar{\mathbf{r}})$. Quaternions are sign-ambiguous which means that $\mathbf{q}$ and $-\mathbf{q}$ represent the same rotation. A geometric interpretation of unit quaternions is the quaternion sphere (the unit sphere S^3 in $\mathbb{R}^4$). If the rotation angle is restricted to $[0, \pi]$, as for the axis-angle representation, the set of unit quaternions is limited to one *hemisphere* of S^3 with quaternions on the *equator* corresponding to rotations by angle π. A mapping to $\mathbf{R}$ is given by:

$$\mathbf{R} = \left(s^2 - \mathbf{v}^T\mathbf{v}\right)\mathbf{I}_{3\times3} + 2\mathbf{v}\mathbf{v}^T + 2s\,[\mathbf{v}]_\times\,. \tag{2.11}$$

Note that this mapping is two-to-one for $\alpha = \pi$. This is easy to prove because if $\alpha = \pi$ then $s = 0$ and (2.11) is reduced to an equation including only terms quadratic in $\mathbf{v}$ that are insensitive to different signs. The computation of a quaternion given a rotation matrix $\mathbf{R}$ is studied in Appendix D.

2.1.4. Rotations - norms and metrics

When working with rotations, it is necessary to define a norm that allows to describe the similarity (or dissimilarity) of two rotations. Each of the representations introduced in the previous section induces a different norm. A well-known norm using matrices is the *Frobenius norm*, which is equal to the Euclidean norm for $\mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$, considering all matrix elements. This leads to the so called *chordal distance*. Having two rotation matrices, $\mathbf{R}_i$ and $\mathbf{R}_j$, the chordal distance is defined as:

$$d_{chordal}\left(\mathbf{R}_i, \mathbf{R}_j\right) = \|\mathbf{R}_i - \mathbf{R}_j\|_F, \tag{2.12}$$

with $\|\cdot\|_F$ denoting the Frobenius norm. Because of the sign ambiguity of quaternions, the *quaternion distance* cannot just be deduced from the chordal distance as a distance in $\mathbb{R}^4$, but an easy extension covers ambiguous signs [Hartley et al., 2013]:

$$d_{quaternion}\left(\mathbf{R}_i, \mathbf{R}_j\right) = \min\left(\|\mathbf{q}_i - \mathbf{q}_j\|_2, \|\mathbf{q}_i + \mathbf{q}_j\|_2\right). \tag{2.13}$$

The most intuitive norm for rotations is the *angular norm* which describes the angle of the relative rotation $\mathbf{R}_{ij} = \mathbf{R}_i^T \mathbf{R}_j$.

$$d_\alpha\left(\mathbf{R}_{ij}\right) = \|\log\left(\mathbf{R}_{ij}\right)\|_2 = \|\mathbf{r}_{ij}\|_2 = \alpha. \tag{2.14}$$

Hartley et al. [2013] show that these different metrics are related to each other:

$$d_{chordal} = 2\sqrt{2}\sin\left(\alpha/2\right), \quad d_{quaternion} = 2\sin\left(\alpha/4\right). \tag{2.15}$$

2.2. Groups and manifolds

Image orientation parameters consist of 3D rotations and translations. As a consequence, one can say that the set of all possible rotation matrices forms a set, which is a subset of all 3×3 matrices. In the following, a brief overview of the basic concepts of *groups*, *Lie groups* and *Lie algebras* will be provided. A comprehensive study of these topics can be found in Gilmore [2008]; Sattinger & Weaver [2013].

Let us assume a set $\mathcal{G} = \{A, B, C\}$. The set $\mathcal{G}$ and a combinatorial operation $\circ$ form an *algebraic group* G, represented by the tupel $(\mathcal{G}, \circ)$ if, and only if, the following four axioms are valid:

(i) Closure under operation: every combination of two elements of the group with the associated operation results in an element of the group: $A \circ B \in G$.

(ii) Associativity: $A \circ (B \circ C) = (A \circ B) \circ C$.

(iii) Identity element: there exists an element I in G that does not change another element of the group when both are combined: $I \circ A = A \circ I = A$.

(iv) Inverse element: for each element A in G there exists an element A^{-1} in G so that a combination results in the identity element: $A^{-1} \circ A = A \circ A^{-1} = I$.

These axioms can be seen as algebraic properties of a group.

2.2.1. The concept of Lie groups

The concept of *Lie groups*, named after Norwegian mathematician Sophus Lie, links these algebraic properties with geometry. Each element in the group is identified with a point in a topological space - a *manifold*. A manifold is a space that appears Euclidean[3] in a local neighborhood but is considerably different on a global scale. For example, the neighborhood of a point on the unit sphere $S^2 \subset \mathbb{R}^3$ looks like a part of the plane $\mathbb{R}^2$, which means S^2 and $\mathbb{R}^2$ are topologically equivalent in that neighborhood. Globally, though, the sphere certainly is not Euclidean but spherical.

A group is a Lie group if, and only if, the following two axioms are valid:

(v) Smoothness of the combination: the combination of two group elements $A \circ B = C$ is differentiable.

(vi) Smoothness of the group inversion: the inversion of a group element A^{-1} is differentiable.

In the following the connection between Lie groups and image orientation parameters, rotation matrices in particular, is presented briefly.

SO(3) - the rotation group

How do the definitions given above apply to 3D rotation matrices? With matrix multiplication as combinatorial operation, axioms (i) and (ii) of an algebraic group are fulfilled. The identity element is the 3D identity matrix $\mathbf{I}_{3\times3}$ and the inverse is equal to the transposed element $\mathbf{R}^{-1} = \mathbf{R}^T$ (axioms (iii) and (iv)). Thus, the 3D rotation matrices form an algebraic group. This group of rotation matrices is called the *special orthogonal group* $SO(n)$ with n denoting the dimension. Hence, in the course of this thesis the focus will be on the group $SO(3)$: $SO(3) = \{\mathbf{R} \in \mathbb{R}^{3\times3} : \mathbf{R}^T\mathbf{R} = \mathbf{I}_{3\times3}, \det \mathbf{R} = 1\}$. This group is a subgroup of the *orthogonal group* $O(3)$, which includes all orthogonal 3×3 matrices, i.e. those matrices that fulfill only the first of the two criteria defined in Section 2.1.3. In contrast to $O(3)$, $SO(3)$ only includes matrices with a determinant equal to 1, which excludes reflections.

Besides the algebraic axioms also the topological axioms (v) and (vi) are fulfilled for rotation matrices. More precisely, matrix multiplication (axiom (v)) and transposition (axiom (vi)) are differentiable functions. This means that $SO(3)$ is a Lie group and the set of 3D rotation matrices is a manifold.

Another group, important for the topic of this thesis is the special Euclidean group $SE(3)$, which also includes translations: $EO(3) = \{\left(\begin{smallmatrix}\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1\end{smallmatrix}\right) \in \mathbb{R}^{4\times4} : \mathbf{R} \in SO(3), \mathbf{t} \in \mathbb{R}^3\}$. In the method that is presented in this work, this group is not used.

[3] A Euclidean space is a space described by real Cartesian coordinates.

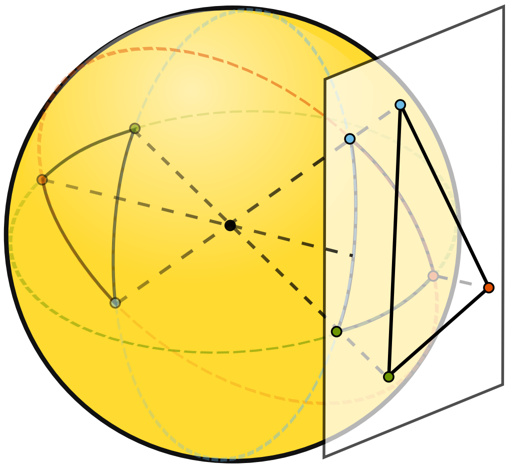


Figure 2.3.: Gnomonic projection of a spheric triangle on the S^2 manifold in $\mathbb{R}^3$ onto a tangent plane. The colored dashed lines show the geodesics. Points on the opposite side of the sphere project to the same point on the plane.

	$SO(3)$	Q	$\mathfrak{so}(3)$
$SO(3)$	-	$1-2$	$1-1$[a]
Q	$2-1$	$-$	$2-1$
$\mathfrak{so}(3)$	$1-1$[b]	$1-2$	$-$

Table 2.1.: Mappings between different rotation representations using the *row* as *from* and the *column* as *to*, e.g. it is a one-to-two mapping from $SO(3)$ to Q.

[a]not determined for $\alpha = \pi$, see Appendix C.3

[b]$2-1$ for $\alpha = \pi$

2.2.2. Lie algebra

A *Lie algebra* $\mathfrak{g}$ is the structure that results from the linearization of a Lie group G at the identity element. It is a linear vector space considering addition and scalar multiplication, thus every linear combination of two elements $\{A, B\} \in \mathfrak{g}$, e.g. $C = a \cdot A + b \cdot B$, is also an element of $\mathfrak{g}$. Moreover, the algebra is closed under commutation, which means that the commutator $C = [A, B] = (AB - BA)$ also lies in $\mathfrak{g}$ (in terms of a Lie algebra, the commutator is in general called the *Lie bracket*). The linear vector space can be seen as a *tangent space* to the Lie group at the identity element which allows various calculations, for instance the computation of an average, which is an important feature for the estimation of rotations (cf. next subsection). The mappings between the Lie group and the Lie algebra, thus the linearization and its inverse, are given by the logarithm map and the exponential map (cf. Section 2.1.3).

$\mathfrak{so}(3)$ - the rotation algebra

The rotation algebra $\mathfrak{so}(3)$ is the linearization of $SO(3)$, which is derived in Appendix C including a proof for the exponential and logarithm map. Their role as transition between group and algebra, in other words between manifold and tangent space, enhances the importance of the two mappings. We have $\log(\cdot) : SO(3) \to \mathfrak{so}(3)$ and $\exp(\cdot) : \mathfrak{so}(3) \to SO(3)$. As a consequence, the axis-angle representation of rotations is the Lie algebra $\mathfrak{so}(3)$ of $SO(3)$, strictly speaking, its skew-symmetric form:

$$[\mathbf{r}]_\times = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix} \in \mathfrak{so}(3). \tag{2.16}$$

Although in practice the limitation $0 \leq \|\mathbf{r}\| \leq \pi$ is valid to limit ambiguity, $\mathfrak{so}(3)$ consists of the set of all 3×3 skew-symmetric matrices, irrespective of its norm (note that the matrix norm of a skew-symmetric matrix is related to the norm of its vectorial components). This underlines the non-injectivity of the exponential map. Due to these limitations of the logarithm and exponential map, quaternions are often used instead of rotation matrices in practical applications (see e.g. Hartley et al. [2011]), where the map is substantially simpler.

The mapping from the quaternion sphere Q to $\mathfrak{so}(3)$ is two-to-one and can be visually interpreted as a gnomonic projection as pointed out in Hartley et al. [2013] and as it is exemplarily shown in Figure 2.3 for a lower dimension. Note that geodesics on the sphere map to straight lines on the plane and that points on the geodesic that lies in the plane parallel to the tangent plane map to infinity. Thus, the mapping plane of $\mathfrak{so}(3)$ is the projective space $\mathbb{P}^3 = \mathbb{R}^3 \cup \Pi_\infty$. Table 2.1 shows the ambiguities in the different mappings. A mapping $SO(3) \to \mathfrak{so}(3)$ via the quaternion sphere is in total one-to-one, although the two individual mappings $SO(3) \to Q$ and $Q \to \mathfrak{so}(3)$ are one-to-two and two-to-one, respectively. In the remainder of this work, the exponential and logarithm map are synonymous for forwards and backwards projection between $SO(3)$ and $\mathfrak{so}(3)$, irrespective of their implementation.

2.3. Convex optimization

This section is dedicated to a brief introduction into the topic of convex optimization and the way optimization problems are formulated in this thesis. A focus lies on convex optimization problems, which can be solved globally. The main requirements, *convex functions and sets*, will be reviewed in 2.3.1, followed by a presentation of the most important *convex optimization problems* in 2.3.2. A general *solving strategy* is outlined in 2.3.3. For a more comprehensive review about convex optimization, the book of Boyd & Vandenberghe [2004] is recommended, the topic of convex relaxation in the context of photogrammetric and computer vision applications is treated in Cremers et al. [2011].

The first question arising when someone is dealing with mathematical optimization is "*why would anyone optimize?*". A problem, for which no unique solution exists, requires optimization. For the given problem of image orientation in principle there exists an infinitely large number of possible solutions. Optimization aims at finding the best solution with respect to a given cost function based on the data that has been observed. More generally, in many applications in computer vision, statistics, or machine learning the goal is to find a model that best fits the given data.

Let there be the following optimization problem:

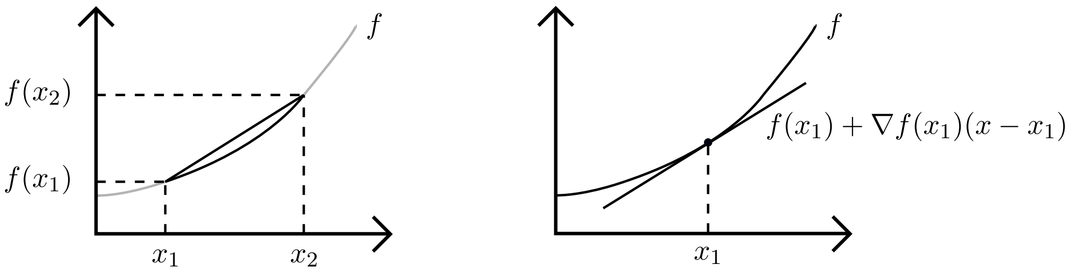
$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) \\ &\text{subject to} && f_i(\mathbf{x}) \leq 0, \;\; i = 1, \dots, m \\ &&& g_j(\mathbf{x}) = 0, \;\; j = 1, \dots, p \;\; . \end{aligned} \tag{2.17}$$

The vector $\mathbf{x} \in \mathbb{R}^n$ contains the unknown parameters of the model that are required to fulfill certain conditions imposed by *inequality constraint functions* $f_1, \dots f_m$ and *equality constraint functions* $g_1, \dots g_p$. Often, these constraints are based on prior knowledge about the parameters. They define the set of feasible parameters $\mathcal{S}$ for the model to be solved, narrowing the search space. Taking the image orientation problem, for example: for a set of images having homologous points, a solution in which the images are captured in a way, which renders any overlap regarding the pictured scene impossible, is not desired.

Among all $\mathbf{x}$ that satisfy these constraints, the best solution is an $\mathbf{x}^\star$ that minimizes the *objective* or *cost function* $f_0(\mathbf{x})$. Usually, the objective function represents the error of the model with respect to the observed data. Thus, in the remainder of this thesis, an optimal solution shows a minimum of the objective function. Hence, the optimal value $\mathbf{y}^\star = f_0(\mathbf{x}^\star)$ of the optimization problem (2.17) is defined as

$$\mathbf{y}^\star = \inf\{f_0(\mathbf{x}) | f_i(\mathbf{x}) \leq 0, i = 1, \dots, m, g_i(\mathbf{x}) = 0, \;\; i = 1, \dots, p\} \;\; .$$

In terms of mathematical optimization, most optimization problems are intractable which means that they cannot be solved. Actually, this merely implies that there is no guarantee to find the *global optimal solution*, i.e. the lowest value of the given objective function $f_0(\mathbf{x})$ satisfying all corresponding constraints. These optimization problems are referred to as *nonconvex optimization problems*. Of course, for many applications, this lack of guarantee is not a substantial problem, because with accurate initialization or prior knowledge, a feasible solution can be found, which is often sufficient for the given purpose. In general, the solution found is a *local optimum*, i.e. an optimal solution within a certain neighborhood in the solution space that is not necessarily the global optimum. However, even if the found solution is globally optimal, it generally lacks proof of this fact. In most cases, the only way for such a nonconvex problem to find a global optimal solution or to prove that a given solution is globally optimal is to perform a parameter space search. Instead of examining the whole parameter space exhaustively, more efficient methods evolved like the tree-based *branch and bound method* [Land & Doig, 1960; Clausen, 1999]. Starting from the root of the tree that represents the whole parameter set, individual branches of the tree are evaluated to compute bounds on the optimal values. Only branches that show an improvement to the current bound are investigated further. With every branch, the search space is narrowed until the necessary level of precision is reached. It is easily imaginable that this search becomes intractable with growing dimension of the parameter space n. Thus, for many complex optimization problems it remains impossible to guarantee that a global optimal solution has been found.



(a) convex function (b) first-order convexity

Figure 2.4.: Visualization of two convexity constraints for a function. Definition of a convex function (2.4a) and the first-order constraint (2.4b).

An exception to these problems are *convex optimization problems*. Convex optimization problems only have one minimum, every local minimum is also the global minimum. Simply put, every gradient descent method will converge to the global optimum. Regarding the general formulation of an optimization problem in (2.17), there are certain requirements for a problem to be convex: First, equality constraints are linear in $\mathbf{x}$ and second, the objective function and the inequality constraints are convex. In the following sections, these requirements are presented in more detail.

2.3.1. Convex functions and convex sets

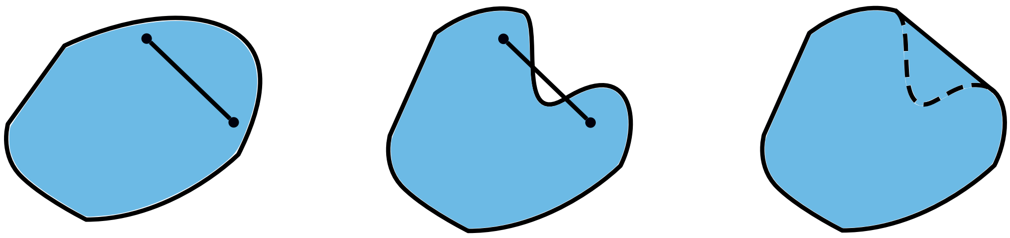
A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if, and only if, the domain of f, $\mathbf{dom}\, f$, is convex and for any $\{\mathbf{x}_1, \mathbf{x}_2\} \in \mathbf{dom}\, f$ it is:

$$f\left(\theta \mathbf{x}_1 + (1-\theta)\, \mathbf{x}_2\right) \;\leq\; \theta f\left(\mathbf{x}_1\right) + (1-\theta)\, f\left(\mathbf{x}_2\right), \quad \theta \in [0,1]\,. \tag{2.18}$$

A graphic interpretation of inequality (2.18) is shown in Figure 2.4a. The left side of the inequality (2.18) describes function f in the interval $[\mathbf{x}_1, \mathbf{x}_2]$, the right side describes a linear connection between $\mathbf{x}_1$ and $\mathbf{x}_2$. Due to positive curvature, the function itself is always below or equal to a linear connection of two points on that function. If f is differentiable, another condition for convexity can be formulated:

$$f(\mathbf{x_1}) + \nabla f(\mathbf{x_1})(\mathbf{x} - \mathbf{x_1}) \leq f. \tag{2.19}$$

This first-order condition is visualized in Figure 2.4b. It states that a first-order Taylor approximation of f at any point $\mathbf{x_1} \in \mathbf{dom}\, f$ is always below or equal to f. Inequality (2.19) reveals an important property of convex functions: it is possible to conclude global from local information. The Taylor approximation at any point is a global underestimator of the function, i.e. it gives a lower bound for the optimal value. For example, if at position $\mathbf{x_1}$ the first derivative is zero, i.e. $\nabla f(\mathbf{x_1}) = \mathbf{0}$, inequality (2.19) reduces to $f(\mathbf{x_1}) \leq f$. The function f lies above or is equal to $f(\mathbf{x_1})$, which proves that in a convex function every local minimum is the global minimum. Thus, $\mathbf{x_1}$ is the global optimal value. Note that a function f is concave (a function with negative curvature only) if the negative function $-f$ is convex. Moreover, it is noteworthy that according to (2.18) and



(a) convex set (b) nonconvex set (c) convex hull

Figure 2.5.: Visualization of convexity constraint for a set (2.5a), a counterexample (2.5b) and a convexification of a nonconvex set via the convex hull (2.5c).

(2.19) a linear function which has zero curvature is both, convex and concave. This implies that an optimization problem with a linear objective function is a convex optimization problem.

A set $\mathcal{D}$ is convex if, and only if, for any $\{\mathbf{x}_1, \mathbf{x}_2\} \in \mathcal{D}$

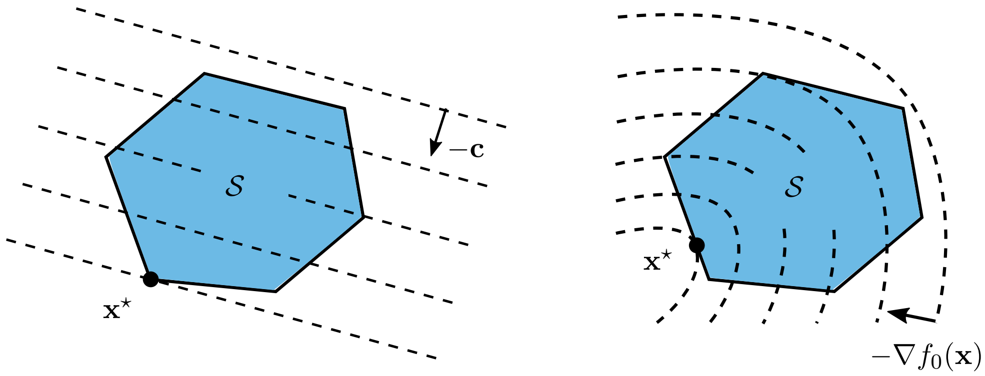
$$\theta \mathbf{x}_1 + (1-\theta)\,\mathbf{x}_2 \in \mathcal{D}, \quad \theta \in [0,1]\,. \tag{2.20}$$

This property is visualized in Figure 2.5a and 2.5b: A connection between any two points in a set must lie entirely inside the set. Besides convex functions and convex sets, the concept of a *convex hull* is relevant for the understanding of this thesis. If a set $\mathcal{D}$ is nonconvex, the convex hull $\mathcal{C} = \mathbf{conv}(\mathcal{D})$ of this set is the smallest convex set that contains $\mathcal{D}$, i.e. $\mathcal{D} \subseteq \mathcal{C}$. It can be defined as the sum of all convex combinations of points d_i in $\mathcal{D}$:

$$\mathbf{conv}(\mathcal{D}) = \left\{ \sum_{i=1}^{k} \beta_i \cdot d_i : d_i \in \mathcal{D}, k \in \mathbb{N}, \sum_{i=1}^{k} \beta_i = 1, \beta_i \geq 0 \right\} \tag{2.21}$$

Dealing with convex optimization, one has to verify whether a function or set is convex (or concave). Whereas this is easy to recognize for some functions or sets like a linear function, for instance, for other functions one might evaluate the criteria defined in (2.18) and (2.20) but often this leads to a large effort. A more elegant way to verify convexity is via so called *convex calculus rules*. These rules describe operations which, applied to convex sets or functions, preserve the convexity properties. For example, the sum or the pointwise maximum of n convex functions or the intersection of n convex sets are convex. Generally, the composition of convex functions leads to a convex function, e.g. $g(f(\mathbf{x}))$ is convex, if g and f are convex. Hence, the goal of convex calculus is to describe a set or function as a selection of sets or functions known to be convex combined with convexity preserving operations.

If an optimization problem is nonconvex, because the objective function or constraints are nonconvex, the only way to derive a convex optimization problem is via *convex relaxation*. Relaxation implies that nonconvex constraints are dropped or simplified leading to convex sets and functions. This modification of the original optimization problem often also affects the optimal solution, which



(a) linear program (b) quadratic program

Figure 2.6.: Geometric interpretation of an LP (2.6a) and a QP (2.6b) in $\mathbb{R}^2$. The feasible set $\mathcal{S}$ is depicted as shaded polyhedron, defined by $m = 6$ inequality constraints $\mathbf{G}\mathbf{x} \leq \mathbf{h}$, and the objective function is visualized using dashed iso-lines. The gradients of the objective functions are $\mathbf{c}$ and $\nabla f(\mathbf{x})$ of the LP and the QP, respectively. The negative gradients point in the direction of maximum descent.

sometimes entails a subsequent adaptation of the results. For instance, in a boolean linear optimization problem, the constraint $x_i \in \{0, 1\}$, which requires the components of $\mathbf{x}$ to be either zero or one, is nonconvex, because 0 and 1 cannot be connected without leaving the set $\{0, 1\}$. A typical relaxation of this optimization problem is to replace the boolean constraint with $x_i \in [0, 1]$, which allows every component to be an element of the interval $[0, 1]$. Often, one is actually interested in boolean values. Hence, after solving the convex relaxed optimization problem the real-valued solution has to be adapted, e.g. using a threshold of 0.5, to provide the desired outcome.

2.3.2. Convex optimization problems

In this section, some fundamental convex optimization problems that can be derived from the standard form of optimization problems given in (2.17) are presented. In the terminology of mathematical optimization, it is common to use the term *program* as a synonym for optimization problem. In a *linear program* (LP), the objective function and the constraint functions are affine (or linear). An LP reads as

$$\begin{aligned} &\text{minimize} && \mathbf{c}^T\mathbf{x} + d \\ &\text{subject to} && \mathbf{G}\mathbf{x} \leq \mathbf{h} \\ &&& \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{2.22}$$

with the constant d, the known vector $\mathbf{c}$, the parameter vector $\mathbf{x}$, $\{\mathbf{c}, \mathbf{x}\} \in \mathbb{R}^n$, $\mathbf{G} \in \mathbb{R}^{m \times n}$ and $\mathbf{A} \in \mathbb{R}^{p \times n}$. Geometrically, an LP can be interpreted as solving a linear objective function constrained by a convex polyhedron as depicted in Figure 2.6a. The optimal solution of an LP always lies at the edge of the polyhedron, i.e. the feasible set. If there is a unique solution, it is at a vertex of the polyhedron.

Convex *Quadratic programs* (QP) are optimization problems with a convex quadratic objective function and affine constraints,

$$\begin{aligned} &\text{minimize} && (1/2)\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{c}^T\mathbf{x} + d \\ &\text{subject to} && \mathbf{G}\mathbf{x} \leq \mathbf{h} \\ & && \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{2.23}$$

where $\mathbf{P}$ is positive semidefinite, i.e. $\mathbf{x}^T\mathbf{P}\mathbf{x} \geq 0$. QP have the same constraints as LP and can be seen as a generalization of LP; with $\mathbf{P} = \mathbf{0}$ every QP can be transformed to an LP. In contrast to LP, the optimal solution can either be situated at the edge of the polyhedron or at the minimum value of the objective function if the minimum value is feasible (see Figure 2.6b).

Another important type of optimization problems is the *semidefinite program* (SDP). A SDP has the form

$$\begin{aligned} &\text{minimize} && \mathbf{c}^T\mathbf{x} + d \\ &\text{subject to} && \mathbf{F}_0 + \sum_{i=1}^{m} x_i\mathbf{F}_i \preceq 0 \\ & && \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{2.24}$$

with symmetric matrices $\mathbf{F}_0, \dots \mathbf{F}_m$. The inequality in (2.24) is a *linear matrix inequality* (LMI) (in contrast to a componentwise inequality) and is denoted by $\preceq$. This LMI means that the left-hand side is negative semidefinite, i.e. all eigenvalues of the matrix are nonpositive. Note that a formulation that requires a positive semidefinite matrix is possible, as well. A SDP is also a generalization of an LP. Both programs minimize a linear objective function. If the matrices $\mathbf{F}_0, \dots \mathbf{F}_m$ are diagonal, the LMI is equivalent to the componentwise inequality constraints of an LP. Likewise, the inequality constraints of an LP can be formulated as an LMI with $\mathbf{G}\mathbf{x} - \mathbf{h}$ as a diagonal matrix. For further information on semidefinite programming the reader is referred to Vandenberghe & Boyd [1996].

2.3.3. Solving convex optimization problems

Several approaches exist for solving a convex optimization problem. For special problems like a QP with a linear relation between observations and unknowns and without equality or inequality constraints, i.e. an unconstrained linear least squares problem, an analytic solution exists given by the *normal equation system*. The same counts for a set of linear homogeneous equations which can be solved using a *singular value decomposition* (SVD). In many cases, however, constrained problems have to be solved. State-of-the-art in solving these problems are so called *interior point methods*, *primal-dual interior point methods*, in particular. These methods can be applied to most types of optimization problems and are solvable in polynomial time [Wright, 2005].

In order to give an insight into these methods, the concept of the *Lagrange dual function* shall be reviewed at this place. The *Lagrange function* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ of an optimization problem such as (2.17), which now is referred to as the *primal* problem, incorporates inequality and equality constraints scaled with the Lagrangian multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$,

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{j=1}^{p} \nu_j g_j(\mathbf{x}). \tag{2.25}$$

One can imagine the individual Lagrangian multipliers λ_i and ν_j as *prices* that charge the Lagrange function if a constraint is violated, i.e. $f_i(\mathbf{x}) > 0, g_j(\mathbf{x}) \neq 0$. For a feasible solution $\tilde{\mathbf{x}} \in \mathcal{S}$ (a solution that fulfills every constraint), $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ provides a lower bound for the optimization problem, if $\boldsymbol{\lambda} \geq 0$, because the terms in the first and second sum are either nonpositive or zero. The Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is the pointwise minimum of L,

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}). \tag{2.26}$$

For a fixed $\mathbf{x}$, $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ comprises the infimum of a set of affine functions of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ and is therefore a concave function. The relation between the Lagrangian dual and the primal objective function reads as:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\mathbf{x}}). \tag{2.27}$$

Knowing this relationship, one can formulate the *dual* problem to the given primal problem in (2.17):

$$\begin{array}{ll} \text{maximize} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \geq 0. \end{array} \tag{2.28}$$

Because the Lagrange dual function is a global underestimator of the primal objective function, one has to maximize $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ in order to find a best lower bound to the primal problem. Let $p^\star$ be the optimal value of the primal and $d^\star$ of the dual problem then the difference $\delta_{dual} = p^\star - d^\star$ is called the *duality gap*.

Assuming a convex optimization problem, there are five conditions the primal and dual parameters $\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star$ have to fulfill in order to be considered as primal and dual optimal with zero duality gap. These conditions are called *Karush-Kuhn-Tucker* (KKT) conditions:

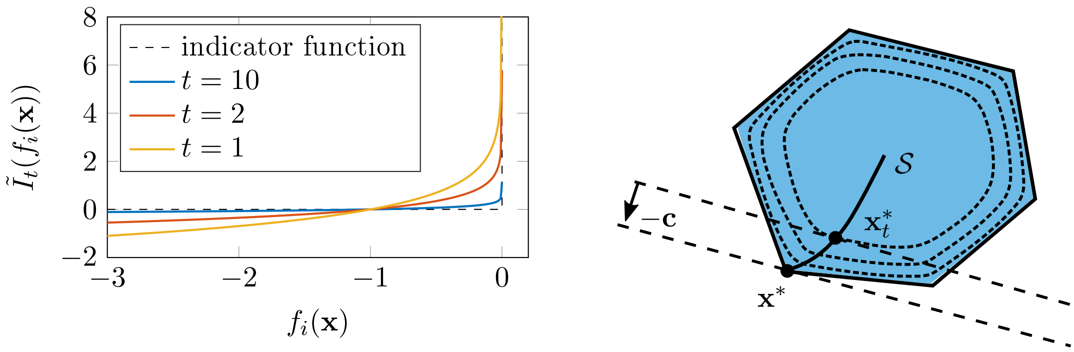
$$f_i(\mathbf{x}^\star) \leq 0, \qquad i = 1, \ldots, m \tag{2.29}$$

$$g_j(\mathbf{x}^\star) = 0, \qquad j = 1, \ldots, p \tag{2.30}$$

$$\lambda_i^\star \geq 0, \qquad i = 1, \ldots, m \tag{2.31}$$

$$\lambda_i^\star f_i(\mathbf{x}^\star) = 0, \qquad i = 1, \ldots, m \tag{2.32}$$

$$\nabla f_0(\mathbf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(\mathbf{x}^\star) + \sum_{j=1}^{p} \nu_j^\star \nabla g_j(\mathbf{x}^\star) = 0 \ . \tag{2.33}$$



(a) *log-barrier function* (b) *central path*

Figure 2.7.: The log-barrier function $\tilde{I}_t(f_i(\mathbf{x}))$ for different t compared to the nondifferentiable indicator function (2.7a). Visualization of the central path for the example in Figure 2.6a (2.7b). The dashed curves inside $\mathcal{S}$ are contour plots of the combined log-barrier function $\phi(\mathbf{x}, t)$ for different t. The central path results from the tangent point of the iso-lines of $f_0(\mathbf{x})$ (depicted as dashed lines with gradient $\mathbf{c}$) at the iso-lines of $\phi(\mathbf{x}, t)$.

Equations (2.29) and (2.30) are the constraints of the primal problem (2.17), Equation (2.31) is the constraint of the dual problem (2.28). The fourth condition (2.32) is known as *complementary slackness*. If $\lambda_i^\star > 0$ then the ith constraint is tight, i.e. $\lambda_i^\star > 0 \Longrightarrow f_i(\mathbf{x}^\star) = 0$. Equivalently, the constraint is lose, i.e. $f_i(\mathbf{x}^\star) < 0$, if $\lambda_i^\star = 0$. In other words, a lose constraint could be dropped without changing the solution of the optimization problem. The fifth condition (2.33) requires the gradient of the Lagrange function to be zero which implies a global optimum of a convex optimization problem.

Interior point methods rely on solving the KKT conditions for an optimization problem with equality constraints only. Given a problem whose objective function and inequality constraint functions are twice differentiable, the goal is to relax this problem and approximate a problem free of inequality constraints in order to apply an unconstrained optimization approach like Newton's method (cf. [Dennis Jr & Schnabel, 1996]). One way of such a relaxation is via a *log-barrier function*,

$$\tilde{I}_t(u) = -(1/t)\log(-u). \tag{2.34}$$

This function converges to ∞ for $u \to 0$ and approximates a binary indicator function $I(u)$, $I(u) = 0$ if $u \leq 0$ and $I(u) = \infty$ if $u > 0$. The indicator function can be interpreted as a cost for the violation of an inequality constraint as in (2.17), shown in Figure 2.7a by a dashed line. For higher t, the approximation becomes more exact as can be seen by the three curves in Figure 2.7a. $\tilde{I}_t(u)$ is convex and differentiable and an approximation to the optimization problem (2.17) is derived as

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) + \phi(\mathbf{x}, t), && \phi(\mathbf{x}, t) = \sum_{i=1}^{m} \tilde{I}_t(f_i(\mathbf{x})) \qquad (2.35) \\ &\text{subject to} && g_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p, \end{aligned}$$

with the problem specific log-barrier function $\phi(\mathbf{x})$.

In practice, the primal problem (2.35) and its dual are solved iteratively with increasing t. For every t, the optimal value $\mathbf{x}_t^\star$ lies inside the feasible set $\mathcal{S}$ and moves towards the global optimum $\mathbf{x}^\star$ for increasing t. The set of all solutions $\mathbf{x}_t^\star$ forms the *central path*, depicted in Figure 2.7b. It intersects the iso-lines of $f_0(\mathbf{x})$ at the tangent points of $f_0(\mathbf{x})$ with $\phi(\mathbf{x}, t)$. The duality gap results in $\delta_{dual,t} = m/t$, hence with $t \rightarrow \infty$ the duality gap vanishes. This allows a formulation of a stopping criterion based on a desired accuracy ϵ, thus the iterative procedure stops if $m/t < \epsilon$.

3. State-of-the-art

The topics related to this thesis can be assigned to two research communities: photogrammetry and computer vision. Image orientation is a problem investigated for a long time in photogrammetry and more recently also by the computer vision community. In recent years, driving forces behind the development of innovative algorithms for image orientation were the growing number of publicly available image data, the need for new vision based navigation applications, e.g. in the field of robotics, and the desire for highly automated large scale mapping. As part of these developments, many different image orientation methods were developed that exhibit various advantages and disadvantages considering the accuracy, efficiency and variability regarding different kinds of image data. Image orientation encounters a growing number of applications including terrestrial or airborne 3D reconstruction for architectural, cultural heritage, archaeological or topographical mapping purposes, quality control of industrial products of different scales but also real time positioning and navigation for moving platforms, so called on-line orientation or SLAM (simultaneous localization and mapping) [Davison et al., 2007; Konolige & Agrawal, 2008; Grisetti et al., 2010]. While SLAM in principle is very similar to the problem discussed in this thesis, the focus of the method described in this work lies on off-line orientation, i.e. not in real-time.

The early beginnings of the orientation of images date back to the second half of the 19th century [Grün & Huang, 2013]. It is well established that the optimal solution to the problem of (off-line) image orientation is found via robust bundle adjustment (see e.g. [Schwidefsky & Ackermann, 1976]). Triggs et al. [2000] give a more recent comprehensive overview over current state of bundle adjustment. It is a nonlinear, nonconvex and unconstrained least squares optimization which can be solved analytically, although the authors mention that the frequent use of robust cost functions like M-estimation functions or other adaptations lead to the fact that the optimization problem significantly differs from an ordinary least squares model. The nonlinearity and nonconvexity of the optimization require an initialization of the unknown orientation parameters and object coordinates. Approaches for the estimation of these initial values can be classified into three categories: *sequential*, *hierarchical* and *global* models.

This chapter is dedicated to give an overview over existing work in the related research that is relevant to this thesis. It is structured according to the workflow of the image orientation procedure and starts with a presentation of the state-of-the-art in the description of image features and computation of relative orientations in Section 3.1. This is followed by a study of the most important relating works on sequential and hierarchical (Section 3.2) as well as global image orientation (Section 3.3).

3.1. Image correspondences and relative orientations

In this section, the state-of-the-art in the computation of image correspondences and relative orientations is reviewed. Image correspondences are the foundation of the model presented in this work, although their computation will not be explained in detail. Nevertheless, the computation of image correspondences is an active field of research. By the end of the last century, various feature descriptors emerged, which allow the automatic detection and matching of point or line-features (e.g. [Harris & Stephens, 1988; Förstner, 1986]). One of the the most widely used descriptors is the scale invariant feature transform, better known as the *SIFT* descriptor [Lowe, 2004], which is invariant to different scales and rotations. Various extensions to SIFT were developed in the following years like the *SURF* descriptor (speeded up robust features) of Bay et al. [2008] that primarily allows a faster computation thanks to the use of integral images. More recently, many works about learning image descriptors were presented (e.g. [Brown et al., 2011; Simonyan et al., 2012; Trzcinski et al., 2015; Chen et al., 2016]). While descriptors like SIFT and SURF have a constant formulation with only few adaptive parameters, the latter methods aim at learning an optimal model for the description of specific feature points that are found by any feature point detector. In terms of accuracy and computation time (without taking the time for training into consideration) they often outperform descriptors like SIFT and SURF.

In the model presented in this work, as in most models for the estimation of image orientations, at a specific point pairwise or tripletwise relative orientations are computed. Centerpieces are linear estimations based on image correspondences like the *fundamental* and *essential matrix* for two views [Faugeras, 1992; Hartley, 1992; Luong & Faugeras, 1996] or the *trifocal tensor* for three views [Shashua, 1995; Hartley, 1997; Ressl, 2000]. Relative orientation parameters can be derived from these entities (e.g. [Hartley & Zisserman, 2003]). Although the estimation of these matrices or tensors is linear and therefore a straight forward process, it suffers from outliers in the image correspondences and the occurrence of degeneracy for specific configurations. It is general practice to eliminate outliers using random sample consensus (*RANSAC*) [Fischler & Bolles, 1981]. There are several attempts to extend this randomized approach and formulate more sophisticated algorithms for this task. Torr et al. [1998] propose a combination of RANSAC with a model selection approach based on a scoring function to distinguish between degenerate and non-degenerate cases. This model is further improved in Torr & Zisserman [2000] and Torr & [2002]. In Torr & Zisserman [2000], the cost function is based on the log likelihood of the solution, instead of maximizing the size of the inlier set. Inliers are thus scored based on their reprojection error, which has a geometrically more meaningful interpretation than only the number of inliers. Torr & [2002] extends this model by directly optimizing the posterior of the Bayesian formulation.

In case of the essential matrix, which requires knowledge of the interior orientation of the camera, the risk of a degenerate case is much smaller. A solution for the essential matrix is found by the five-point algorithm presented in Nistér [2004] and extended in Stewenius et al. [2006], which is often applied in combination with RANSAC. In the five-point algorithm an algebraic objective function is minimized, whereas a distinction between inliers and outliers is favorably based on the geometric reprojection error. Moreover, the heuristic nature of RANSAC lacks the guarantee of the solution to be optimal. Yang et al. [2014], for example, state that this may lead to inferior solutions. Several approaches propose a globally optimal estimation of the essential matrix, e.g. [Hartley & Kahl, 2007; Enqvist et al., 2011a; Yang et al., 2014]. Often, the global optimum is reached with a branch-and-bound search over different parameter spaces representing the essential matrix. Hartley & Kahl [2007] were inspired by the idea that once the relative rotation is known, the relative translation can be estimated in a quasiconvex optimization using the L_∞ norm of the reprojection error. They perform a branch-and-bound search over the space of rotations $SO(3)$ and evaluate the corresponding translations using second order cone programming (SOCP) [Boyd & Vandenberghe, 2004]. Yang et al. [2014] propose a model that finds a global optimal solution with respect to a geometric error using a branch-and-bound search over the essential manifold, which is defined explicitely as a 5D manifold $SO(3) \times (SO(3) \backslash SO(2))$. This manifold can be interpreted as the independent images representation of the relative orientation.

Iterative methods for the estimation of the relative orientation parameters in general lead to more accurate results than can be derived from the essential matrix. Horn [1990] describes an iterative algorithm using unit quaternions that is also applicable for the case in which no initial guess is available. He shows that given a large number of correspondences this method will reach the global optimum. Moreover a detailed discussion about critical surfaces is given. An iterative estimation in a Gauss-Helmert-model is described in Förstner et al. [2004].

Synthesis - image correspondences and relative orientations

In the method proposed in this thesis, an essential matrix is estimated by the five-point algorithm based on SIFT feature points. For this algorithm efficient implementations are available. Because the focus lies on precise relative orientation parameters, a subsequent iterative estimation in a Gauss-Helmert model [Förstner et al., 2004] is performed. It is shown in Section 4.2.1, that the basin of convergence of this iterative estimation is large and does not necessarily require a globally optimal essential matrix.

3.2. Sequential and hierarchical image orientation

In this section, related work for two important strategies for the computation of image orientation parameters is reviewed: sequential and hierarchical image orientation. Sequential image orientation starts with a subset of images and sequentially adds further images, typically using alternating spatial intersection and resection until all images are oriented. These methods usually apply intermediate bundle adjustments to reduce a drift of the solution. Hierarchical models process several small sets of images and compute the orientation in a respective local coordinate system. Afterwards, these local reconstructions are merged so that all images are oriented in a single coordinate system. In particular the sequential orientation model can be seen as the common practice in photogrammetry, outlined in the respective textbooks (e.g. Kraus [1997], pp. 48, Hartley & Zisserman [2003], pp. 435 or Pollefeys [2007], pp. 393). In the last twenty years, various models have been presented that are based on these two different procedures. In the following, the most important works are presented.

3.2.1. Sequential image orientation

Pollefeys et al. [2004] describe a sequential model for the orientation of a hand-held camera using self-calibration. They identify an initial pair of images based on the number of correspondences and an image-based distance that avoids images that are captured too close together and compute the two-view geometry in the uncalibrated case. Further images are added based on projective resection and intersection. The solution subsequently is computed in a Euclidean coordinate system using self-calibration in order to determine the focal length in a final bundle adjustment. The automatic orientation of uncalibrated image sequences in general does not allow highly accurate solutions [Remondino & El-Hakim, 2006]. An automatic model for image orientation using calibrated images is presented in Läbe & Förstner [2006]. Their model sequentially adds images to an initial pair based on their pairwise relative orientation, hence without explicit computation of 3D coordinates of homologous points. These orientations are checked for correctness both beforehand, based on a concatenation of rotations and the coplanarity of base vectors in every triplet combination of three respective pairwise orientations, and during integration using linear epipolar constraints.

More recently, related work concentrated on the orientation of very large sets of images, for example from image-hosting websites, taken with many different types of cameras, mostly for the purpose of 3D visualization [Snavely et al., 2008; Agarwal et al., 2009; Wu, 2013]. Snavely et al. [2008] construct a view-graph of images and pairwise relative orientations and tracks of observations over the respective edges. As in Pollefeys et al. [2004], the initial pair is found based on the number of correspondences but subject to the constraint that there is no single homography that describes these matches, in order to avoid singularities. Moreover, they only use calibrated images. Images are added sequentially via spatial intersection and direct linear transform (DLT) [Marzan & Karara, 1975], in an order based on the number of observations corresponding to already estimated 3D points. After every extension of the set of images and object points, all recently added unknown parameters are refined in a bundle adjustment. Agarwal et al. [2009] extend the image orientation model of Snavely et al. [2008] to deal with even larger sets of images. The sequential orientation is performed only on a skeleton graph, a subgraph considering only those images that significantly contribute to an extension of the reconstruction. Remaining images are then included subsequently. The method of these papers is implemented in the open source software *Bundler*[1]. Another open source software called *VisualSFM*[2] is outlined in Wu [2013]. The main concern of this work is to derive a $\mathcal{O}(n)$ formulation of the image orientation problem. In order to reduce a drift of the solution, after some

[1]This software can be downloaded at `http://www.cs.cornell.edu/~snavely/bundler/`.
[2]This software can be downloaded at `http://ccwu.me/vsfm/`.

sequences a re-estimation of the 3D coordinates of points is proposed, which were discarded initially because of inaccurate relative orientations. This is similar to a loop closure without an explicit loop detection.

In contrast to large and unordered image datasets, the model proposed in Pierrot-Deseilligny & Clery [2011] has a different goal. The authors present an orientation software, *Apero*[3], that aims at a precise photogrammetric reconstruction while pursuing a complete automation of the workflow. They allow various camera models such as fisheye lenses with more complex distortion models than the aforementioned approaches. The orientation procedure includes a combination of the essential matrix, spatial intersection and resection and intermediate bundle adjustment.

3.2.2. Hierarchical image orientation

Parallel to sequential models, hierarchical image orientation approaches emerged. Fitzgibbon & Zisserman [1998] proposed a hierarchical model based on image triplets. After an estimation of the trifocal tensor and the spatial intersection of image correspondences, overlapping triplets are merged via a best fit homography either minimizing a cost function regarding the distance of 3D points (which necessitates a Euclidean normalization) or the reprojection error in image space. Either one-image or two-images overlapping triplets can be applied. The one-image overlap requires correspondences tracked over at least five images but leads to superior efficiency. The two-images overlap does not necessarily require any intersected 3D points and has the advantage that erroneous matches can be detected. A generalization to this work is presented in Nistér [2000]. Before triplets are merged, they are organized in a binary tree, with the root covering the whole sequence and for every child the sequence is cut in half (i.e. each of the leaves spans over three subsequent images of the sequence). A threshold for the number of correspondences controls the maximum range of triplets, trying to find the best compromise between base length and number of homologous points. When the optimal set of triplets is found, the triplets are merged and intermediate images are added. This approach depends on prior knowledge about the image location, e.g. implicitly given by an ordered image sequence.

The approach of Havlena et al. [2009] is also applicable to unordered images. It avoids an exhaustive matching of image pairs and instead computes a similarity matrix using SURF features [Bay et al., 2008], which primarily affects efficiency. From this, the best pairwise match is selected sequentially and three possible image triplets are constructed from each pair that are then scored based on the quality of a local reconstruction. The quality is assessed using the intersection angles of the projection rays at the 3D points. The resulting triplets are merged and remaining images are oriented afterwards using 3D-2D correspondences.

Gherardi et al. [2010] present a hierarchical orientation model using pairs of images, which are organized in a binary cluster tree. Starting at the leaves, single branches are oriented in a fashion similar to sequential image orientation forming clusters of images that are merged when two branches

[3] This software can be downloaded at `http://logiciels.ign.fr/?Micmac`.

meet. With an additional balancing of the tree, the efficiency of the model is improved by one order of magnitude which is shown in a comparison to *Bundler*.

Ni & Dellaert [2012] formulate the image orientation problem as a bipartite visibility graph with object points and images being vertices, connected by an edge if an observation of the point exists in the respective image. This graph is converted to a hypergraph, in which a hyperedge, representing a 3D point, connects all those vertices, i.e. images, in which the respective point is observed. The hypergraph is partitioned where only a small number of hyperedges connecting subsets of images exists. When the hypergraph is partitioned, the subsets of images are oriented starting with the smallest subsets of images. Then these subsets are aligned using the points corresponding to the separating hyperedges.

Another hierarchical method for the orientation of large datasets is presented in Mayer [2014]. Starting from image triplets, the hierarchical merging is performed with a fixed overlap of two images, i.e. two triplets are merged to a quadruplet and so on. The focus of this work lies on efficiency improvement, evaluating different strategies for the reduction of points used for the hierarchical merging. According to his findings, only a random deletion of points leads to reliable and precise results.

Synthesis - sequential and hierarchical image orientation

Various models for the sequential and hierarchical estimation of image orientation parameters exist that often work well on various kinds of datasets. Considering the general strategy of these approaches, the solution suffers from the distribution and accumulation of errors. In sequential approaches, with every added image, the drift of the solution increases, which can only be counteracted by a regular intermediate bundle adjustment, which leads to serious problems regarding the efficiency for large image sets. Moreover, the result regarding both, sequential and hierarchical methods, depends on heuristic decisions like the selection of an initial pair of images (e.g. [Thormählen et al., 2004]) and the order of adding images or merging subsequences.

3.3. Global image orientation

In the previous section, sequential and hierarchical approaches for the computation of initial orientations were presented. In this section, the focus is on global image orientation models. Instead of incrementally enlarging the set of parameters, as in the sequential and hierarchical approaches, these models follow the precept of simplifying the estimation by a division into subproblems for different *types* of parameters. Thus, in general, rotations and translations are estimated separately based on pairwise relative orientations. The term *global* in this context means that the entire set of relative information is taken into account at once. This leads to the conceptual advantage that errors are not accumulated but distributed more equally over all orientation parameters. In particular, loops in the image sequence are implicitly taken care of, because all redundant relative orientations are used simultaneously.

Although the evolution of global image orientation models dates back to the work of Govindu [2001], it has not yet reached the state of a commonly used approach, for instance in available photogrammetric software packages. However, it recently gained increasing interest in the computer vision community. The concept of the view-graph introduced in Levi & Werman [2003] is often used in this context. In the view-graph, images or global orientations of images are represented as vertices and two vertices are connected by an edge if a relative orientation has been established between the two images. This concept is also applied in the robotics community (see e.g. Olson [2009]). Although often only two dimensional orientations are considered, the principle is similar. In Olson [2009] a graphical model is applied to find a cluster of consistent edges using a weighted adjacency matrix and a binary indicator vector. To derive an efficient solution, the problem is relaxed allowing a continuous valued binary vector and discretized later on.

In the following, the focus lies on three dimensional orientations. Important contributions in the field of global image orientation are presented, divided into the estimation of rotations and translations.

3.3.1. Estimation of rotations

In almost all global image orientation approaches, rotation parameters are estimated first. This is basically due to two major reasons. First, a relative rotation can be estimated precisely, irrespective of the base-to-distance ratio that affects the estimation of relative translations [Enqvist et al., 2011b]. Second, having a set of global rotations the estimation of global translations can be formulated as a convex optimization problem.

The idea of estimating global rotations from a redundant set of pairwise relative rotations can be traced back to the fundamental work of Govindu [2001]. Rotations, represented as quaternions, are estimated in an unconstrained linear least-squares optimization. The simplified problem formulation lacks a treatment of the ambiguous nature of quaternions, neither the sign ambiguity nor the norm is taken into account (cf. Section 2.1.3). A few years later, Govindu proposed an extension [Govindu, 2004], which uses the results of Govindu [2001] as initialization for an iterative refinement in the Lie algebra of $SE(3)$ (the special Euclidean group of dimension 3). This formulation allows the estimation of an intrinsic average that minimizes a cost function defined on the motion manifold. Primary results of this work show that, while the accuracy is approximately three times worse than after bundle adjustment, the computation is significantly more efficient. A further extension to the model proposed in Govindu [2004] is presented in Govindu [2006], which incorporates robustness to the estimation of orientations. Here, the problem is formulated as an over-determined view-graph estimation, for which a minimum solution exists with an arbitrary minimum spanning tree (MST). In the manner of a RANSAC algorithm, random MST are chosen and the solution with the maximum number of inliers, i.e. edges that are consistent with the solution of the respective MST, is selected.

Experiments are conducted only on one dataset, for which the approach was performed in a sliding window, so without considering the whole view-graph, which implies a limitation concerning the size of the view-graph. Almost half of the relative orientations were considered outliers.

Estimation using rotation matrices

Martinec & Pajdla [2007] present an unconstrained linear least-squares estimation using rotation matrices. The homogeneous system of linear equations is solved using eigenvalue decomposition and the orthonormality constraint is enforced afterwards with a mapping to the closest orthonormal matrix with respect to the Frobenius norm. The authors also describe a solution using quaternions in an unconstrained linear optimization similar to Govindu [2001] and reveal that it should not be applied in practice because the difference of the estimated values to the manifold of valid quaternions is too large. Due to the unconstrained optimization, these discrepancies also occur using rotation matrices but are generally smaller. Arie-Nachimson et al. [2012] present a similar solution by spectral decomposition which performs better than the method of Martinec & Pajdla [2007], in particular for larger outlier rates. Their model also allows the formulation of a SDP for which the problem is cast into a trace maximization for a product of symmetric matrices composed of relative and global rotation matrices. The optimal matrix is required to be positive semidefinite and to have $3{\times}3$ identity matrices on its main diagonal in order to derive global rotations via factorization. Additional linear constraints are introduced which are equivalent to nonlinear determinant constraints to make the relaxation tighter. Accuracy and efficiency are evaluated on various datasets and show comparable and partly better results than state-of-the-art sequential approaches of that time. Recently, Horowitz et al. [2014] and Saunderson et al. [2014, 2015] demonstrated how the convex estimation in a SDP can be further improved by requiring the global rotation to lie in the convex hull of $SO(3)$. They describe the convex hull as a linear matrix inequality constraint. This linear matrix inequality describes a spectrahedron, a natural generalization of a polyhedron and a convex shape expressed as the intersection of the convex cone of positive semidefinite matrices with an affine subspace. They show that this constraint leads to exact rotation matrices more often than the orthogonal constraints of Arie-Nachimson et al. [2012] does, because the feasible solution space is tighter.

Graph based detection of outliers in the relative rotations

Zach et al. [2010] use a Bayesian framework on the view-graph to detect outliers in the relative orientations. For all cycles in the view-graph, conditional probabilities for existence or absence of an outlier are derived based on their deviation to the identity. Using loopy belief propagation, inference is spread to the Bayesian network. The binary constraint for edges (outlier or inlier) is relaxed to a real-valued constraint to derive a convex optimization problem. In practice, only a subset of all cycles is considered due to tractability, which requires a careful selection based on a sequence of spanning trees. Because noise in the relative orientations is not explicitly modeled with respect to the cycle length, the maximum length of a cycle is limited to six edges. Enqvist et al. [2011b] show examples where this restriction is not sufficient, for instance in long loopy sequences where erroneous edges between far away images exist. They propose a similar approach using a most reliable MST, weighted by the number of correspondences. Sequentially, remaining edges are added and for each edge the occurring cycle is checked for consistency. Additional search heuristics cope with the case, in which the initial MST contains outliers. The estimation of global rotations is performed with the method described in Govindu [2001] using initial rotations of the sequential graph consistency check to resolve the sign-ambiguity problem of quaternions. They compare their model to Bundler and show several examples in which Bundler only achieves degenerate solutions. The problem of taking an outlier containing initial MST is covered in Arrigoni et al. [2014b]. They propose an extension to Enqvist et al. [2011b] using cycle bases in the vector space formed by cycles in the graph. Cycles are classified into inliers and outliers and outlier edges in the respective cycle are identified via summation. This has a positive effect on the false negative rate, especially for weakly connected image sets.

Estimation with robust cost functions

Instead of finding outliers *before* the actual estimation of global rotations, Crandall et al. [2011] use a truncated cost function, combining a pairwise constraint between two cameras or camera and object point and a cost function depending on prior information such as geotags. The view-graph, extended by vertices for object points and edges from images to points if an observation exists, is viewed as a Markov Random Field (MRF). Using discrete Belief Propagation, initial rotations are estimated and refined in a non-linear optimization. In order to ease the computation, rotations are assumed to be not tilted, which sometimes is a reasonable constraint considering practical applications.

Hartley et al. [2011] and Hartley et al. [2013] achieve robustness within the estimation using a L_1 cost function. They apply the *Weiszfeld* algorithm ([Weiszfeld, 1937] (in French) or [Weiszfeld & Plastria, 2009]) in the Lie algebra $\mathfrak{so}(3)$ to find the geodesic median of a set of rotations. While this approach finds its primary application in the estimation of *one* rotation from several estimates, the authors also describe a sequential approach on the view-graph, in which one rotation is estimated at a time. The information is spread in the manner of a distributed consensus. Besides, Hartley et al. [2013] give a comprehensive overview of different metrics and representations as well as the convexity of manifolds, in particular $SO(3)$. Chatterjee & Govindu [2013] note that with increasing size of the graph, this method scales poorly and occasionally requires many iterations until convergence. They present a robust two-step model in which, firstly, a L_1 solution in $\mathfrak{so}(3)$ is computed and, secondly, refined in an iterative reweighted least-squares rotation averaging, similar to the approach in Govindu [2004]. The weighting of individual relative rotations is based on the residuals, i.e. the fitting error to the estimated global rotations. The approach is compared to the method of Hartley et al. [2011] and next to better convergence it also shows comparable and partly higher accuracy on two different datasets.

Arrigoni et al. [2014a, 2015b] solve the problem of Martinec & Pajdla [2007] and Arie-Nachimson et al. [2012] as a matrix completion and decomposition (e.g. [Candès & Tao, 2010]). The cost function is extended by two sparse matrices covering an outlier term and a matrix for missing data, which are then treated separately.

Synthesis - global rotation estimation

There are various approaches to estimate global rotations from a set of relative rotations, all of which have certain advantages and disadvantages. Unconstrained estimation, in general, does not provide an intrinsic estimation, i.e. the solution has to be mapped to the rotation manifold afterwards. Graph based outlier detection often has problems with large view-graphs whereas an estimation with a robust cost function still suffers from the presence of outliers, although to a smaller extent. Finally, intrinsic methods like Govindu [2004] require initialization of the unknowns and deliver only a local solution. Chatterjee & Govindu [2013] can be seen as a current state-of-the-art method for a robust, accurate and efficient computation of global rotations, which is why this method is used in several recent publications on global image orientation [Wilson & Snavely, 2014; Ozyesil & Singer, 2015; Cui et al., 2015].

The method presented in this work relies on an elimination of outliers before the estimation based on a new and effective graph-based algorithm. Rotations are then computed in a combined estimation, first, solving a convex SDP as proposed by Saunderson et al. [2014] and second, in an iterative averaging in the Lie algebra, similar to Govindu [2004]. The SDP provides accurate initial values for the Lie algebraic averaging, which allows a fast convergence and a high probability for a globally optimal solution.

3.3.2. Estimation of translations

Using the estimated global rotations and the relative translation directions, global translations are estimated. Related work on this topic can be roughly divided into two categories: First, translations and object points are estimated jointly, often in a quasiconvex L_∞ optimization. The second category comprises approaches, in which translations and structure are treated separately.

Quasiconvex L_∞ estimation

The first work that studies the characteristic of the L_∞ cost function in the scope of geometric reconstruction is Hartley & Schaffalitzky [2004]. The authors show that the L_2 reprojection error of one point moving in front of a camera has only one global minimum but is not convex (i.e. quasiconvex, which includes negative curvature). Considering the problem of spatial intersection, the L_2 optimal solution involves a summation of individual quasiconvex functions (one for each image), which is not necessarily convex. The maximum operator preserves quasiconvexity, thus the

L_∞ cost function, which describes the maximum residual, leads to a quasiconvex problem. The optimization is extended to be applied to the estimation of 3D points and global translations of multiple cameras using the convex cheirality constraint, which guarantees that all points lie in front of all cameras. Kahl [2005] extends this work in several ways. Firstly, he describes additional problems for which the approach can be applied and, secondly, he introduces the formulation as a second order cone program (SOCP), a common class of convex optimization programs. Moreover, the work also comprises an evaluation on real datasets, which reveals that the L_∞ solution is well suited for the initialization of a nonlinear bundle adjustment. A significant problem of the L_∞ optimization is its sensitivity to outliers. Sim & Hartley [2006] tackle this problem and present a heuristic method that tries to achieve robustness by iteratively solving the L_∞ problem and, after each iteration, eliminating the observations with maximal residual. They prove that the set of observations with maximal residual, the so called support set, in quasiconvex optimization, in particular given a SOCP, always contains at least one outlier. When a satisfactory solution is found, the iteration is stopped. For large problems with many second order cone constraints, this approach leads to a considerable amount of computation time. In order to cope with outliers, Ke & Kanade [2007] minimize the mth smallest residual, which can be compared with a least-median method. However, this formulation lacks quasiconvexity and finding the global optimal solution requires to solve a set of individual programs which is costly for large scale problems.

The approach of Martinec & Pajdla [2007] aims to improve both, the sensitivity to outliers and the inefficiency. Initially, they discard most likely mismatches, which are assumed to lie apart from clusters in image space. A predefined amount of correspondences is discarded based on the distance to a Gaussian fitted to all points in the image. From the remaining points, four of those that are farthest from the Gaussian are selected to serve for translation and structure estimation using the method of Kahl [2005]. This considerable reduction of data has a significant positive effect on efficiency and it also reduces redundancy and, thus, reliability.

Translation estimation based on pairwise constraints

Next to rotation parameters, Govindu [2001] also describes a linear framework for the estimation of global translations from pairwise translation directions. In an unweighted design, this approach leads to unstable results, because error terms depend on the distance between images. In order to improve the solution, Govindu presents an iterative estimation that performs an adjustment of weights with the goal of a uniform weighting of individual constraints. On the one hand, the method does not handle gross errors and, therefore, is very sensitive to outliers in the pairwise relative translations. On the other hand, an algebraic error is minimized, which is known to be sub-optimal regarding the geometric characteristic of translations. As mentioned above, Govindu [2004] studies a joint estimation of rotations and translations as manifold-averaging in the Lie group $SE(3)$. This local method is efficient in finding the intrinsic average but is dependent on good initialization. In Crandall et al. [2011], translations are estimated using the MRF and discrete belief propagation, already described for the case of rotation estimation (cf. Section 3.3.1), again using a truncated cost function, which allows the introduction of prior information from geotags. Sinha et al. [2010] formulate linear constraints from pairwise point reconstruction, assuming global rotations to be known. They align two individual pairwise reconstructions in an image triplet, which means they estimate the remaining four parameters, i.e. scale and translation, using sample consensus [Torr & Zisserman, 2000]. From these tripletwise parameters, relative scales and translations are estimated for a set of image pairs in an over-determined system of linear equations covering all images that are to be oriented. It has been shown in subsequent works (e.g. [Arie-Nachimson et al., 2012; Ozyesil et al., 2015]) that this method does not produce very accurate results. Arie-Nachimson et al. [2012] propose a spectral linear method based on a novel decomposition of the essential matrix, depending on global rotations and translations. Assuming rotations to be known, translations are recovered via eigenvalue decomposition. Like the model in Govindu [2001], this pairwise method degenerates for images in a collinear arrangement. In Ozyesil & Singer [2015] it is shown that the estimation of global translations is only well posed for a parallel rigid set of relative translations (for more information about parallel rigidity see e.g. [Servatius & Whiteley, 1999]). Thus, locations are estimated for the maximum parallel rigid set of images using a convex L_1 optimization, similar to the cost function in Moulon et al. [2013] (cf. next subsection). Although the method is robust against outliers, it is shown in Cui & Tan [2015] that the approach like the work of Arie-Nachimson et al. [2012] produces degenerate solutions for collinear images.

Translation estimation based on triplet wise constraints

To cover this degeneracy, Jiang et al. [2013] present a linear model based on triplet constraints. Geometrically, these constraints can be described as a *triangulation* of the location of a third image from two overlapping images and their respective relative translation directions. Robustness against outliers in the relative translations is derived by several verification steps on the view-graph, including a comparison of relative translation directions before and after registration. Experiments reveal a more accurate estimation of locations of collinear images than in Arie-Nachimson et al. [2012] and precise results for the benchmark dataset presented in Strecha et al. [2008]. In Moulon et al. [2013], relative translation directions are derived from triplets of images. Then they formulate a linear optimization problem using the L_∞ norm including a linear cheirality constraint. The objective function describes the geometric distance of translations in object space. Results show a high accuracy regarding the benchmark datasets and a low run time for this approach.

Cui et al. [2015] present a linear method that uses point tracks over at least three images to build a system of constraints for the scale of the relative translations. The idea is that a pairwise locally reconstructed point must be equivalent in all possible combinations of image pairs in which the point is observed. Using a set of point tracks that covers the whole set of images, linear constraints are formulated and solved using L_1 optimization, which is shown to be more robust to outliers in the point observations. The model is robust against a collinear arrangement and weakly associated images and produces very accurate results.

Translation estimation based on n images

In Cui & Tan [2015], a model is presented, in which to every vertex in the view-graph, all adjacent vertices are used to compute a sparse depth image. Then, for each vertex, these depth images are merged in order to compute individual scales for each relative translation. Translation parameters for all the images are estimated in a linear L_1 optimization. Experiments show good results in terms of accuracy and run time.

Outlier detection using relative translations

Wilson & Snavely [2014] present a method for global translation estimation, in particular for large and unordered image sets. They perform an outlier search, similar in spirit to Zach et al. [2010], but using translations only. The three dimensional relative translation directions are projected to a random set of one-dimensional directions and the consistency problem reduces to fulfilling ordering constraints (called *minimum feedback arc set* in graph theory). Using heuristics, the originally NP-hard problem is approximated. Global translations are estimated in a non-linear least-squares optimization that, even with random initialization, achieves good results on large image sets from image-hosting websites[4]. It is shown in Cui et al. [2015] that this approach performs poorly on weakly associated images as well as in sequential image arrangements.

Synthesis - global translation estimation

Except for Govindu [2004], all presented approaches assume global rotations to be estimated beforehand, which allows various convex formulations of the translation estimation. Quasiconvex L_∞ estimation suffers from inefficiency, especially for large problems, and its sensitivity to outliers. The estimation based on pairwise constraints, while often having problems with certain acquisition geometries such as collinear images, are more efficient. More robust methods based on image triplets generally have advantages considering robustness against outliers in relative translations and collinear image arrangement.

The approach proposed in this work is based on the method of Cui et al. [2015], because it achieves very accurate results. In contrast to Cui et al. [2015], who perform a L_1 optimization, robustness is achieved by a new outlier elimination based on pairwise spatial intersection and a triplet wise scale constraint. The distinction between inliers and outliers is very discriminative so that the method is robust against outliers. Moreover, points are selected additionally based on their distribution in image space, which in theory leads to more stable results.

[4] These data is also provided by the authors at `http://www.cs.cornell.edu/projects/1dsfm/`.

3.4. Alternative approaches

Besides bundle adjustment, there are few works that comprise different strategies. Fusiello & Crosilla [2015] present an approach that solves orientation parameters via generalized anisotropic Procrustes analysis. This method does not depend on initialization of the unknowns but merely builds on alternately estimating image orientation parameters and scales for the intersection rays. The geometric cost function describes the distance of intersecting rays in object space. While achieving similar accuracy to bundle adjustment, the convergence of this approach tends to be quite slow.

4. A new method for the global image orientation

This chapter is devoted to a detailed explanation of the method newly developed for global image orientation. After a brief introduction into global image orientation and a comparison of the objective of this thesis with the state-of-the-art (4.1) the model is presented in three main parts: the estimation of relative orientations from homologous points (4.2), a combined convex estimation of global rotations (4.3) and the estimation of global translations based on local spatial intersections (4.4).

In the first part, an accurate and robust method to estimate relative orientations from homologous points is discussed. After the presentation of a constrained M-estimation (4.2.1), the focus lies on the robustness of the relative orientations regarding the contextual information from connected images. The image orientation problem is viewed as a graphical structure, for which certain constraints can be established that allow the detection of gross errors (4.2.2).

The second part comprises the estimation of global rotations and is divided into two subsequent optimizations. Firstly, the optimization is relaxed to find a SDP-formulation that allows to solve the problem in the convex hull of $SO(3)$ (4.3.1). Secondly, the solution serves as an accurate initialization for a subsequent iterative estimation in the Lie algebra $\mathfrak{so}(3)$ (4.3.2).

The third part is dedicated to the estimation of global translations with known rotations. Using rotated relative translations, linear homogeneous constraints are formulated based on local point reconstructions (4.4.1). This step requires a selection of suitable point tracks (4.4.2) and a detection of outliers in the image coordinates of the homologous points, for which a novel combination of the reprojection error and a triplet-scale constraint is introduced (4.4.3). Thereafter, a solving strategy is discussed (4.4.4).

The last section comprises information about the final robust bundle adjustment (4.5).

A detailed schematic illustration of the workflow is depicted in Figure 4.1. The four main parts (Sections 4.2-4.5) are highlighted by the gray areas. The three columns correspond to pointwise information (left), relative orientations (center) and global orientations (right).

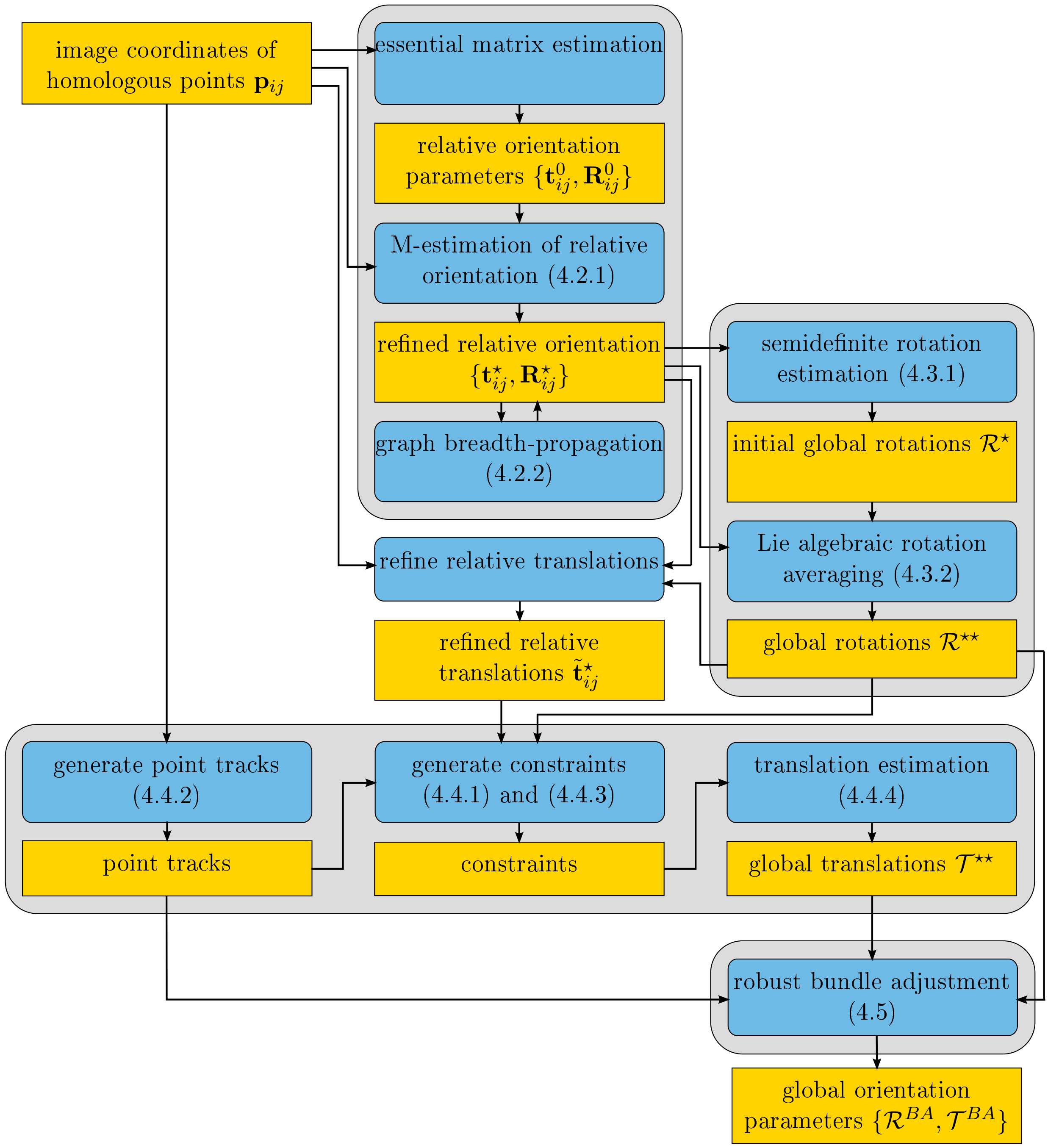


Figure 4.1.: Schematic illustration of the workflow. Blue boxes represent the individual steps of the method, yellow boxes the respective resulting parameters. Gray zones highlight the individual parts.

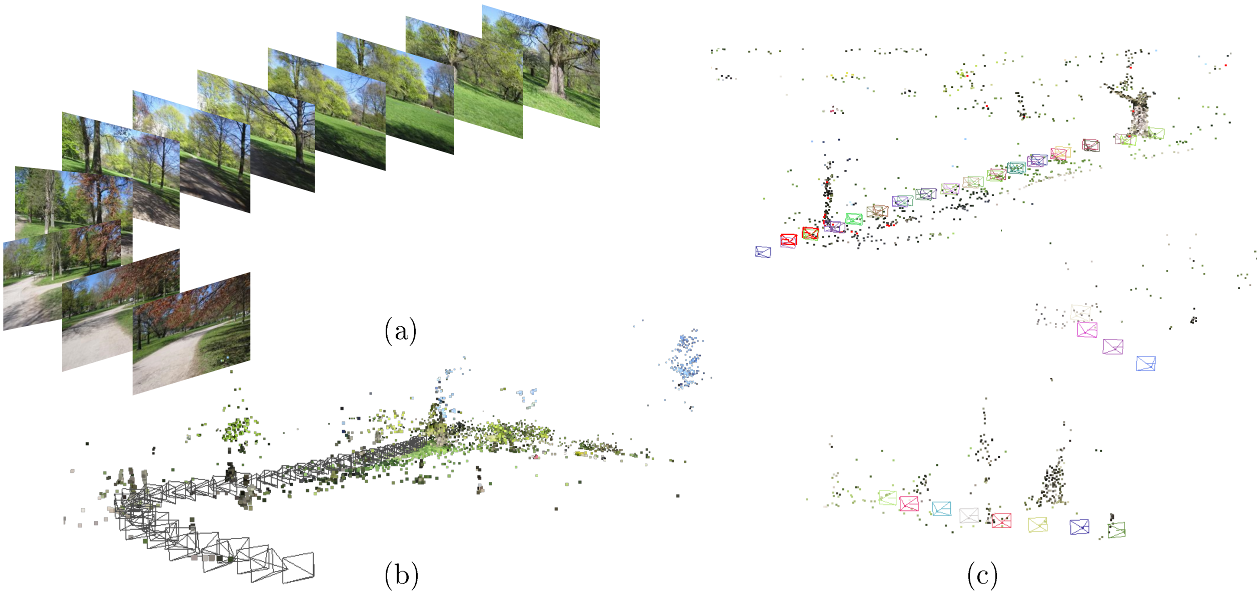


Figure 4.2.: Reconstruction of a park scene (4.2a) using the global image orientation proposed in this work (4.2b) and an incomplete reconstruction of the same scene using *VisualSFM* [Wu, 2013] (4.2c).

4.1. Global image orientation and objective revisited

This section gives a brief introduction to the proposed global image orientation approach and compares the objective of this thesis with available approaches from the state-of-the-art. The most characteristic feature of global image orientation methods is that they use all relative information that is available at once. This requires an extensive matching of images which is time consuming if no prior information regarding image position and rotation and thus possible overlap is available. The availability of this kind of information strongly depends on the type of images that shall be oriented. Images from aerial photogrammetry, for instance, come with quite accurate position and rotation information from differential GNSS and IMU and are taken in a regular pattern such that the overlapping regions are basically known in advance. Terrestrial moving platforms often use time stamps that lead to the assumption that consecutive images picture a similar scene. The inverse assumption does not hold because images, despite lying far apart on the time scale, may of course capture the same object. This is known as the loop-closure problem (e.g. [Meidow, 2012]). Images from terrestrial photogrammetry or from unordered datasets from image-hosting websites (e.g. [Wilson & Snavely, 2014]) often do not contain prior information of any kind. Whereas images taken for a photogrammetric purpose in general aim for precise measurements and therefore exhibit a suitable geometric setup regarding overlap and coverage of the object, images from hosting websites often have none of these characteristics but a high redundancy in terms of acquisitions from almost the same view point.

Sequential and hierarchical approaches (e.g. [Havlena et al., 2009; Wu, 2013]) in general only use a subset of the relative information available and depend on computationally demanding intermediate bundle adjustment. This limitation might lead to poor reconstruction results as is exemplarily depicted in Figure 4.2. In this example a set of collinear images was acquired and oriented using the method proposed in this work (4.2b) and the free software *VisualSFM* [Wu, 2013] (4.2c). The reason why *VisualSFM* was not able to compute the correct orientation cannot be determined with certainty.

In general, this degeneracy is more likely to happen for image sets whose acquisition was not planned for photogrammetric reconstruction. Nevertheless, in terms of error propagation, a joint estimation of all unknowns is always favorable because errors are distributed more equally.

In Section 1.1, one major objective of this thesis is presented: The proposed method shall allow the estimation of accurate image orientation parameters (1), it shall provide robustness against outliers in the relative orientations and image coordinates of homologous points (2) and it shall be applicable to a variety of different data, as mentioned in the previous paragraph (3). Before this method is studied in the following sections, its main directions are compared with the state-of-the-art.

A high accuracy is a feature desired by probably all image orientation methods, though perhaps not the main focus. Many approaches are merely driven by a reduction of computation time (e.g. [Govindu, 2004; Chatterjee & Govindu, 2013; Jiang et al., 2013]), which is not the main purpose of this work. A simple but effective method to enhance the accuracy is to improve relative orientations by a subsequent nonlinear M-estimation in a Gauss-Helmert-model (cf. Section 4.2.1). This additional step was not explicitly conducted before in the context of global image orientation. The detection and elimination of erroneous relative orientations in the proposed method is carried out within a new graph-based breadth-propagation algorithm (cf. Section 4.2.2), similar in spirit to the approaches in Zach et al. [2010] and Enqvist et al. [2011b] but without any limitations, e.g. regarding the length of cycles. Another advantage in comparison to methods based on robust cost functions like Hartley et al. [2011]; Chatterjee & Govindu [2013] is that outliers are eliminated and do not distort the result in any way.

The estimation of global rotations is based on the method of Govindu [2004] (cf. Section 4.3.2). It is efficient, uses the Lie group structure of rotations and, thus, sticks to a solution space on the rotation manifold and provides a maximum likelihood solution in the Lie algebra. Its dependency on initial rotations is covered by the estimation of initial rotations based on the SDP presented in Saunderson et al. [2014] (cf. Section 4.3.1). The estimation is augmented by prior knowledge in form of weights deduced from the variances, which are available thanks to the M-estimation. This often leads to more accurate results. Although such a confidence weighting is mentioned in some related works, it is always based on the number of correspondences, which is not necessarily a reliable indicator for the actual quality of the relative orientation. For instance, the distribution of homologous points in image space is not taken into account.

Translation estimation is based on Cui et al. [2015] (cf. Section 4.4). This method has advantages regarding collinear camera arrangements and weakly associated images and therefore is suitable for various types of data. However, in Cui et al. [2015], outliers in the image coordinates of the homologous points are detected by a robust L_1 estimation, which does not lead to an unbiased maximum likelihood solution. In the proposed method, robustness is achieved by introducing further constraints (cf. Section 4.4.3), so that outliers are eliminated before the estimation allowing a statistically more favorable L_2 estimation (cf. Section 4.4.4). Moreover, a novel strategy of selecting suitable points for the construction of constraints is used, which takes a favorable distribution in image space into account (cf. Section 4.4.2).

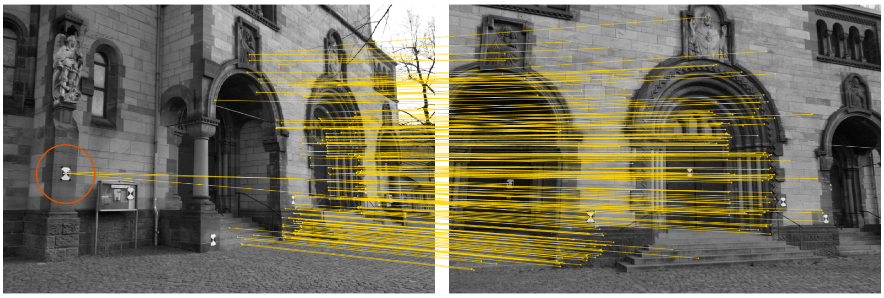
4.2. Preprocessing

In this section, the first part of the method is described, which is dedicated to provide a set of accurate relative orientations. For a set of n images $\mathcal{I} = \{\mathsf{I}_1, \dots \mathsf{I}_n\}$ it is assumed that image coordinates of homologous points $\mathbf{p}_{ij}, \forall (i,j) \in \{1, \dots, n\}, i < j$ have been found in case two images $\{\mathsf{I}_i, \mathsf{I}_j\}$ overlap. Every $\mathbf{p}_{ij}$ consists of a set of point tuples $\left(\mathbf{p}_i^l, \mathbf{p}_j^l\right)$, in which each tuple encompasses the observations in images I_i and I_j that are likely to represent the same point $\mathbf{P}^l$ in object space, whereas *likely* in this context refers to the metric of the SIFT descriptor [Lowe, 2004]. In practical applications, these points contain a considerable amount of outliers. For this set of n images, there exist at most $(n(n-1))/2$ possible pairwise combinations ij. For each of these combinations, a relative orientation is computed if the number of homologous points in $\mathbf{p}_{ij}$ is larger than a given threshold, i.e. $|\mathbf{p}_{ij}| > \tau_{|\mathbf{p}|}$. The number of relative orientations is denoted by m. Initial values for these m relative orientations $\{\mathbf{t}_{ij}^0, \mathbf{R}_{ij}^0\}$ are computed based on the essential matrix, which is estimated using the 5-point algorithm [Nistér, 2004] in a random sampling approach [Torr & , 2002], using a subsequent factorization [Hartley & Zisserman, 2003], pp. 239. This step will not be discussed in the remainder of this work. Most of the outliers in the homologous points are detected and excluded as shown exemplarily in Figure 4.3. However, the heuristic nature of such a random sampling approach is that the solution is not necessarily optimal and will vary if computed repeatedly. The optimal solution for the relative orientation is derived by a subsequent nonlinear maximum likelihood M-estimation (see e.g. [Huber et al., 1964; Hampel, 1968], or [Förstner et al., 2004], pp. 812), which is briefly studied in Section 4.2.1. Thereafter, a novel graph based approach to enhance the reliability of the relative orientations is presented in Section 4.2.2. A previous version of this approach is presented in Reich & Heipke [2016].

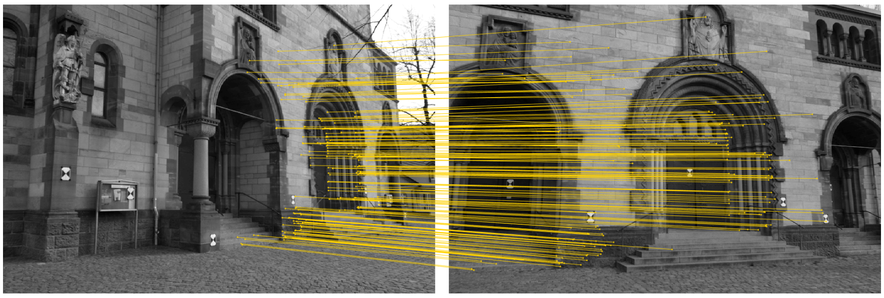
4.2.1. Constrained M-estimation of the relative orientation

As introduced in Section 2.1.2, the relative orientation is parameterized by the dependent-images representation. The functional model for the estimation is derived from the pointwise nonlinear coplanarity condition in equation (2.7), assuming normalized observations and ignoring the superscript ${}^{n}\cdot$:

$$c^l_{\{\mathbf{t}_{ij}, \mathbf{R}_{ij}\}} = \mathbf{p}_j^{l,T} \mathbf{R}_{ij}^T \left[\mathbf{t}_{ij}\right]_\times \mathbf{p}_i^l \stackrel{!}{=} 0. \tag{4.1}$$



(a) before RANSAC



(b) after RANSAC

Figure 4.3.: Homologous points before (4.3a) and after (4.3b) estimation of the initial relative orientation using the 5-point algorithm and RANSAC. The red circle highlights two outliers that have been eliminated.

This condition is solved in the sense of a *Gauss-Helmert model*. The partial derivatives of the condition equations with respect to the unknowns, stored in the Jacobian $\mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$, and to the observations, stored in matrix $\mathbf{B}$, are computed using the initial relative orientation $\{\mathbf{t}^0_{ij}, \mathbf{R}^0_{ij}\}$. For point l they read:

$$\underset{1\times 6}{\mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}^{\,l} = \frac{\partial c^l_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}{\partial \{\mathbf{t}_{ij}, \mathbf{R}_{ij}\}} = \left[\left(\left[\mathbf{R}_{ij}\, \mathbf{p}^l_j \right]_\times \mathbf{p}^l_i \right)^T, \left(\left[\mathbf{R}^T_{ij} \left[\mathbf{t}_{ij} \right]_\times \mathbf{p}^l_i \right]_\times \mathbf{p}^l_j \right)^T \right] \tag{4.2}$$

$$\underset{1\times 6}{\mathbf{B}}^{\,l} = \frac{\partial c^l_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}{\partial \left(\mathbf{p}^l_i, \mathbf{p}^l_j \right)} = \left[\mathbf{p}^{l,T}_j \mathbf{R}^T_{ij} \left[\mathbf{t}_{ij} \right]_\times, \left(\mathbf{R}^T_{ij} \left[\mathbf{t}_{ij} \right]_\times \mathbf{p}^l_i \right)^T \right]. \tag{4.3}$$

The linearized functional model of the Gauss-Helmert model reads as:

$$\mathbf{B}\mathbf{v} + \mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}\mathbf{x} + \mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} = \mathbf{0}, \tag{4.4}$$

with $\mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$, $\mathbf{B}$ and $\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$ representing all homologous points, constructed by stacking (4.1)-(4.3) for each homologous tuple in $\mathbf{p}_{ij}$, respectively. The vector $\mathbf{v}$ contains the residuals of the image coordinates in image space and the parameter vector $\mathbf{x} = [t_{ij,x}, t_{ij,y}, t_{ij,z}, \omega_{ij}, \varphi_{ij}, \kappa_{ij}]$ contains the unknowns, using the Euler angle representation. Note, that any other rotation representation like quaternions can also be used. A maximum likelihood estimation of the relative orientation parameters is derived by minimizing the sum of the squared and weighted residuals. Thus, the following

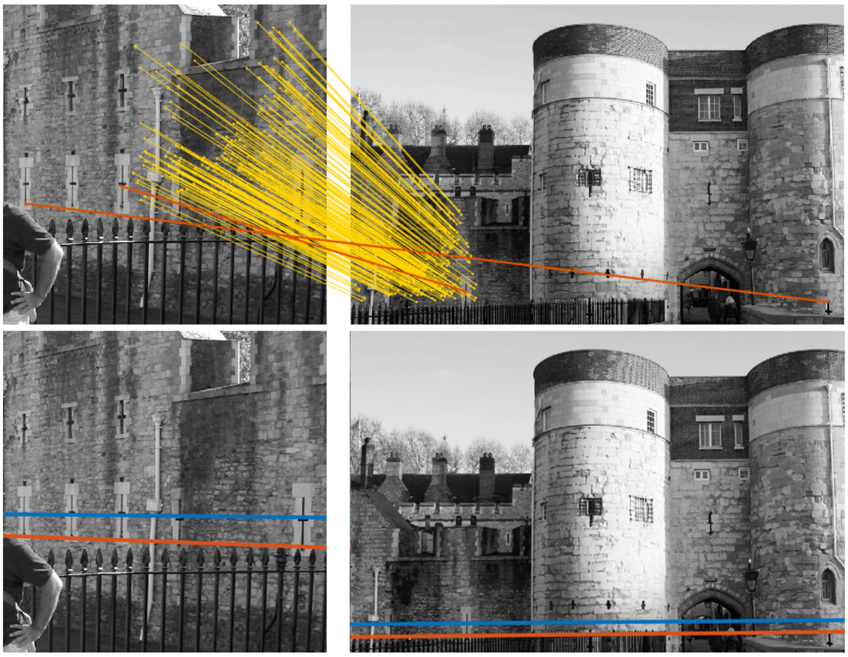


Figure 4.4.: Visualization of two outliers lying close to the epipolar plane. Top: inlier correspondencs after estimation of the essential matrix. After M-estimation of the relative orientation 2 outliers were found (red). Bottom: epipolar lines for both outliers after M-estimation of the relative orientation.

optimization problem is formulated:

$$\begin{aligned} &\text{minimize} && \mathbf{v}^T \mathbf{W}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} \mathbf{v} \\ &\text{subject to} && \mathbf{B}\mathbf{v} + \mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} \mathbf{x} + \mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} = \mathbf{0} \\ & && \left[\mathbf{t}_{ij}^T, \underset{1\times 3}{\mathbf{0}}\right] \mathbf{x} - 1 = 0 \ . \end{aligned} \tag{4.5}$$

The weight matrix $\mathbf{W}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}$ is applied for an iterative reweighting of the individual condition equations, which leads to the constrained M-estimation. Initially, observations in image space are considered equally accurate an uncorrelated. Then, individual weights for every point l are computed by $w^l_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} = \sigma_0^2 \left(\mathbf{B}^l \mathbf{\Sigma}_{\mathbf{p}_{ij}} \mathbf{B}^{l,T}\right)^{-1}$ using the covariance matrix of the observations $\mathbf{\Sigma}_{\mathbf{p}_{ij}}$. The second constraint in (4.5) eliminates the scale from the estimation and requires the length of the relative translation to be equal to one. Such a constrained optimization problem is generally solved using Lagrangian multipliers (cf. Section 2.3.3).

The estimation is iterated until in iteration b the factor $\frac{\|\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}\|^{b-1} - \|\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}\|^{b}}{\|\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}\|^{b-1} + \|\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}\|^{b}}$, which describes the change of the condition equation (Equation (4.1)) from one iteration to the next, is below a given threshold $\tau_{\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}$. Observations whose residual is higher than $3\hat{\sigma}_0^2$, using the empirical variance factor $\hat{\sigma}_0^2 = \left(\mathbf{v}^T \mathbf{W}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} \mathbf{v}\right) / \left(|\mathbf{p}_{ij}| - 5\right)$, are considered outliers and excluded from further computations.

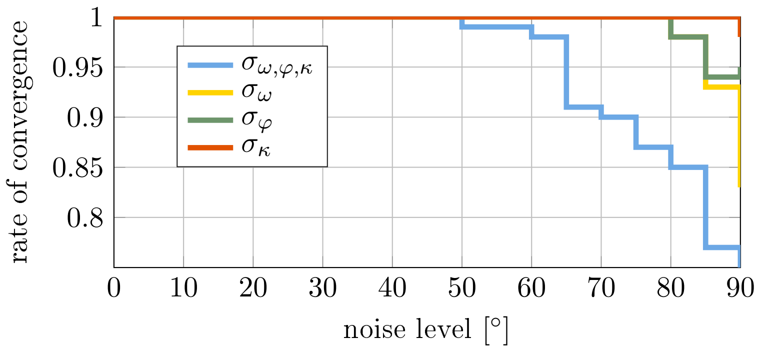


Figure 4.5.: Rate of convergence to the global optimum of M-estimation for different noise rates on individual Euler angles (yellow, green, red) and jointly on all Euler angles (blue).

Especially in scenes with repetitive structure it happens that two correspondences fulfill the epipolar constraint while they do not represent the same point in object space. Sometimes, these outliers are detected during the M-estimation because of the refined relative orientation. An example of such points can be seen as red matches in Figure 4.4. Although the correspondences still lie very close to the respective epipolar lines, the differences compared to the correct matches shown in yellow are high enough to be detected as outliers.

Finally, using the set of homologous points $\mathbf{p}^{\star}_{ij}$, from which these remaining outliers have been deleted, an estimation with unit weights is performed to derive the improved relative orientation $\{\mathbf{t}^{\star}_{ij}, \mathbf{R}^{\star}_{ij}\}$. From the inverse normal equation matrix the empirical covariance matrix of the relative orientation is computed:

$$\mathbf{\Sigma}_{\{\mathbf{t}^{\star}_{ij},\mathbf{R}^{\star}_{ij}\}} = \hat{\sigma}_0^2 \left(\mathbf{J}^T_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} \left(\mathbf{B}^T \mathbf{\Sigma}_{\mathbf{p}_{ij}} \mathbf{B} \right)^{-1} \mathbf{J}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}} \right)^{-1} . \tag{4.6}$$

This matrix parameterizes the confidence for each relative orientation and is used for a weighting during the estimation of global rotations and translations.

The optimization problem outlined in this section is not convex due to the condition equation (4.1) which is bivariate in the unknowns. The radius of convergence to reach the global optimum depends on the number and the distribution of observations in image space. Figure 4.5 shows a Monte-Carlo analysis for the estimation of the relative orientation, exemplarily for the image pair of Figure 4.3, varying the initial rotations. Euler angles are drawn uniformly in the interval $[-t5°, t5°]$, $0 \leq t \leq 18$, around the true Euler angles (see noise level on the x-axis). For each t, 100 independent trials are conducted using the modified Euler angles as initialization for the M-estimation of the relative orientation. The rate of convergence on the y-axis shows the proportion of trials, at which the optimization converged to a globally optimal solution. Until $\pm 45°$ variation in $(\omega, \varphi, \kappa)$ the estimation always reaches the global optimum. A variation in the translation direction only has a negligible influence on the convergence to the optimal relative orientation. Thus, in practice, the initialization from a sub-optimal essential matrix is accurate enough to derive the global optimum.

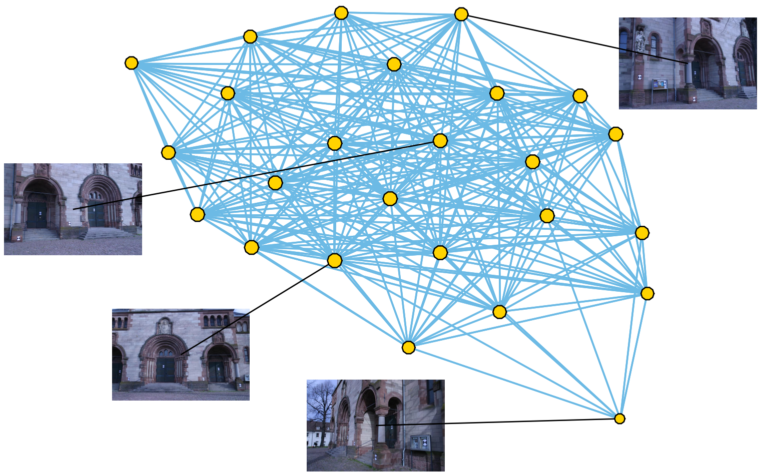


Figure 4.6.: View-graph of *Herz-Jesu-P25* (see Section 5.1 for a description of the data).

4.2.2. Propagation of the view-graph

In the previous section, it was shown how the precision of the relative orientation is improved so that they are optimal regarding pairwise correspondence constraints. In order to derive a set of accurate relative orientations, this is not sufficient. What if, for a pair of images $\{\mathsf{I}_i, \mathsf{I}_j\}$, the majority of homologous point tuples in $\mathbf{p}^{\star}_{ij}$ supports a relative orientation which is intrinsically justified considering the epipolar constraint but erroneous regarding the true underlying geometry? This may happen if the images picture scenes with repetitive structure. Further problems are planar scenes, for which the estimation of the relative orientation may cause problems, or an unfavorable distribution of correspondences in the image that might lead to unstable results.

In this section, a novel graph based algorithm is presented that detects outliers in the relative orientations using breadth-propagation. Besides outlier detection, initial values for the unknown global rotations are also generated. Breadth-propagation is conducted using relative rotations only, whereas the underlying graphical structure is defined for both, relative translation and rotation.

Definition 1. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph in which a vertex $\mathsf{V}_i \in \mathcal{V}$ represents the global orientation of image I_i and two vertices V_i and V_j are connected by an edge $\mathsf{E}_{ij} \in \mathcal{E}$ if a relative orientation between images I_i and I_j has been estimated. n is the number of vertices, $n = |\mathcal{V}|$, and m is the number of edges, $m = |\mathcal{E}|$. It is $|\mathcal{V}| = n$ and $|\mathcal{E}| = m$. This graph is called *view-graph*.

The concept of the view-graph is used frequently to describe the relation between relative and global orientation parameters [Govindu, 2006; Martinec & Pajdla, 2007; Sinha et al., 2010; Arie-Nachimson et al., 2012]. It allows an intuitive visualization of the orientation problem as depicted in Figure 4.6. Yellow vertices depict individual images, connected by blue edges if a relative orientation is estimated. In this example, the image corresponding to the lowest vertex has considerably less connections because it is taken from a relatively sharp angle.

The general problem of finding global orientation parameters can also be described in terms of

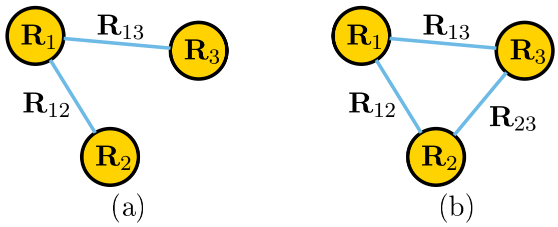


Figure 4.7.: Example of view-graphs. MST, all vertices are connected (4.7a) and redundant graph with a minimum cycle (4.7b).

a graphical model: *determine the vertices given edges, annotated with relative orientations, and a propagation constraint*. In case of global rotations, this model is applied using the functional relation between relative and global rotations given in Equation (2.4). Starting from an arbitrary vertex V_1 with fixed rotation $\mathbf{R}_1$, global rotations can be propagated to adjacent vertices by $\mathsf{V}_j \leftarrow \mathbf{R}_j = \mathbf{R}_1\mathbf{R}_{1j}$. Having exactly $m = n-1$ edges so that all n vertices are connected, the graph is a so-called *minimum spanning tree* (MST) (see Figure 4.7a). In this MST all global rotations can be estimated by successive propagation from one vertex to the next and there is only one solution. In the general case, the number of edges m is higher than $n-1$, at maximum $(n\,(n-1))\,/2$, which on the one hand makes propagation more complicated, but on the other hand induces redundancy that can be used to examine the correctness of the underlying graph, i.e. the relative rotations.

The redundancy in the view-graph reveals itself in terms of cycles. Propagating along a cycle until the starting vertex is reached induces a constraint to all relative rotations in the cycle. Assuming the smallest cycle possible, formed by the set $\{\mathbf{R}_{12}, \mathbf{R}_{23}, \mathbf{R}_{13}\}$, outlined in Figure 4.7b, the propagation constraint is written as:

$$\mathbf{R}_1 = \mathbf{R}_1\mathbf{R}_{12}\mathbf{R}_{23}\mathbf{R}_{13}^T \;\;\Leftrightarrow\;\; \mathbf{R}_{12}\mathbf{R}_{23}\mathbf{R}_{13}^T = \mathbf{I}_{3\times3}. \tag{4.7}$$

Note that $\mathbf{R}_{13}^T = \mathbf{R}_{31}$. Due to noise in the relative rotations, in practice, the condition in (4.7) is not fulfilled in general. Deviations above the expected imprecision of the relative rotations give evidence of the presence of an outlier in the relative rotations. From one cycle alone, only the existence of an outlier can be identified, whereas a decision, which rotation is wrong, is not possible. Zach et al. [2010] use loopy belief propagation in a Bayesian network, Enqvist et al. [2011b] successively add edges to a most reliable MST in order to identify erroneous edges. In the breadth-propagation algorithm presented in this section, cycles are used implicitly: Starting from an arbitrary vertex, rotations are propagated in the manner of a graph breadth-first-search [Moore, 1959; Gould, 1988]. This means that all adjacent vertices are estimated before a subsequent vertex is chosen to propagate from. Hence, the algorithm can be divided into $|\mathcal{V}| = n$ sequences. In the following, a superscript always refers to the sequence. The starting vertex of the first sequence, which is denoted as V^1, and the order in which the algorithm selects consecutive vertices are partly subject to heuristic decision. Of course, only vertices, which have an estimated rotation, can be selected as the starting vertex. It is a justified assumption that the propagation of noisy relative rotations causes a drift in the estimation of global rotations, i.e. for every concatenation of rotation matrices, the uncertainties are propagated as well. The selection of the first starting vertex V^1, proposed in the following, takes this into account.

Let $d_i^1 \in \mathbf{d}^1 \in \mathbb{Z}^n : 0 \leq d_i^1 \leq n-1$ denote the distance from the starting vertex V^1 to the vertex V_i, which is defined as the minimum number of edges between V^1 and V_i. If $\mathsf{V}^1 = \mathsf{V}_i$ then $d_i^1 = 0$ and if $\mathcal{G}$ is a MST then the maximum distance $\max\left(\mathbf{d}^1\right) = n-1$. In general, $\max\left(\mathbf{d}^1\right) \ll n-1$. Let the matrix $\mathbf{D}$ be an $n \times n$ distance matrix composed by stacking all $n \times 1$ vectors $\mathbf{d}^i$. This matrix contains all pairwise distances, including the zero-distance to the reference vertex itself. For example, the view-graph of *Herz-Jesu-P25* shown in Figure 4.6 has a maximum distance of 2, irrespective of the starting vertex V^1, so every vertex can be reached from every other vertex using at most two edges. Regarding a minimum drift, the task is to find a vertex V^1 so that $\max\left(\mathbf{d}^1\right)$ is minimal considering all possible vertices as starting vertex:

$$\begin{aligned} &\text{find} && \mathsf{V}^1 \\ &\text{subject to} && \max\left(\mathbf{d}^1\right) = \min\left(\max\left(\mathbf{D}\right)\right). \end{aligned} \tag{4.8}$$

In (4.8) the function $\max(\cdot)$ applied to a matrix returns the column-wise maximum value. This problem is solved using the adjacency matrix of the view-graph by applying the *Floyd-Warshall algorithm* [Floyd, 1962; Warshall, 1962; Gould, 1988] (cf. Appendix E for a pseudo code and a toy example).

Given the distance matrix $\mathbf{D}$, there might be more than one possible solution $\mathsf{V}^\star$ for V^1, i.e. more than one column with minimal maximal distance. Out of these solutions, the vertex with maximum degree is selected: $\mathsf{V}^1 \in \mathsf{V}^\star : |\mathsf{V}^1| = \max|\mathsf{V}^\star|$. If still there is no unique candidate, a random selection between the remaining vertices is made. From these algebraic distances, one might want to extrapolate a geometric interpretation of the image arrangement. It is likely that the image corresponding to the selected starting vertex lies approximately in the geometric barycenter of all images but this is, however, merely speculative.

The order of the breadth-propagation is defined by the number of examined edges incident to a vertex. Let $s \in \{1, 2 \ldots, n\}$ denote the current sequence of the algorithm and $\mathcal{E}_i \subseteq \mathcal{E}$ be the set of edges incident to vertex V_i. Then $\mathcal{E}_i^s \subseteq \mathcal{E}_i$ is the set of all *examined* edges at sequence s incident to V_i. From $s = 2$ on, remaining vertices are sorted in decreasing order in $|\mathcal{E}_i^{s-1}|$ with index i covering all vertices in $\mathcal{V} \setminus \{\mathsf{V}^1, \ldots, \mathsf{V}^{s-1}\}$. The starting point for a sequence $s \geq 2$ is the topmost element in this sorted list. If there is not a unique vertex V_i with maximum $|\mathcal{E}_i^{s-1}|$ the number of all incident edges $|\mathcal{E}_i|$ is used.

This procedure is visualized in Figure 4.8 for a small example. Sequences are numbered from I – VIII. The numbering of the vertices reflects the order in which each vertex is considered as starting vertex in the subsequent sequence and is updated in every sequence based on the number of examined incident edges. The first starting vertex V^1 is selected according to (4.8) which can

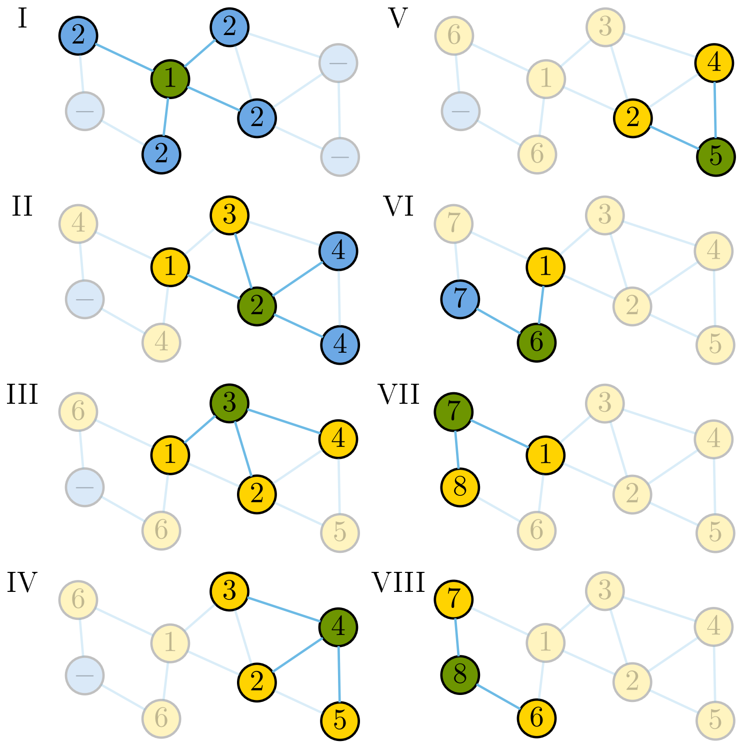


Figure 4.8.: Simulation of a graph breadth-propagation, starting from top left proceeding column wise from I to VIII. Propagation from the starting vertex (green) to non-estimated vertices (blue) and to already estimated vertices (yellow).

be easily verified because it is the only vertex with $\max\left(\mathbf{d}^1\right) = 2$. The actual starting vertex in every sequence is depicted in green. Vertices that have not been estimated yet are pictured in blue and vertices that have been estimated are shown in yellow; a distinction of these two cases will be given in the following paragraph. Only the highlighted part is affected in the respective sequence. After sequence I there are four candidates in $\mathsf{V}^\star$ for the starting vertex of sequence II. Then, the number of all incident edges to vertices in $\mathsf{V}^\star$ is decisive for the selection of the starting vertex V^2. In sequences VI and VII a random selection for the starting vertex is made. After $n = 8$ sequences the breadth-propagation stops. As can be seen, all vertices are yellow and all edges have been consulted twice.

If a vertex has not been estimated, i.e. the vertex is blue, a rotation is propagated from the starting vertex using (2.4). The case in which a vertex has been estimated before, i.e. the vertex is yellow, is equivalent to closing a cycle in the graph. In this case, the difference between the estimated rotation of vertex V_j, denoted as ${}^1\mathbf{R}_j$, and the propagation from V^s, denoted as ${}^2\mathbf{R}_j = \mathbf{R}^s\mathbf{R}^s_j$, is computed using the angular distance, $d_\alpha\left({}^1\mathbf{R}_j^T\,{}^2\mathbf{R}_j\right)$, defined in (2.14). Based on a threshold τ_α two different situations are distinguished:

1. If $d_\alpha\left({}^1\mathbf{R}_j^T\,{}^2\mathbf{R}_j\right) \leq \tau_\alpha$: The two rotations are similar, it is assumed that the cycle is free from outliers. Both rotations are averaged using single rotation averaging (cf. Appendix B)

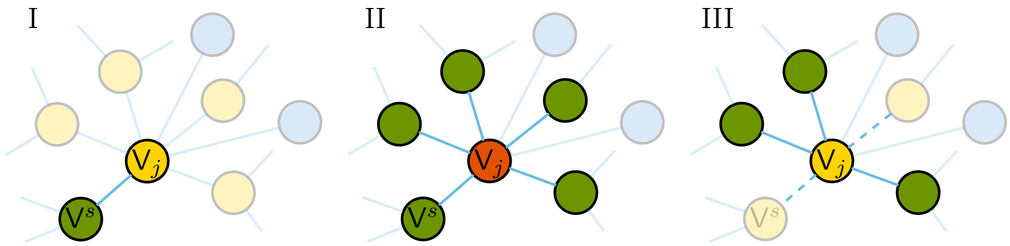


Figure 4.9.: Simulation of an inconsistent propagation. I: propagation from V^s to V_j that has been estimated before. II: both rotations are inconsistent, all estimated adjacent vertices (green) are consulted to make a propagation to V_j. III: V_j is averaged from the largest consistent set of propagated rotations (green), inconsistent edges (dashed lines) are considered to be classified as $\mathcal{E}^-$ (see text for more information).

2. If $d_\alpha\left({}^1\mathbf{R}_j^T\,{}^2\mathbf{R}_j\right) > \tau_\alpha$: The two rotations are inconsistent, it is assumed that the cycle is affected by outliers. Redundant information is consulted to identify the outlier (see the following paragraph).

Let the set of edges $\mathcal{E}$ be divided into two subsets $\mathcal{E}^+$ and $\mathcal{E}^-$ such that $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$ and $\mathcal{E}^+ \cap \mathcal{E}^- = \emptyset$. $\mathcal{E}^-$ is the set of all edges considered to be outliers and $\mathcal{E}^+$ are the inliers. In the beginning, all edges in $\mathcal{G}$ are classified as $\mathcal{E}^+$. If, during propagation to a vertex V_j, an inconsistency in a cycle is discovered, i.e. the second case is reached, classification of the involved edges into $\mathcal{E}^-$ and $\mathcal{E}^+$ is necessary. In order to find the erroneous relative rotation, the following sub-procedure, outlined in Figure 4.9, is applied. In case of an inconsistency (step I), all estimated vertices adjacent to V_j are consulted to make a propagation (step II). This expands the set of rotation estimates for V_j, $\mathcal{R}_j = \{{}^1\mathbf{R}_j, {}^2\mathbf{R}_j\}$, by additional $|\mathcal{E}_i^s| - 1$ estimates that allow a division of $\mathcal{R}_j$ into $\mathcal{R}_j^+$ and $\mathcal{R}_j^-$ with $\mathcal{R}_j = \mathcal{R}_j^+ \cup \mathcal{R}_j^-$, $\mathcal{R}_j^+ \cap \mathcal{R}_j^- = \emptyset$. The largest consistent subset $\mathcal{R}_j^+ \subseteq \mathcal{R}_j$, in which every element ${}^i\mathbf{R}_j \in \mathcal{R}_j^+$ is similar to every other with respect to the threshold τ_α, is taken to compute an average solution by single rotation averaging (step III). If there is not a unique largest subset, the rotation is not propagated. In future work the decision making could be enhanced by introducing prior information from the covariance matrices of the relative rotations.

Edges corresponding to $\mathcal{R}_j^-$, i.e. those connected to vertices from which the propagation did not lead to a consistent solution, are considered to be classified as outlier edges $\mathcal{E}^-$. A final classification requires the validation of two criteria: Firstly, $|\mathcal{R}_j^+|/|\mathcal{R}_j^-| \geq \tau_c$ must hold, thus the ratio between the number of consistent and inconsistent rotation estimates must be above a given threshold τ_c. This constraint avoids a premature classification of edges, for which a reliable decision cannot be made. τ_c is set to 1.5. Secondly, the respective complement vertices, corresponding to $\mathcal{R}_j^-$, must have been estimated redundantly by at least two additional adjacent vertices. This is reasonable because otherwise the complement vertex itself could be also erroneous and it could not be differentiated, which of the two examined incident edges is the outlier. These criteria are best explained considering Figure 4.9, step III. Consistent vertices are connected to V_j with a solid line, inconsistent vertices with a dashed line. The first criterion is fulfilled, because $|\mathcal{R}_j^+|/|\mathcal{R}_j^-| = 1.5$. Regarding the second criterion, the starting vertex V^s, besides its connection to V_j, has two further examined incident edges whereas the second complement vertex to the upper right of V_j only has one. Therefore, only the edge E_j^s is classified as $\mathcal{E}^-$. After the propagation, a final classification is conducted by computing the angular distance considering all edges $\mathcal{E}^+$ and vertices $\mathcal{V}$.

For this breadth-propagation to work properly, the following requirements are imposed that have to be fulfilled:

Lemma 1 (Redundancy)**.** *The redundancy of relative rotations with respect to global rotations is high enough. More specifically, the degree of any vertex $|\mathsf{V}_i|$ must be at least equal to three to be able to identify an outlier.*

A vertex with degree one is estimated only once and will never be in line for a second estimation. A vertex of degree two is estimated twice and thus the angular distance between both estimations is computed at some point in the algorithm. However, if $d_\alpha > \tau_\alpha$ a decision, which rotation is correct and which one is incorrect cannot be taken. If the degree is larger or equal to three, the possibility of a largest consistent subset and a rigorous classification of incident edges exists. In case no such set exists, i.e. all rotation estimates in $\mathcal{R}_j$ differ, or no final decision is reached, the relative rotation with smallest trace of its covariance matrix is assumed to be an inlier.

Lemma 2 (Outlier distribution)**.** *Outliers are distinguished from inliers based on the angular distance between two individual estimations $\{{}^1\mathbf{R}_j, {}^2\mathbf{R}_j\}$ and a threshold τ_α. It is unlikely that a propagation using two or more incident erroneous relative rotations leads to a consistent solution.*

In statistics, it is common to assume a uniform distribution for outliers (see e.g. Triggs et al. [2000]), which supports Lemma 2. The case, in which a consistent solution is derived from outliers only, however, cannot be precluded entirely in practice. It has to be noted that no outlier detection model can be considered immune to this issue.

The derivation of the breadth-propagation algorithm above and the two lemmata lay the ground for the following theorem.

Theorem 1 (Convergence of the breadth-propagation)**.** *Once the breadth-propagation is finished, every vertex reachable from V^1 has been estimated. Every edge $\mathsf{E}_{ij} \in \mathcal{E}^+$ has been examined at least twice.*

The proof of Theorem 1 is based on the proof of Theorem 2.1.3 in Gould [1988], p. 36.

Proof. First, the proof is given without considering noise. Suppose any vertex V_k that has an estimated rotation $\mathbf{R}_k$ and is reachable from V^1. Then, by applying the breadth-propagation explained above, there must exist an adjacent vertex V_j with the rotation $\mathbf{R}_k\mathbf{R}_{jk}^T$, which, in turn, is adjacent to a vertex V_i with $\mathbf{R}_k\mathbf{R}_{jk}^T\mathbf{R}_{ij}^T$. This is continued until, eventually, the vertex V^1 is reached with $\mathbf{R}_k\mathbf{R}_{jk}^T\mathbf{R}_{ij}^T \ldots \mathbf{R}_h^{1,T} = \mathbf{I}_{3\times3}$. For the noisy case the equality does not hold and has to be substituted with

$$\frac{d_\alpha\left(\mathbf{R}_j^T\mathbf{R}_k\mathbf{R}_{jk}^T\right)}{d_{jk}} < \tau_\alpha, \tag{4.9}$$

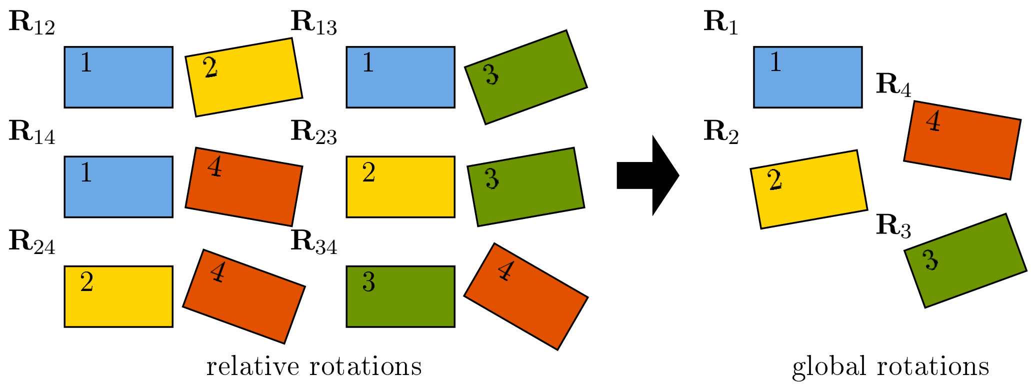


Figure 4.10.: Simplified visualization of of the problem of estimating global rotations from relative rotations in the two-dimensional case.

with d_{jk} being the number of edges between vertices j and k. Because every edge $\mathsf{E}_{ij} \in \mathcal{E}^+$ is incident to two vertices V_i and V_j, it is examined at least twice by applying the propagation rules defined above. $\square$

The breadth-propagation algorithm delivers a set of m^+ inlier relative rotations corresponding to $\mathcal{E}^+$ and initial global rotations $\mathcal{R} = \{\mathbf{R}_1 \ldots \mathbf{R}_n : \mathbf{R}_i \in SO(3)\}$. The classification of the relative rotations into $\mathcal{E}^+$ and $\mathcal{E}^-$ is applied to relative translations as well. This follows from the correlation of relative rotation and translation: a defective relative rotations that does not represent the underlying global geometry induces a defective relative translation.

4.3. Estimation of rotations

In the previous section, the focus was on the construction of a set of consistent and precise relative orientations using M-estimation and graph breadth-propagation. This chapter is dedicated to the first of the two types of image orientation parameters: global rotations. These have already been estimated as a byproduct of the breadth-propagation presented in Section 4.2.2. However, the propagated rotations $\mathcal{R}$ depend on the order of propagation. Moreover, the redundant relative information is used sequentially while in each sequence the majority of the global rotations is assumed to stay fixed. In this way, the rotations are estimated in the sense of a distributed consensus [Hartley et al., 2011, 2013] but not until convergence. Thus, the distribution of errors is not optimal.

The method presented in this chapter treats all global rotations at the same time using all relative orientations at once, which leads to a more equal distribution of errors. This problem is visualized in Figure 4.10, simplified for the two-dimensional case: From redundant pairwise relative rotations a set of global rotations is estimated in a joint coordinate system. The solution to this problem is twofold: Primarily, a relaxed convex optimization is conducted that delivers a solution in the convex hull of $SO(3)$ (Section 4.3.1). This solution is used as initialization for a subsequent estimation in the Lie algebra $\mathfrak{so}(3)$, described in Section 4.3.2.

4.3.1. Semidefinite estimation of rotations

The main purpose of the semidefinite estimation of global rotations described in this section is to derive initial global rotations $\mathcal{R}^\star$ for a subsequent nonlinear least-squares adjustment that are not dependent of the characteristics of breadth-propagation and represent a globally optimal solution regarding a relatively mildly relaxed optimization problem. The approach presented in this section is inspired by recent developments in optimization on algebraic groups [Horowitz et al., 2014; Saunderson et al., 2014], in which the problem is solved using a constrained convex relaxation. In the following, the nonconvex optimization is described followed by a presentation of the convex relaxation.

A nonconvex cost function can be derived from Equation (2.4):

$$d_{\mathbf{R}_{ij}} = \|\mathbf{R}_i^T\mathbf{R}_j - \mathbf{R}_{ij}\|_F, \tag{4.10}$$

which corresponds to the chordal distance defined in Equation (2.12). This cost function is nonconvex because of its bivariate part $\mathbf{R}_i^T\mathbf{R}_j$. It constitutes the cost in $\mathbb{R}^9$ without any intuitive geometrical interpretation. A least-squares solution is found by solving the following optimization problem:

$$\begin{aligned} &\underset{\{\mathbf{R}_1,\ldots,\mathbf{R}_n\}}{\text{minimize}} && \sum_{\forall(i,j)\in\mathcal{E}^+} d^2_{\mathbf{R}_{ij}} \\ &\text{subject to} && \mathbf{R}_1,\ldots,\mathbf{R}_n \in SO(3)\ . \end{aligned} \tag{4.11}$$

As already mentioned, neither the objective function nor the manifold $SO(3)$ are convex [Hartley et al., 2013], which makes a global estimation impossible. Even with accurate initial values for the unknowns, the inclusion of the determinant constraint, inherent to the $SO(3)$ manifold, is cumbersome and therefore often dropped [Martinec & Pajdla, 2007; Arie-Nachimson et al., 2012; Carlone et al., 2015].

In order to derive a convex optimization problem, both, the objective function and the $SO(3)$ constraint, have to be relaxed. For the former, a common strategy is to substitute the minuend in (4.10) with a 3×3 matrix $\mathbf{M}_{ij}$. This matrix is a submatrix of the $3n \times 3n$ symmetric matrix $\mathbf{M}$, which is constructed by a comprehensive multiplication of the global rotation matrices. Let $\mathbf{R} = [\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_n]$ be a $3 \times 3n$ matrix composed by stacking the global rotation matrices, then it is:

$$\mathbf{M} = \mathbf{R}^T\mathbf{R} = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{R}_1^T\mathbf{R}_2 & \cdots & \mathbf{R}_1^T\mathbf{R}_n \\ \mathbf{R}_2^T\mathbf{R}_1 & \mathbf{I}_{3\times3} & \cdots & \mathbf{R}_2^T\mathbf{R}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_n^T\mathbf{R}_1 & \mathbf{R}_n^T\mathbf{R}_2 & \cdots & \mathbf{I}_{3\times3} \end{bmatrix}\ . \tag{4.12}$$

$\mathbf{M}$ is a *Gramian* matrix and by definition positive semidefinite, i.e. $\mathbf{M} \succeq 0$, and $\mathbf{rank}\,(\mathbf{M}) = \mathbf{rank}\,(\mathbf{R}) = 3$. Furthermore, note that the main diagonal of $\mathbf{M}$ is composed of 3×3 identity matrices.

This is meaningful regarding the multiplication pattern because, recalling the group axioms no. (iii) and (iv) in Section 2.2 a rotation multiplied by its inverse (or transposed) gives the identity matrix.

Analogously, the subtrahend of (4.10) is substituted with $\mathbf{M}^0_{ij}$, a submatrix of $\mathbf{M}^0$, similar in design to $\mathbf{M}$:

$$\mathbf{M}^0 = \begin{bmatrix} \mathbf{I}_{3\times3} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1n} \\ \mathbf{R}_{12}^T & \mathbf{I}_{3\times3} & \cdots & \mathbf{R}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{1n}^T & \mathbf{R}_{2n}^T & \cdots & \mathbf{I}_{3\times3} \end{bmatrix} . \tag{4.13}$$

Again, the main diagonal is composed of identity matrices. The cost function (4.10) can now be written as:

$$d_{\mathbf{M}_{ij}} = \|\mathbf{M}_{ij} - \mathbf{M}^0_{ij}\|_F \ , \tag{4.14}$$

which is linear in the new unknown matrix $\mathbf{M}_{ij}$.

The matrix (4.13) is composed of the relative rotation matrices corresponding to $\mathcal{E}^+$. In general, this matrix is not fully occupied because not every pair of images overlaps, which is the fundamental requirement for the existence of a relative orientation. Thus, the cost function (4.14) is only deployed for those combinations for which an element in $\mathbf{M}^0$ is available. Matrix $\mathbf{M}$ in (4.12), though, must be fully occupied given the fact that every global rotation in $\mathbf{R}$ can be estimated. This problem is directly related to *matrix completion*, which deals with computing a full low rank matrix from a subset of its entries [Chen & Suter, 2004; Candès & Tao, 2010; Arrigoni et al., 2015b].

The cost function (4.14) is further simplified using the trace equivalence:

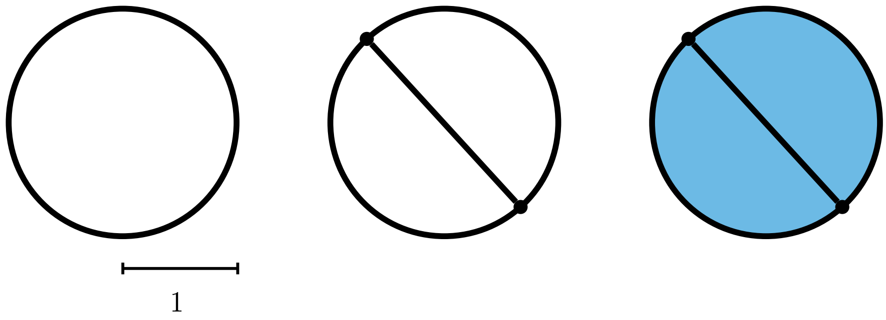
$$\begin{aligned} d_{\mathbf{M}_{ij}} = \|\mathbf{M}_{ij} - \mathbf{M}^0_{ij}\|_F &= \sqrt{\mathbf{tr}\left(\left(\mathbf{M}_{ij} - \mathbf{M}^0_{ij}\right)\left(\mathbf{M}^T_{ij} - \mathbf{M}^{0\,T}_{ij}\right)\right)} \\ &= \sqrt{\mathbf{tr}\left(2\mathbf{I} - 2\mathbf{M}_{ij}\mathbf{M}^{0\,T}_{ij}\right)} \\ \text{because} \quad \mathbf{tr}\left(\mathbf{M}_{ij}\mathbf{M}^{0\,T}_{ij}\right) &= \mathbf{tr}\left(\mathbf{M}^T_{ij}\mathbf{M}^0_{ij}\right) . \end{aligned} \tag{4.15}$$

Without loss of generality, Equation (4.15) can be squared and the constant part can be extracted in order to derive the final convex cost function,

$$d_{\mathbf{M}_{ij}} = -\mathbf{tr}\left(\mathbf{M}_{ij}\mathbf{M}^{0\,T}_{ij}\right) \ . \tag{4.16}$$

The maximum trace of a rotation matrix is equal to 3. In case of noise-free relative rotations, the product in (4.16) yields the 3×3 identity matrix and the minimum cost would be equal to $-3m^+$. In practice, however, this assumption does not hold but gives a lower bound on the optimal value.

Before the convex optimization problem is established, the $SO(3)$ constraint has to be relaxed. One way of doing this is to drop the determinant constraint and require orthogonality only, which extends the set of feasible matrices to $O(3)$, the set of orthogonal 3×3 matrices. These matrices



(a) (b) (c)

Figure 4.11.: Convex relaxation of $SO(2)$. Set of $SO(2)$ (4.11a). Connection between two elements of $SO(2)$ (4.11b). Convex hull of $SO(2)$ (4.11c).

can either have a determinant equal to 1 or -1; matrices with a negative determinant correspond to reflections. A correction of those matrices with determinant equal to -1 could be achieved after optimization by negation but there is no guarantee that these negated matrices lead to the optimal solution.

A tighter relaxation is proposed in Horowitz et al. [2014] and Saunderson et al. [2014], in which the matrices are required to be in the convex hull of $SO(3)$. It was shown in Section 2.3.1 that the convex hull of a nonconvex set is the minimum convex set that contains the set itself (cf. Equation (2.21)).

In the following, the derivation of the convex hull is outlined for the 2D case, in which an intuitive visualization is possible (Figure 4.11). Rotations in 2D are defined as:

$$SO(2) = \left\{ \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} : \theta \in [0, 2\pi) \right\}. \tag{4.17}$$

They include a well known analogy to unit-length complex numbers, which form the 1-sphere in two-dimensional space. Thus, an alternative description is:

$$SO(2) = \left\{ \begin{bmatrix} x & y \\ -y & x \end{bmatrix} : x^2 + y^2 = 1 \right\}. \tag{4.18}$$

This 1-sphere is visualized in Figure 4.11a. Applying the criterion for a convex set, developed in Section 2.3.1, one sees that $SO(2)$ is nonconvex. This is shown in Figure 4.11b, where a shortest connection of two elements in $SO(2)$ is almost entirely outside the set. The convex hull of $SO(2)$ is the unit disk, pictured in Figure 4.11c:

$$\mathbf{conv}\,(SO(2)) = \left\{ \begin{bmatrix} x & y \\ -y & x \end{bmatrix} : x^2 + y^2 \leq 1 \right\}. \tag{4.19}$$

3-dimensional rotations cannot be visualized as easily. A derivation of the convex hull is possible via unit quaternions, which form a 3-sphere in $\mathbb{R}^4$. This derivation according to Horowitz et al. [2014] and Sanyal et al. [2011] is provided in Appendix A.

Horowitz et al. [2014] and Saunderson et al. [2014] show that the convex hull of $SO(3)$ can be described as an LMI, which is the solution space of SDP (cf. Section 2.3.2). Explicitly, $\mathbf{conv}\,(SO(3))$ is defined as:

$$\mathbf{conv}\,(SO(3)) = \Bigg\{\mathbf{R} \in \mathbb{R}^{3\times 3} : \begin{bmatrix} 1+R_{11}+R_{22}+R_{33} & R_{32}-R_{23} & R_{13}-R_{31} & R_{21}-R_{12} \\ R_{32}-R_{23} & 1+R_{11}-R_{22}-R_{33} & R_{21}+R_{12} & R_{13}+R_{31} \\ R_{13}-R_{31} & R_{21}+R_{12} & 1-R_{11}+R_{22}-R_{33} & R_{32}+R_{23} \\ R_{21}-R_{12} & R_{13}+R_{31} & R_{32}+R_{23} & 1-R_{11}-R_{22}+R_{33} \end{bmatrix} \succeq 0 \Bigg\}. \tag{4.20}$$

The R_{ij} refer to the respective values in $\mathbf{R}$.

Now that the objective function and the constraint of optimization problem (4.11) are relaxed, the final convex optimization problem is established:

$$\begin{aligned} \underset{\mathbf{M}_{ij}}{\text{minimize}} \quad & \sum_{\forall (i,j)\in\mathcal{E}^+} w_{\mathbf{M}_{ij}} d_{\mathbf{M}_{ij}} \\ \text{subject to} \quad & \mathbf{M} \succeq 0 \\ & \mathbf{M}_{ii} = \mathbf{I}_{3\times 3} \\ & \mathbf{M}_{ij} \in \mathbf{conv}\,(SO(3)), \;\; \forall (i,j) \in \{1,\dots,n\}, i<j. \end{aligned} \tag{4.21}$$

The weights $w_{\mathbf{M}_{ij}}$ are derived from the covariance matrix of the relative rotations $\mathbf{\Sigma}_{\mathbf{R}^\star_{ij}}$ (see Equation (4.6)),

$$w_{\mathbf{M}_{ij}} = \frac{1}{1+\mathbf{tr}\left(\mathbf{\Sigma}_{\mathbf{R}^\star_{ij}}\right)}, \tag{4.22}$$

which results in weights $w_{\mathbf{M}_{ij}} \in (0,1]$. Note that the rank constraint for $\mathbf{M}$ is dropped because it is nonconvex. While the cost function is only deployed for combinations of (i,j) for which a relative rotation exists, the constraints are applied to all submatrices in $\mathbf{M}$. The first constraint is given by the condition that $\mathbf{M}$ is a Gramian matrix. The second constraint requires that every $\mathbf{R}_i$ is orthogonal, thus $\mathbf{R}_i \in O(3)$, whereas the third constraint states that every $\mathbf{R}_i$ is in the convex hull of $SO(3)$, thus $\mathbf{R}_i \in \mathbf{conv}\,(SO(3))$. It is proven in Saunderson et al. [2014] that the orthogonality constraint in combination with the convex hull constraint leads to a solution in $SO(3)$:

$$\mathbf{conv}\,(SO(3)) \cap O(3) = SO(3) \tag{4.23}$$

A detailed workflow of the semidefinite estimation of rotations is depicted in Figure 4.12. The optimization problem (4.21) is a SDP (cf. Section 2.3.2), thus its standardized formulation allows the

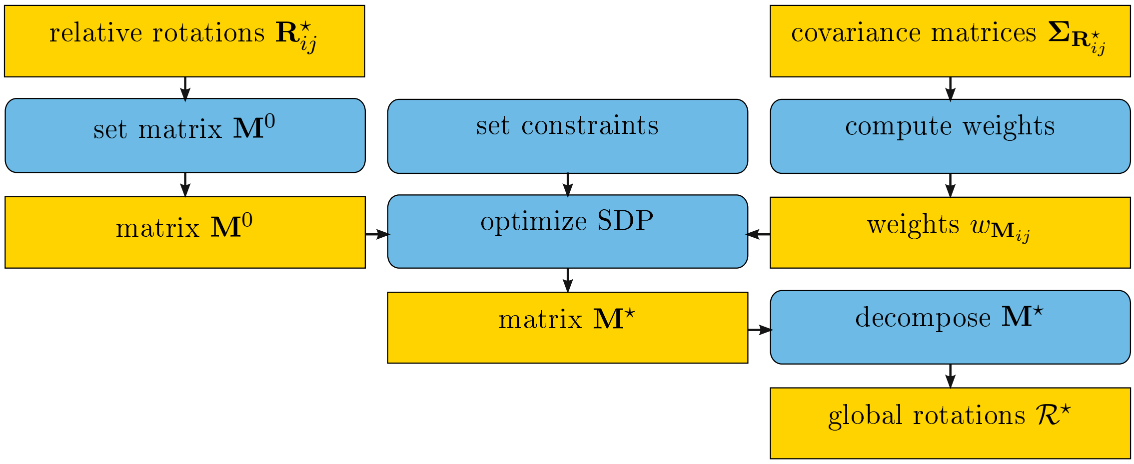


Figure 4.12.: Detailed workflow of semidefinite estimation of rotations. Blue boxes show the individual steps, yellow boxes the parametric outcome.

usage of various solvers. Once an optimal solution $\mathbf{M}^\star$ is found, the global rotations $\mathbf{R}^\star_i \in \mathcal{R}^\star$ need to be extracted. This is performed by a factorization, inverse to Equation (4.12), hence by an eigenvalue decomposition of $\mathbf{M}^\star$. The estimated global rotations are given by the three leading eigenvectors, corresponding to the nonzero eigenvalues. These matrices are multiplied by $\left(\mathbf{R}^{\star,1}\right)^T$ corresponding to V^1 in order to derive a solution compatible with the solution of the breadth-propagation of Section 4.2.2, in which the rotation of the starting vertex V^1 is equal to the identity matrix.

Remark. Solving the optimization problem defined in (4.21) becomes computationally intense when the problem size grows because the number of constraints grows quadratically in n. In order to approach this issue it might be sufficient to apply only a subset of the constraints $\mathbf{M}_{ij} \in \mathbf{conv}\left(SO(3)\right)$. How this affects the solution will be examined in Section 5.2.1.

4.3.2. Lie algebraic rotation averaging

In the previous section, it was shown how global rotations for all n images are estimated in a SDP using all relative rotations at once. This leads to a globally optimal solution $\mathcal{R}^\star$ with respect to a relaxed objective function and solution manifold. The minimum of this objective function, given the necessary simplifications, does not comprise the maximum likelihood solution of the original problem, however. The goal of this section is to approach this issue. In the following, an iterative least-squares optimization is presented, in which the domain of the objective function is defined in $SO(3)$ and which is solved in the respective Lie algebra $\mathfrak{so}(3)$. The solution is a set of optimal global rotations $\mathcal{R}^{\star\star}$. This procedure is similar to the approach of Govindu [2004], where it was defined for the Euclidean group $SE(3)$ that contains translation parameters as well. In addition to Govindu [2004], the approach presented in this thesis introduces prior information in form of weights computed from the covariance matrices of the relative rotations.

Again, the functional model is derived from Equation (2.4):

$$\Delta\mathbf{R}_{ij} = \mathbf{R}_i \mathbf{R}^\star_{ij} \mathbf{R}_j^T \quad . \tag{4.24}$$

In contrast to $d_{\mathbf{R}_{ij}}$ or $d_{\mathbf{M}_{ij}}$ in (4.10) and (4.16), respectively, whose domains are $\mathbb{R}^+$, in (4.24) we have $\mathbf{dom}\ \Delta\mathbf{R}_{ij} = SO(3)$. If the relative rotations were noise-free, $\Delta\mathbf{R}_{ij}$ would be the 3×3 identity matrix. Equation (4.24) is nonlinear in the unknown global rotations. In order to linearize it let $\Delta\mathbf{R}_i^{\star\star} = \mathbf{R}_i^{\star\star}\mathbf{R}_i^{\star T}$ be the parameter offset between the initial rotation $\mathbf{R}_i^\star \in \mathcal{R}^\star$ (cf. Section 4.3.1) and the optimal global rotation $\mathbf{R}_i^{\star\star} \in \mathcal{R}^{\star\star}$. Furthermore, let $\Delta\mathbf{R}_{ij}^0 = \mathbf{R}_i^\star\mathbf{R}_{ij}^\star\mathbf{R}_j^{\star T}$ be the reduced relative rotation. Using these definitions and applying $\mathbf{R}_i^{\star\star}$ for the unknown global rotations, (4.24) can also be written as:

$$\Delta\mathbf{R}_{ij} = \underbrace{\mathbf{R}_i^{\star\star}\mathbf{R}_i^{\star T}}_{\Delta\mathbf{R}_i^{\star\star}}\ \underbrace{\mathbf{R}_i^\star\mathbf{R}_{ij}^\star\mathbf{R}_j^{\star T}}_{\Delta\mathbf{R}_{ij}^0}\ \underbrace{\mathbf{R}_j^\star\mathbf{R}_j^{\star\star T}}_{\Delta\mathbf{R}_j^{\star\star T}} . \tag{4.25}$$

Note that $\mathbf{R}_i^{\star T}\mathbf{R}_i^\star = \mathbf{R}_j^{\star T}\mathbf{R}_j^\star = \mathbf{I}_{3\times3}$.

The next step of linearization considers the geometric properties of $SO(3)$ as a differentiable manifold (cf. Section 2.2.1). A mapping to the tangent space, the Lie algebra, $SO(3) \rightarrow \mathfrak{so}(3)$, is performed by applying the logarithm map:

$$\log\left(\Delta\mathbf{R}_{ij}\right) = \log\left(\mathbf{R}_i^{\star\star}\mathbf{R}_i^{\star T}\mathbf{R}_i^\star\mathbf{R}_{ij}^\star\mathbf{R}_j^{\star T}\mathbf{R}_j^\star\mathbf{R}_j^{\star\star T}\right) . \tag{4.26}$$

The linearized cost function is obtained by an approximation to (4.26) using the *Baker-Campbell-Hausdorff* formula, which gives the solution to $\log(\mathbf{XY})$ for any $(\mathbf{X},\mathbf{Y})$ in an arbitrary Lie group [Gilmore, 1974]. Neglecting higher order terms, a first order approximation is given by $\log(\mathbf{XY}) = \log(\mathbf{X}) + \log(\mathbf{Y})$. Accordingly, (4.26) can be approximated by:

$$\underbrace{\log\left(\Delta\mathbf{R}_{ij}\right)}_{\Delta\mathbf{r}_{ij}} \approx \underbrace{\log\left(\mathbf{R}_i^{\star\star}\mathbf{R}_i^{\star T}\right)}_{\Delta\mathbf{r}_i^{\star\star}} + \underbrace{\log\left(\mathbf{R}_i^\star\mathbf{R}_{ij}^\star\mathbf{R}_j^{\star T}\right)}_{\Delta\mathbf{r}_{ij}^0} - \underbrace{\log\left(\mathbf{R}_j^{\star\star}\mathbf{R}_j^{\star T}\right)}_{\Delta\mathbf{r}_j^{\star\star}} \tag{4.27}$$

$$\Delta\mathbf{r}_{ij} \approx \Delta\mathbf{r}_i^{\star\star} + \Delta\mathbf{r}_{ij}^0 - \Delta\mathbf{r}_j^{\star\star}, \tag{4.28}$$

with (4.28) being linear in the unknown parameter offsets. Note that $\log\left(\mathbf{X}^T\right) = -\log(\mathbf{X})$. This approximation is exact only in case the Lie bracket, i.e. the commutator $[\mathbf{X},\mathbf{Y}] = \mathbf{XY} - \mathbf{YX}$, is zero, which is in general not true for elements of $SO(3)$ or $\mathfrak{so}(3)$, because rotations are not commutative. Moreover, using (4.28) as a global metric to define distances between two rotations is not recommended, because it is not invariant to rotations, thus $\log(\mathbf{ZX}) + \log(\mathbf{ZY}) \neq \log(\mathbf{X}) + \log(\mathbf{Y}), \mathbf{Z} \in SO(3)$ [Hartley et al., 2013]. However, in the course of the Lie algebraic averaging algorithm presented in this section, offsets between the initial rotations $\mathbf{R}_i^\star$ and the optimal values $\mathbf{R}_i^{\star\star}$ are assumed to be small, i.e. only a small region in $SO(3)$ close to the identity is considered.

Having defined a linearized cost function (4.28), the optimization problem can be established. For that reason, Equation (4.28) is written in vectorial terms of a *Gauss-Markov model* $\mathbf{v} = \mathbf{J}\Delta\mathbf{x} - \Delta\mathbf{l}$.

Let $\Delta\mathbf{r}$ be a $3m^+ \times 1$ vector of residuals, in which the individual costs of (4.28) are stacked and let $\Delta\mathbf{r}^{\star\star} = [\Delta\mathbf{r}_1^{\star\star}, \ldots, \Delta\mathbf{r}_n^{\star\star}]^T$ be a $3n \times 1$ vector of parameters, in which the unknown global rotations in axis-angle representation are stacked. Analogously, $\Delta\mathbf{r}^0$ is a $3m^+ \times 1$ vector of reduced relative rotations. The $3m^+ \times 3n$ Jacobian matrix $\mathbf{J} = \left[\partial\Delta\mathbf{r}_{ij}/\partial\Delta\mathbf{r}_i^{\star\star}, \partial\Delta\mathbf{r}_{ij}/\partial\Delta\mathbf{r}_j^{\star\star}\right], \forall(i,j) \in \mathcal{E}^+$ is composed by positive and negative 3×3 identity matrices. Then, ignoring the approximate, (4.28) is written as:

$$\Delta\mathbf{r} = \mathbf{J}\Delta\mathbf{r}^{\star\star} + \Delta\mathbf{r}^0. \tag{4.29}$$

This leads to the unconstrained quadratic optimization problem

$$\underset{\{\Delta\mathbf{r}_1^{\star\star},\ldots,\Delta\mathbf{r}_n^{\star\star}\}}{\text{minimize}} \qquad \Delta\mathbf{r}^T\mathbf{W}_{\Delta\mathbf{r}}\Delta\mathbf{r} \ . \tag{4.30}$$

$\mathbf{W}_{\Delta\mathbf{r}}$ is a $[3m^+ \times 3m^+]$ weight matrix, computed from each individual covariance matrix of the relative rotation $\boldsymbol{\Sigma}_{\mathbf{R}_{ij}^\star}, \forall(i,j) \in \mathcal{E}^+$ in order to control its influence on the solution, using a sigmoidal weighting function [Krarup et al., 1980]:

$$\begin{aligned} \mathbf{W}_{\Delta\mathbf{r}_{ij}} &= \frac{1}{1 + \left(\lambda\ \mathbf{tr}\left(\boldsymbol{\Sigma}_{\mathbf{R}_{ij}^\star}\right)\right)^\nu}\mathbf{I}_{3\times3} \\ \lambda &= \frac{1}{w_{0.5}}, \quad \nu = 4. \end{aligned} \tag{4.31}$$

Parameters λ and ν control the weighting, in particular, regarding the position of the *halfweight* $w_{0.5}$, the point where the weight is equal to 0.5, and the steepness at that point. These values have to be adjusted respecting the expected accuracy of the relative rotations. ν is set to a common value in photogrammetric estimation (see e.g. Rottensteiner [2001]) and $w_{0.5}$ is set according to the maximum trace of the involved covariance matrices:

$$w_{0.5} = \max_{\forall(i,j)\in\mathcal{E}^+} \left(\mathbf{tr}\left(\boldsymbol{\Sigma}_{\mathbf{R}_{ij}}\right)\right) \ , \tag{4.32}$$

which encourages a rather soft weighting and avoids large discrepancies between individual weights.

The averaging is performed iteratively and, after each iteration, the parameters are updated by $\mathbf{R}_i^{\star\star} = \exp\left(\Delta\mathbf{r}_i^{\star\star}\right)\mathbf{R}_i^\star$. The updated parameters serve as initialization for the subsequent iteration. This corresponds to an alternating map between the manifold $SO(3)$ and its Lie algebra $\mathfrak{so}(3)$. A detailed workflow of the Lie algebraic averaging is depicted in Figure 4.13. Note that the Jacobian $\mathbf{J}$ is constant consisting of only identity- and zero-submatrices and only has to be computed once. Thus, only the reduced relative rotations have to be recomputed in every iteration taking the updated global rotations into account. Iteration is performed until convergence, i.e. the factor $\frac{(\Delta\mathbf{r}^T\mathbf{W}_{\Delta\mathbf{r}}\Delta\mathbf{r})^{b-1} - (\Delta\mathbf{r}^T\mathbf{W}_{\Delta\mathbf{r}}\Delta\mathbf{r})^b}{(\Delta\mathbf{r}^T\mathbf{W}_{\Delta\mathbf{r}}\Delta\mathbf{r})^{b-1} + (\Delta\mathbf{r}^T\mathbf{W}_{\Delta\mathbf{r}}\Delta\mathbf{r})^b}$ in iteration b must be below a given threshold $\tau_{\Delta\mathbf{r}}$.

Remark. The iterative design of the presented algorithm in principle allows for a reweighting based on the respective residual for the purpose of detecting outliers in the relative rotations as proposed in Chatterjee & Govindu [2013]. This may further extend the outlier detection during the view-graph

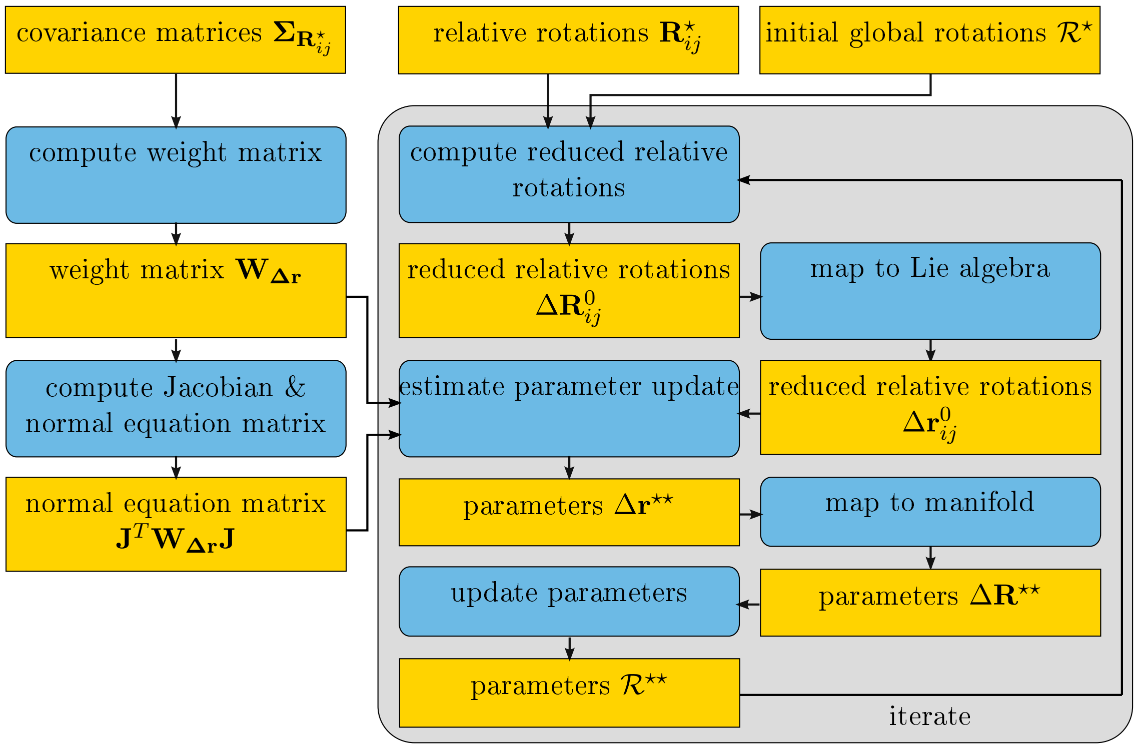


Figure 4.13.: Detailed workflow of Lie algebraic rotation averaging. Blue boxes show the individual steps, yellow boxes the parametric outcome.

breadth-propagation. In this case, the computation of the weight matrix and the Jacobian (see Figure 4.13) must be conducted within each iteration.

4.4. Estimation of translations

The previous chapter was dedicated to the estimation of global rotations. In this chapter, the focus lies on the second type of parameters that is necessary to describe the image orientation: global translations. In most approaches of global image orientation, the estimation of translations is founded on known rotations (e.g. [Ke & Kanade, 2007; Martinec & Pajdla, 2007; Arie-Nachimson et al., 2012; Arrigoni et al., 2015b; Cui et al., 2015]). A reason for this is that the task reduces to a convex optimization problem.

The functional model in form of linear homogeneous constraints, proposed in Section 4.4.1, is similar to the model presented in Cui et al. [2015] using local pairwise intersections. A visualization of this problem is shown in Figure 4.14: Given pairwise local reconstructions of one point, rotated by the global rotations, the goal is to derive a consistent solution, which corresponds to a unique point position. In the course of this thesis, the model of Cui et al. [2015] is extended in two significant ways: Firstly, a *new strategy to find homologous point tracks*, taking also the point distribution in image space into account, is explained in Section 4.4.2. This is followed by a *novel outlier detection* based

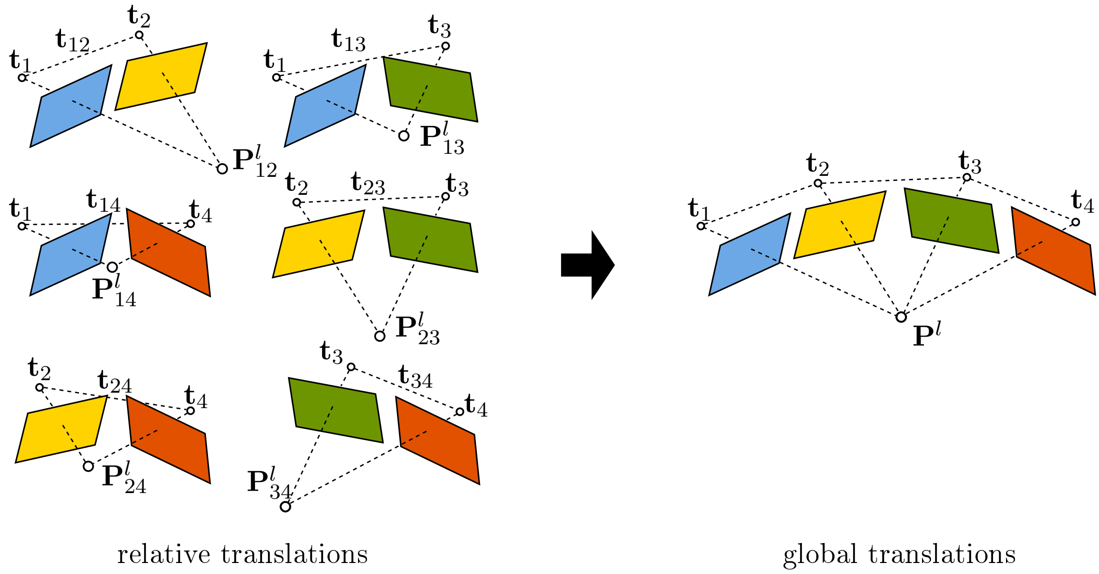


Figure 4.14.: Visualization of the problem of estimating global translations from global rotations, relative translations, and a homologous point.

on a combination of a reprojection of pairwise spatial intersections and a triplet-scale-constraint, derived in Section 4.4.3, which overcomes the frequent occurrence of erroneous homologous points. In this way, also outliers that cannot be found in a pairwise setting, e.g. because they fulfill the epipolar constraint, are detected. Finally, in Section 4.4.4 the optimization problem will be established and the strategy to solve it will be given.

4.4.1. Construction of linear constraints

In this section, it is shown how linear constraints for the estimation of the global translation parameters are derived from global rotations, relative translation directions and homologous points. It is assumed that the relative translation directions are rotated by the global rotations so that they represent the relative translations in the global instead of the local pairwise coordinate system, i.e. $\tilde{\mathbf{t}}^{\star}_{ij} = \mathbf{R}^{\star\star}_i \mathbf{t}^{\star}_{ij}, \forall (i,j) \in \mathcal{E}^+$. The only requirement in order to formulate these constraints is that there are points, which have been observed in at least three images. In the following, these constraints will be deduced using the example of the image triplet $\mathcal{I} = \{\mathsf{I}_i, \mathsf{I}_j, \mathsf{I}_k\}$.

Let there be one object point $\mathbf{P}^l$ that is observed in the image triplet $\mathcal{I}$. Given the global rotations $\{\mathbf{R}^{\star\star}_i, \mathbf{R}^{\star\star}_j\}$ and the relative translation $\tilde{\mathbf{t}}^{\star}_{ij}$, for which the functional relation to the unknown global translations is given by $\tilde{\mathbf{t}}^{\star}_{ij} = \frac{\mathbf{t}_j - \mathbf{t}_i}{\|\mathbf{t}_j - \mathbf{t}_i\|}$, point $\mathbf{P}^l$ can be estimated by spatial intersection. Noisy rotations and relative translations in general make a unique solution impossible, hence the rays do not intersect in object space. A common way to derive an approximate solution is to take the average, i.e. the mid-point of the common perpendicular of both rays as depicted in Figure 4.15a, which is called the *mid-point-method* (e.g. [Hartley & Sturm, 1997]). The spatial intersection equation is

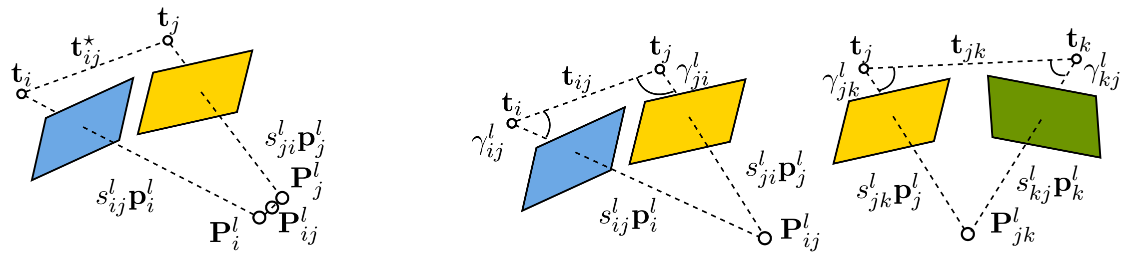


Figure 4.15.: Pairwise spatial intersection leads to two solutions $\{\mathbf{P}^l_i, \mathbf{P}^l_j\}$ that are closest to each other in object space (4.15a). Two pairwise spatial intersections of the same point $\mathbf{P}^l$ (4.15b).

derived from equation (2.2) by inversion:

$$\mathbf{P}^l_i = \mathbf{t}_i + s^l_{ij}\mathbf{p}^l_i, \qquad \mathbf{P}^l_j = \mathbf{t}_j + s^l_{ji}\mathbf{p}^l_j, \quad \text{with} \tag{4.33}$$
$$\mathbf{p}^l_i = \mathbf{R}^{\star\star}_i {}^n\mathbf{p}^l_i, \qquad \mathbf{p}^l_j = \mathbf{R}^{\star\star}_j {}^n\mathbf{p}^l_j,$$
$$\mathbf{P}^l_{ij} = \frac{1}{2}\left(\mathbf{P}^l_i + \mathbf{P}^l_j\right). \tag{4.34}$$

s^l_{ij} and s^l_{ji} are unknown scale factors, which are lost by projection to image space and which can be estimated in an unconstrained linear program using the mid-point-method. It shall be noted that this method does not give an optimal solution in the sense of least squared residuals in image space, which is true in particular for a projective coordinate frame in which common perpendicularity and mid-point are meaningless, i.e. they are not invariant to projective transformation. In the course of this thesis, an Euclidean (and thus metric) coordinate frame is used. For the ease of reading in the following, all point observations are assumed to be rotated and the superscript ${}^n\cdot$ is ignored, as it is defined in (4.33). Moreover, speaking of general geometric properties, specific instances of the relative translations, i.e. $\cdot^\star$, are ignored and homologous rays are assumed to intersect in object space.

The estimated object coordinates of $\mathbf{P}^l_{ij}$ are computed relative to images I_i and I_j and only determined up to a 7 parameter 3D Helmert transformation (see also Figure 2.1 in Chapter 2). Because global rotations $\mathcal{R}^{\star\star}$ are assumed to be known, three of these seven parameters are given. The four remaining parameters are three translations and one scale, thus a linear transformation which applies to the triangle $\Delta\left(\mathbf{t}_i\mathbf{t}_j\mathbf{P}^l_{ij}\right)$ with the three edges $\{\mathbf{t}_{ij}, s^l_{ij}\mathbf{p}^l_i, s^l_{ji}\mathbf{p}^l_j\}$ as depicted on the left side in Figure 4.15b.

In order to formulate a constraint from (4.33) regarding the global translations of all images in $\mathcal{I}$, point $\mathbf{P}^l$ is reconstructed by another pairwise local spatial intersection using the third image I_k

(see Figure 4.15b on the right side):

$$\mathbf{P}_j^l = \mathbf{t}_j + s_{jk}^l \mathbf{p}_j^l, \qquad \mathbf{P}_k^l = \mathbf{t}_k + s_{kj}^l \mathbf{p}_k^l, \text{ with} \tag{4.35}$$

$$\mathbf{P}_{jk}^l = \frac{1}{2}\left(\mathbf{P}_j^l + \mathbf{P}_k^l\right). \tag{4.36}$$

By assuming that all observations $\{\mathbf{p}_i^l, \mathbf{p}_j^l, \mathbf{p}_k^l\}$ correspond to the very same point in object space, equations (4.34) and (4.36) can be equated, neglecting the multiplication by 1/2:

$$\mathbf{t}_i + s_{ij}^l \mathbf{p}_i^l + \mathbf{t}_j + s_{ji}^l \mathbf{p}_j^l = \mathbf{t}_j + s_{jk}^l \mathbf{p}_j^l + \mathbf{t}_k + s_{kj}^l \mathbf{p}_k^l. \tag{4.37}$$

While there are only three independent point observations, four different scales $\{s_{ij}^l, s_{ji}^l, s_{jk}^l, s_{kj}^l\}$ have to be computed. Constraint (4.37) could be used to form a linear equation system in the manner of $\mathbf{Jx} = \mathbf{b}$, in which the Jacobian is populated by positive and negative 3×3 identity matrices. However, $\mathbf{t}_j$ would drop out of the equation, which is not desirable. Instead, every constant factor in Equation (4.37) can be also written in terms of the unknown global translations. Let us consider the triangle $\Delta\left(\mathbf{t}_i \mathbf{t}_j \mathbf{P}_{ij}^l\right)$ exemplarily. Because the directions of all edges in this triangle are known in the global coordinate system, both observations $\mathbf{p}_i^l$ and $\mathbf{p}_j^l$ can be described in terms of the relative translation:

$$\mathbf{p}_i^l = \mathbf{R}_{\gamma_{ij}^l} \mathbf{t}_{ij} \quad \mathbf{p}_j^l = \mathbf{R}_{\gamma_{ji}^l}(-\mathbf{t}_{ij}). \tag{4.38}$$

$\mathbf{R}_{\gamma_{ij}^l}$ and $\mathbf{R}_{\gamma_{ji}^l}$ are two rotation matrices that can be computed from angle $\gamma_{ij}^l = \arccos\left(\mathbf{t}_{ij} \cdot \mathbf{p}_i^l\right)$ and axis $\bar{\mathbf{r}}_{ij}^l = \mathbf{t}_{ij} \times \mathbf{p}_i^l$ and likewise for $\mathbf{R}_{\gamma_{ji}^l}$ using Equation (2.9). With $\mathbf{t}_{ij} = \frac{\mathbf{t}_j - \mathbf{t}_i}{\|\mathbf{t}_j - \mathbf{t}_i\|}$, Equation (4.37) changes to:

$$\mathbf{t}_i + s_{ij}^l \mathbf{R}_{\gamma_{ij}^l} \frac{\mathbf{t}_j - \mathbf{t}_i}{\|\mathbf{t}_j - \mathbf{t}_i\|} + \mathbf{t}_j + s_{ji}^l \mathbf{R}_{\gamma_{ji}^l} \frac{\mathbf{t}_i - \mathbf{t}_j}{\|\mathbf{t}_j - \mathbf{t}_i\|} = \mathbf{t}_j + s_{jk}^l \mathbf{R}_{\gamma_{jk}^l} \frac{\mathbf{t}_k - \mathbf{t}_j}{\|\mathbf{t}_k - \mathbf{t}_j\|} + \mathbf{t}_k + s_{kj}^l \mathbf{R}_{\gamma_{kj}^l} \frac{\mathbf{t}_j - \mathbf{t}_k}{\|\mathbf{t}_k - \mathbf{t}_j\|}. \tag{4.39}$$

Using
$$\begin{aligned} \mathbf{C}_{ij}^l &= \frac{s_{ij}^l}{\|\mathbf{t}_j - \mathbf{t}_i\|} \mathbf{R}_{\gamma_{ij}^l}, \quad & \mathbf{C}_{ji}^l &= \frac{s_{ji}^l}{\|\mathbf{t}_j - \mathbf{t}_i\|} \mathbf{R}_{\gamma_{ji}^l}, \\ \mathbf{C}_{jk}^l &= \frac{s_{jk}^l}{\|\mathbf{t}_k - \mathbf{t}_j\|} \mathbf{R}_{\gamma_{jk}^l}, \quad & \mathbf{C}_{kj}^l &= \frac{s_{kj}^l}{\|\mathbf{t}_k - \mathbf{t}_j\|} \mathbf{R}_{\gamma_{kj}^l}, \end{aligned}$$
(4.39) reduces to:

$$\begin{aligned} & \mathbf{t}_i + \mathbf{C}_{ij}^l (\mathbf{t}_j - \mathbf{t}_i) + \mathbf{t}_j + \mathbf{C}_{ji}^l (\mathbf{t}_i - \mathbf{t}_j) = \mathbf{t}_j + \mathbf{C}_{jk}^l (\mathbf{t}_k - \mathbf{t}_j) + \mathbf{t}_k + \mathbf{C}_{kj}^l (\mathbf{t}_j - \mathbf{t}_k) \\ \Leftrightarrow \quad & \left(\mathbf{C}_{ij}^l - \mathbf{C}_{ji}^l\right)(\mathbf{t}_j - \mathbf{t}_i) + \mathbf{t}_i + \mathbf{t}_j = \left(\mathbf{C}_{jk}^l - \mathbf{C}_{kj}^l\right)(\mathbf{t}_k - \mathbf{t}_j) + \mathbf{t}_j + \mathbf{t}_k. \end{aligned} \tag{4.40}$$

The derivation of (4.40) is exemplary for the combination $\{\mathsf{l}_i \mathsf{l}_j\}$ and $\{\mathsf{l}_j \mathsf{l}_k\}$. Of course, it can be also formulated for the two remaining combinations $\{\mathsf{l}_i \mathsf{l}_j\}$,$\{\mathsf{l}_i \mathsf{l}_k\}$ and $\{\mathsf{l}_i \mathsf{l}_k\}$,$\{\mathsf{l}_j \mathsf{l}_k\}$. By putting everything on one side of the equation, the three following linear homogeneous constraint functions

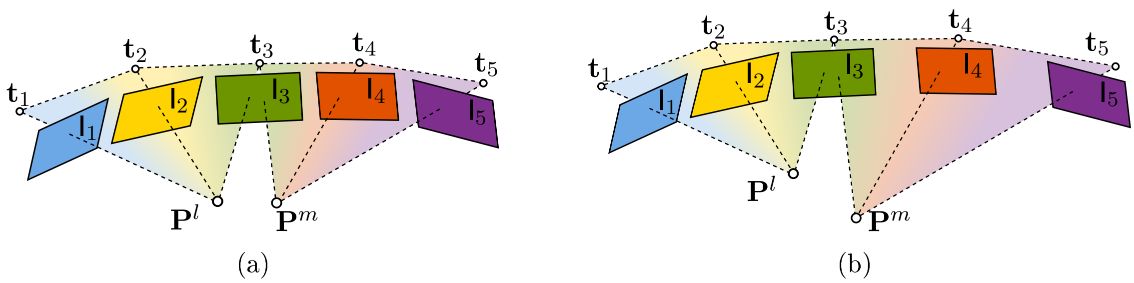


Figure 4.16.: Non-overlapping point track and two possible solutions

are established:

$$\left(\mathbf{I}_{3\times3} - \mathbf{C}^l_{ij} + \mathbf{C}^l_{ji}\right)\mathbf{t}_i + \left(\mathbf{C}^l_{ij} - \mathbf{C}^l_{ji} + \mathbf{C}^l_{jk} - \mathbf{C}^l_{kj}\right)\mathbf{t}_j + \left(-\mathbf{I}_{3\times3} - \mathbf{C}^l_{jk} + \mathbf{C}^l_{kj}\right)\mathbf{t}_k = \mathbf{0}. \tag{4.41}$$

$$\left(-\mathbf{C}^l_{ij} + \mathbf{C}^l_{ji} + \mathbf{C}^l_{ik} - \mathbf{C}^l_{ki}\right)\mathbf{t}_i + \left(\mathbf{I}_{3\times3} + \mathbf{C}^l_{ij} - \mathbf{C}^l_{ji}\right)\mathbf{t}_j + \left(-\mathbf{I}_{3\times3} - \mathbf{C}^l_{ik} + \mathbf{C}^l_{ki}\right)\mathbf{t}_k = \mathbf{0}. \tag{4.42}$$

$$\left(\mathbf{I}_{3\times3} - \mathbf{C}^l_{ik} + \mathbf{C}^l_{ki}\right)\mathbf{t}_i + \left(-\mathbf{I}_{3\times3} + \mathbf{C}^l_{jk} - \mathbf{C}^l_{kj}\right)\mathbf{t}_j + \left(\mathbf{C}^l_{ik} - \mathbf{C}^l_{ki} - \mathbf{C}^l_{jk} + \mathbf{C}^l_{kj}\right)\mathbf{t}_k = \mathbf{0}. \tag{4.43}$$

$$\text{with}\quad \mathbf{C}^l_{ik} = \frac{s^l_{ik}}{\|\mathbf{t}_k - \mathbf{t}_i\|}\mathbf{R}_{\gamma^l_{ik}}, \qquad \mathbf{C}^l_{ki} = \frac{s^l_{ki}}{\|\mathbf{t}_k - \mathbf{t}_i\|}\mathbf{R}_{\gamma^l_{ki}}.$$

Only two of these three equations are linearly independent (e.g. (4.43) results from the difference of (4.41) and (4.42)). In Section 4.4.3, it will be shown how these redundant constraints are used to detect outliers in the homologous points. Note that the only unknowns in (4.41) - (4.43) are the global translations $\{\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_k\}$, no object coordinates of $\mathbf{P}^l$ are estimated. Often, a point is visible in more than three images. In this case, constraints can be formulated for every independent triplet of images that can be composed.

Given a set of n images $\mathcal{I} = \{\mathsf{I}_1 \dots \mathsf{I}_n\}$, it is unlikely that there is *one* point, which is visible in *every* image. In order to be able to estimate the global translations of all images, points have to be found that fulfill a certain overlapping-criterion. A set of points that covers all images is called an *overlapping point track*.

Definition 2. An overlapping point track is a set of points, so that in every image in $\mathcal{I}$, at least one of these points is observed. If the overlapping point track has more than one element, every point in this set must be observed in at least two images, in which one or more other points are also observed.

It is important that two points within an overlapping point track share at least two images in which they are observed. A counterexample is visualized in Figure 4.16. Both subfigures 4.16a and 4.16b show the same point track. Triplets are marked in light colors. The object points $\mathbf{P}^l$ and $\mathbf{P}^m$ only share image I_3, in which they both are observed. Thus, the scale information cannot be transferred between both triplets and both solutions in 4.16a and 4.16b are possible. In consequence, the problem cannot be solved.

In Figure 4.17, an example of an overlapping point track is shown. Both triplets share the edge

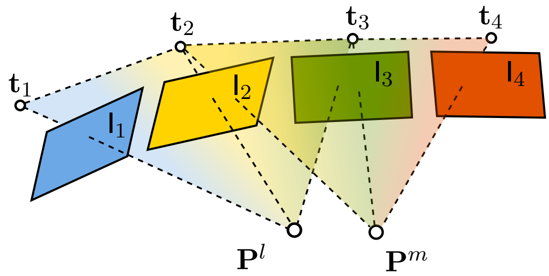


Figure 4.17.: Overlapping point track. Only one solution is possible

between $\mathbf{t}_2$ and $\mathbf{t}_3$ and there is only one possible solution. Hence, when constructing point tracks, this requirement has to be taken into account. A strategy to select appropriate points is examined in the following section.

4.4.2. Selection of points

In the previous section, it was explained how linear homogeneous constraints for the unknown global translations are derived from pairwise relative spatial intersections. It was shown that an overlapping point track has to be found in order to estimate global translations for all n images. Considering all existing homologous points would be computationally too demanding, making a selection of points inevitable. This section examines the criteria under which these points are selected.

These criteria are heuristic, twofold and originate from certain point characteristics that, given the knowledge about how the functional constraints are established, tend to lead to an accurate solution. On the one hand, a point is favored the more images it is observed in. These points are indicative of being trustworthy, bear a strong constraint over a large number of images and many different triplet combinations can be used to build constraints, enhancing the robustness (cf. Section 4.4.3). On the other hand, the distribution of points in image space has an impact on the estimation of global translations. The second criterion is not obvious, because the estimation of the location of the image is not related to the direction of rays, as, for instance, when considering spatial resection. It is derived via scale relations in image triplets, which per se do not depend on the diversity of ray directions. Positive effects of a wide distribution of observations, however, reveal themselves indirectly in form of a higher variety regarding depth and visibility in different sets of images. It follows that points observed in a very narrow region in image space are more likely have similar depth and are observed in similar image sets. Consequences this may entail are pictured in Figures 4.18c and 4.18d. A depth, large compared to the base length, leads to a glancing intersection and an imprecise estimation of the scales, shown with the error ellipses in Figures 4.18a and 4.18c. This reflects on the constraint functions (4.41)-(4.43) and, thus, on the estimation quality of the global translations. Not considering the distribution of points in image space might lead to a situation shown in Figure 4.18d, in which, for instance, points are chosen only based on the number of images they are measured in. A counterexample (Figure 4.18b) shows that next to a higher variety of depths also further triplet combinations may occur. This clearly enhances the statistical accuracy

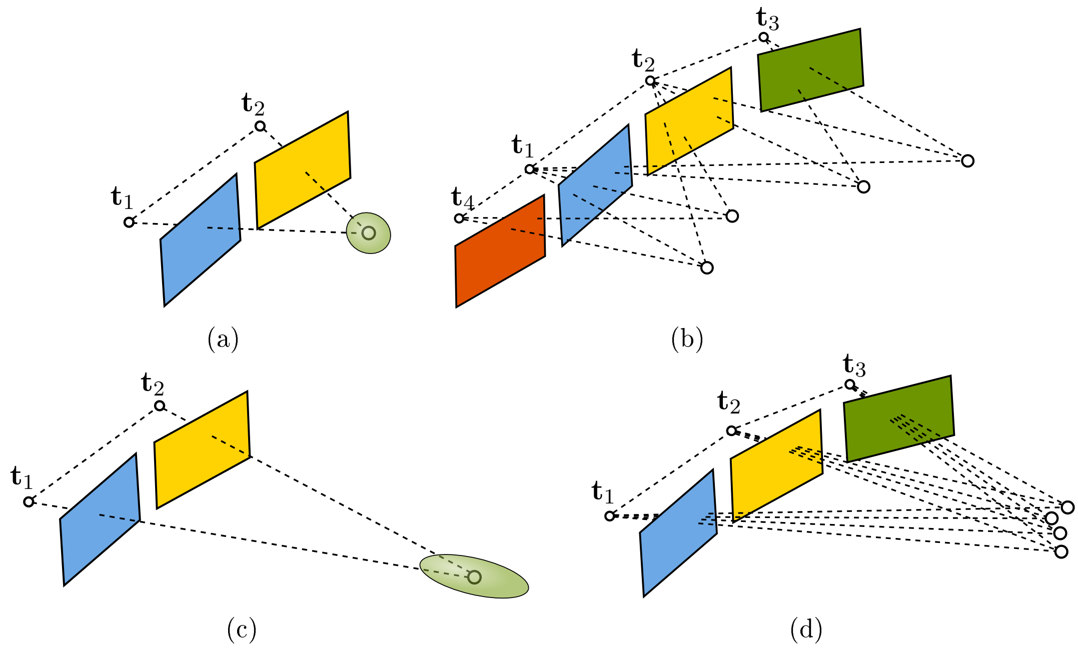


Figure 4.18.: Different selection of points: suitable intersection (a) and points with wide distribution in image space (considering image I_1) (b); glancing intersection (c) and points with narrow distribution in image space (d).

and robustness, because the constraints are founded on different relative translations.

A set of point tracks scoring highest in both criteria while accomplishing the overlapping requirement derived in Section 4.4.1, in general, is highly unlikely, because often points with the largest number of observations tend to be in similar areas of the images. Furthermore, evaluating the distribution of points in image space while considering a new candidate is cumbersome: firstly, the number of points already added might vary from image to image and, secondly, the distribution might be suitable in one image and inappropriate in another. Finding a proper threshold that distinguishes between various distributions, the number of points per image and the total number of images in which the point is observed, is delicate. Thus, in order to avoid to unnecessarily complicating the point selection, the following twofold strategy is used: In the first part, a fixed number of point tracks is constructed while the distribution criterion is only applied to the image with minimum distance to all other images regarding the view-graph, i.e. the image corresponding to the starting vertex V^1 (cf. Section 4.2.2). In the following, this image will be referred to as the starting image. In the second part, the distribution of points in all images is examined successively and additional points are added if the distribution is not good enough according to the distribution criterion.

All points are sorted by the number of images they are measured in. The image space is divided into five regions as shown in Figure 4.19. In this way, four points in four different regions are necessary to guarantee that at least one point lies apart from the others. For the first point track, the topmost point observed in any region in the starting image is taken. The region this point is observed in is flagged. The point track is then completed by stepping down the sorted point list

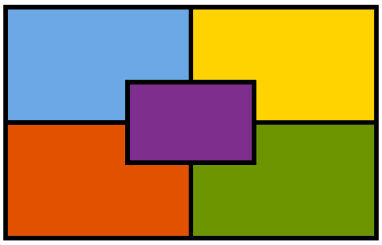


Figure 4.19.: An image is divided into five different regions

adding further points that are observed in images, not yet covered by points in the track, while only fulfilling the overlapping requirement. From the second point track on, the topmost point is taken that is observed in the region with minimal point count. The track is completed as described before. After a certain number of point tracks is constructed in this way, in the second part of the point selection, the point distribution is analyzed successively in every image. If one image only has points in three or less different regions, additional points are selected until at least four of the five regions are covered. This number is derived by inspection of Figure 4.19: Points in three different regions might by coincidence lie close together. If four different regions are covered at least one point must lie apart considerably.

Note that these additional points do not necessarily form a complete point track but eventually allow the construction of further constraints with new combinations of images. The overlapping requirement is guaranteed by the existing complete point tracks. In the end, points are selected so that in every image at least four different regions are covered by at least one point.

4.4.3. Outlier detection

In this section, a strategy is explained to detect outliers in the selected point tracks. Until now, point correspondences have only been filtered using the coplanarity constraint (cf. Section 4.2.1). Erroneous points that lie in epipolar planes of two overlapping images are still present. Moreover, because global rotations have been estimated (see Section 4.3), some erroneous points might not even lie in an epipolar plane anymore.

A spatial intersection of these points leads to wrong scales and, consequently, to inconsistent constraints. Therefore, a detection is necessary before the optimization is solved. In the following, two individual filters are presented that aim at different types of outliers. The first one uses the reprojection error of the pairwise spatial intersection in image space, i.e. the normalized difference between the observation and the reprojected point using the collinearity equations (2.1). For example, for the image pair $\{\mathsf{I}_i, \mathsf{I}_j\}$ it is required:

$$\Delta\bar{\mathbf{p}}_{ij}^{l} = \|\mathbf{p}_i^l - \bar{\mathbf{p}}_{ij}^{l,i}\| + \|\mathbf{p}_j^l - \bar{\mathbf{p}}_{ij}^{l,j}\| \leq \tau_r, \tag{4.44}$$

with $\bar{\mathbf{p}}_{ij}^{l,i}$ and $\bar{\mathbf{p}}_{ij}^{l,j}$ being the projection of $\mathbf{P}_{ij}^l$ in images i and j, respectively, and τ_r being another threshold. The reprojection error comes as a byproduct of the spatial intersection that has to be performed in order to establish the constraints in (4.41)-(4.43). Thus, all pairwise combinations for which Equation (4.44) is not fulfilled are rejected.

In general, erroneous points lying almost in one epipolar plane are not detected by a high reprojection error because both rays nearly intersect. The second filter aims at detecting these cases that often occur when the images contain repetitive structure. Given a triplet of images $\{\mathsf{I}_i, \mathsf{I}_j, \mathsf{I}_k\}$ with a point $\mathbf{P}^l$ observed in all three images, all homogeneous constraint functions (4.41)-(4.43) can be established. Now, the triplet is transformed to a local coordinate system by setting $\mathbf{t}_i = \mathbf{0}_{3\times 1}$. The locations of the two remaining images $\mathbf{t}_j$ and $\mathbf{t}_k$ are determined up to scale. For every constraint (4.41)-(4.43) the following tripletwise linear homogeneous equation system is established, using the matrices $\mathbf{C}_i^l, \mathbf{C}_j^l, \mathbf{C}_k^l$ as substitutions for the respective matrices in (4.41)-(4.43):

$$\begin{aligned} \begin{bmatrix} \mathbf{C}_i^l & \mathbf{C}_j^l & \mathbf{C}_k^l \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{t}_j \\ \mathbf{t}_k \end{bmatrix} &= \mathbf{0} \\ \Leftrightarrow \qquad \mathbf{C}_j^l \mathbf{t}_j &= -\mathbf{C}_k^l \mathbf{t}_k. \end{aligned} \tag{4.45}$$

Now, without loss of generality, the distance between I_i and I_j is assumed to be equal to one, so that a local coordinate system is defined. Then, (4.45) can be written as:

$$\begin{aligned} \mathbf{C}_j^l \mathbf{t}_{ij} &= -s^l_{\mathbf{t}_{ijk}} \mathbf{C}_k^l \mathbf{t}_{ik} \\ \Leftrightarrow \qquad s^l_{\mathbf{t}_{ijk}} &= -\frac{\mathbf{C}_j^l \mathbf{t}_{ij}}{\mathbf{C}_k^l \mathbf{t}_{ik}}, \end{aligned} \tag{4.46}$$

by introducing the scale $s^l_{\mathbf{t}_{ijk}}$, which describes the ratio between the two distances $\|\mathbf{t}_j - \mathbf{t}_i\|$ and $\|\mathbf{t}_k - \mathbf{t}_i\|$. In this way, a consistency constraint is established, which leads to the following theorem.

Theorem 2 (Triplet consistency constraint). *By assuming $\mathbf{t}_i = \mathbf{0}_{3\times 1}$, $\mathbf{R}_i = \mathbf{R}_i^{\star\star}$ and $\|\mathbf{t}_j - \mathbf{t}_i\| = 1$ to define a local coordinate system, the scale $s^l_{\mathbf{t}_{ijk}}$ describes the distance $\|\mathbf{t}_k - \mathbf{t}_i\|$ in this local system. If the observations of a point $\mathbf{P}^l$, i.e. $\{\mathbf{p}_i^l, \mathbf{p}_j^l, \mathbf{p}_k^l\}$, are homologous and it is $\mathbf{t}_{ij} = \frac{\mathbf{t}_j - \mathbf{t}_i}{\|\mathbf{t}_j - \mathbf{t}_i\|}$, $\mathbf{t}_{ik} = \frac{\mathbf{t}_k - \mathbf{t}_i}{\|\mathbf{t}_k - \mathbf{t}_i\|}$, $\mathbf{t}_{jk} = \frac{\mathbf{t}_k - \mathbf{t}_j}{\|\mathbf{t}_k - \mathbf{t}_j\|}$, then is must be:*

$$\frac{-\Big(\mathbf{C}_{ij}^l - \mathbf{C}_{ji}^l + \mathbf{C}_{jk}^l - \mathbf{C}_{kj}^l\Big)\mathbf{t}_{ij}}{\Big(-\mathbf{I}_{3\times 3} - \mathbf{C}_{jk}^l + \mathbf{C}_{kj}^l\Big)\mathbf{t}_{ik}} = \frac{-\Big(\mathbf{I}_{3\times 3} + \mathbf{C}_{ij}^l - \mathbf{C}_{ji}^l\Big)\mathbf{t}_{ij}}{\Big(-\mathbf{I}_{3\times 3} - \mathbf{C}_{ik}^l + \mathbf{C}_{ki}^l\Big)\mathbf{t}_{ik}} = \frac{-\Big(-\mathbf{I}_{3\times 3} + \mathbf{C}_{jk}^l - \mathbf{C}_{kj}^l\Big)\mathbf{t}_{ij}}{\Big(\mathbf{C}_{ik}^l - \mathbf{C}_{ki}^l - \mathbf{C}_{jk}^l + \mathbf{C}_{kj}^l\Big)\mathbf{t}_{ik}} = s^l_{\mathbf{t}_{ijk}}. \tag{4.47}$$

Proof. Equation (4.47) follows from (4.45)-(4.46) applied to all three constraints in(4.41)-(4.43), so it remains to show that these constraint functions are valid. If the observations and relative translations are correct then the estimated scales $\{s_{ij}^l, s_{ji}^l, s_{ik}^l, s_{ki}^l, s_{jk}^l, s_{kj}^l\}$ (see Equation (4.37)) and rotations $\{\mathbf{R}_{\gamma_{ij}^l}, \mathbf{R}_{\gamma_{ji}^l}, \mathbf{R}_{\gamma_{ik}^l}, \mathbf{R}_{\gamma_{ki}^l}, \mathbf{R}_{\gamma_{jk}^l}, \mathbf{R}_{\gamma_{kj}^l}\}$ (see Equation (4.38)) are exact and (4.47) follows from

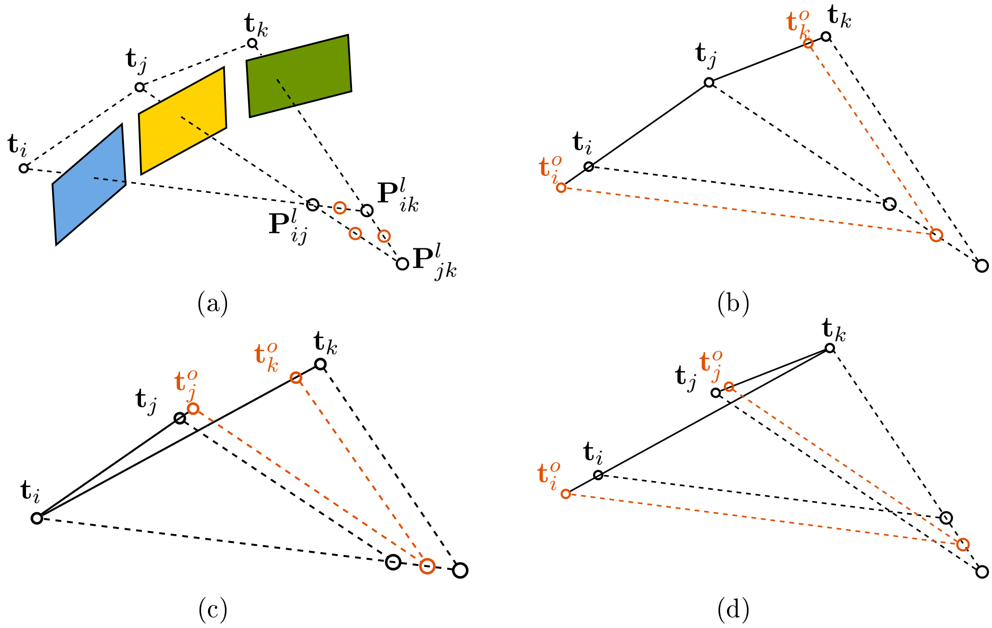


Figure 4.20.: Erroneous observation in one image: 4.20a: three different pairwise spatial intersections (black dots), dots in red show the respective mean position using (4.36). 4.20b-4.20d: small dots in black show the true locations of the three images which are connected according to the three possible combinations $\{\mathsf{l}_i\mathsf{l}_j\}$, $\{\mathsf{l}_j\mathsf{l}_k\}$ (4.20b), $\{\mathsf{l}_i\mathsf{l}_j\}$, $\{\mathsf{l}_i\mathsf{l}_k\}$ (4.20c) and $\{\mathsf{l}_i\mathsf{l}_k\}$, $\{\mathsf{l}_j\mathsf{l}_k\}$ (4.20d). Small dots in red with superscript $\cdot^o$ show the locations of the images estimated according to the spatial intersection. The ratio of the bases in all three figures is different.

(4.41)-(4.43) and (4.45)-(4.46). If one observation or relative translation is an outlier, there will be three different pairwise spatial intersections of $\mathbf{P}^l$ as depicted in Figure 4.20a. The constraint functions (4.41)-(4.43) are fulfilled only if at least one out of $\{\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_k\}$ is changed (Figures 4.20b-4.20d). Note that in these figures always two translations are changed according to Equation (4.34) (see the small red dots with superscript $\cdot^o$). This, however, changes the ratio between $\|\mathbf{t}_j - \mathbf{t}_i\|$ and $\|\mathbf{t}_k - \mathbf{t}_i\|$, which is a contradiction. $\square$

In general, having noisy observations, relative translations and estimated global rotations, Equation (4.47) is not fulfilled. Thus, a suitable threshold that allows the detection of contradicting triplets needs to be found. In order to do that, all three estimations of $s^l_{\mathbf{t}_{ijk}}$ are stacked into a vector $\mathbf{s}^l_{\mathbf{t}_{ijk}}$, which is normalized and of which the standard deviation $\sigma_{\mathbf{s}^l_{\mathbf{t}_{ijk}}}$ is computed. Then, after the scales have been computed for all triplet compbinations, the following condition must be true:

$$\sigma_{\mathbf{s}^l_{\mathbf{t}_{ijk}}} \leq \tau_s, \tag{4.48}$$

with $\sigma_{\mathbf{s}^l_{\mathbf{t}_{ijk}}}$ being the standard deviation of the three elements in $\mathbf{s}^l_{\mathbf{t}_{ijk}}$ and τ_s being another threshold computed from the vector over all normalized standard deviations $\mathbf{s}^l_{\mathbf{t}}$. For every triplet of images, the constraints (4.41)-(4.43) are established if, and only if, Equations (4.44) and (4.48) are fulfilled.

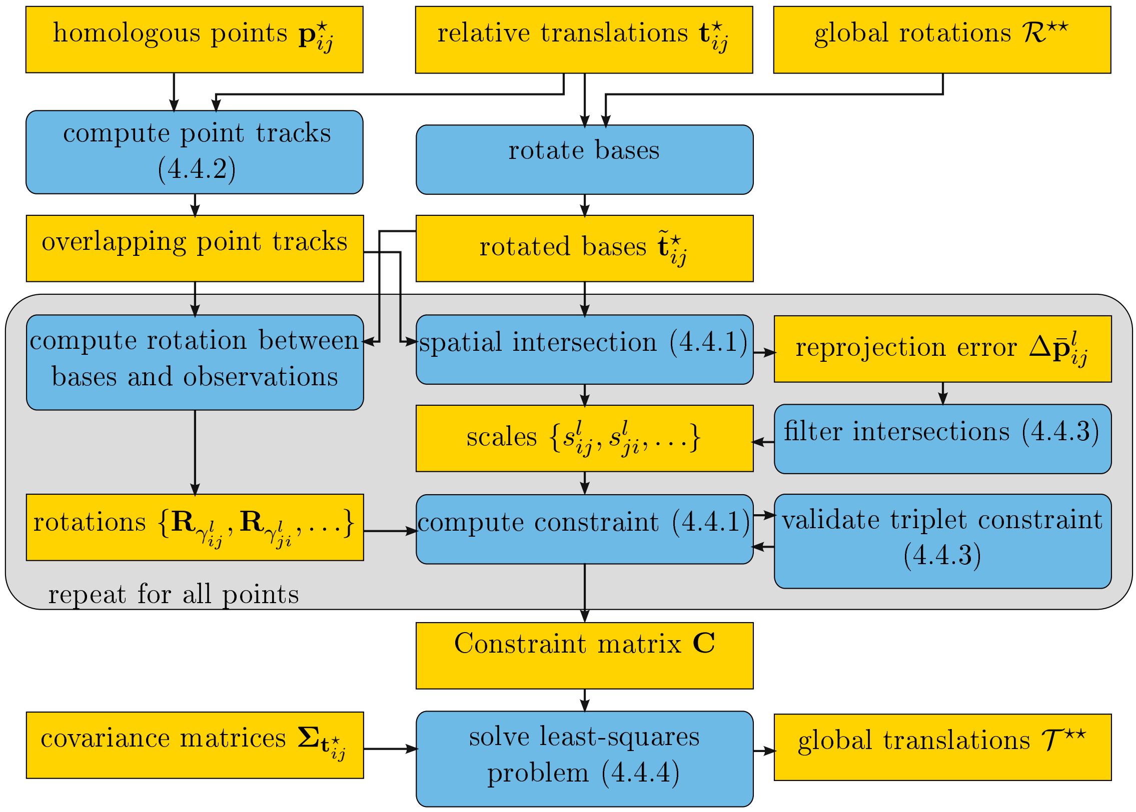


Figure 4.21.: Detailed workflow of global translation estimation. Blue boxes show the individual steps, yellow boxes the parametric outcome.

4.4.4. Solving the homogeneous linear equation system

In the previous sections, it was shown how linear constraints for individual global translations are established, how a suitable set of point tracks is found and how outliers in the image coordinates of the homologous points are detected. A detailed schematic overview of the workflow of translation estimation, summarizing these steps, is visualized in Figure 4.21. In this section, the estimation of global translations is examined.

Given a point $\mathbf{P}^l$ that is observed in $n_{\mathbf{P}^l}$ images, in total $\binom{n_{\mathbf{P}^l}}{3}$ different triplet combinations are possible. For each of these triplets, two linear independent constraint functions are established. Note that while all three linear constraints are computed in order to check triplet consistency (Equation (4.47)), just two of these are taken for the final estimation using a randomized selection. These are stacked into a single matrix $\mathbf{C}$ so that the following optimization problem is established:

$$\begin{aligned} \underset{\{\mathbf{t}_1,\ldots,\mathbf{t}_n\}}{\text{minimize}} \quad & \|\mathbf{w}_{\mathbf{t}}^T \mathbf{C}\mathbf{t}\| \\ \text{subject to} \quad & \|\mathbf{t}\| = 1 \ , \end{aligned} \tag{4.49}$$

with $\mathbf{t}$ being a $3n \times 1$ vector consisting of all unknown global translations. Vector $\mathbf{w}_{\mathbf{t}}$ contains weights that are computed for every triplet combination based on a combination of the covariance matrices of the respective relative translations. The weight for a triplet $\{\mathsf{I}_i, \mathsf{I}_j, \mathsf{I}_k\}$ is defined as:

$$w_{\mathbf{t},ijk} = \frac{1}{1 + \left(\lambda \ \mathbf{tr}\left(\bar{\mathbf{\Sigma}}_{\mathbf{t}^\star_{ij}}\right)_{\Delta(ijk)}\right)^{\nu}}, \tag{4.50}$$

with $\left(\bar{\mathbf{\Sigma}}_{\mathbf{t}^\star_{ij}}\right)_{\Delta(ijk)}$ being the mean covariance matrix for all three relative translations in the triangle $\Delta\left(\mathsf{I}_i\mathsf{I}_j\mathsf{I}_k\right)$. λ and ν are defined as in Equation (4.31) and the halfweight is set to

$$w_{0.5} = \max_{\forall(i,j)\in\mathcal{E}^+}\left(\mathbf{tr}\left(\mathbf{\Sigma}_{\mathbf{t}_{ij}}\right)\right). \tag{4.51}$$

The trivial solution of the objective function is $\mathbf{t} = \mathbf{0}$, which is avoided by the constraint $\|\mathbf{t}\| = 1$. The length of $\mathbf{t}$ is set to one because of simplicity (see below), any other nonzero value is also possible. The feasible solution manifold to (4.49) is the nullspace of $\mathbf{C}$. It is computed via a SVD of $\mathbf{C}$:

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T. \tag{4.52}$$

The orthonormal $3n \times 3n$ matrix $\mathbf{V}$ contains the orthogonal basis vectors of the rowspace of $\mathbf{C}$. A least-squares solution to the problem (4.49), provided a regular matrix $\mathbf{C}$, is given by the rightmost column of $\mathbf{V}$, i.e. $\mathbf{t}^\star = [0, 0 \ldots, 0, 1]^T \mathbf{V}$. Because $\mathbf{V}$ is orthogonal, the constraint $\|\mathbf{t}\| = 1$ is fulfilled. A derivation of this solution is given in Appendix F.

At the same time, the constraint $\|\mathbf{t}\| = 1$ fixes the scale of the estimated image orientation parameters. This, in combination with the already estimated global rotations, defines four of the seven datum parameters. Thus, a solution of the optimization problem (4.49) is only given up to a three-dimensional translation. Because of this datum defect, matrix $\mathbf{C}$ is rank deficient and only of rank $3n - 3$. There are two ways of approaching this issue. Firstly, one could eliminate three columns of $\mathbf{C}$, which solves the rank defect. The global translation parameters relative to the deleted columns are then set to 0. A second approach, which is utilized in this thesis, bears on not taking the rightmost column of $\mathbf{V}$ but the $3n - 3$rd column to take the rank defect into account. In this case, the barycenter of the estimated global translations lies at 0.

4.5. Bundle adjustment

The final step of the global image orientation workflow is a bundle adjustment. In the previous sections of Chapter 4, the focus was on the estimation of initial values for the exterior orientation parameters for this nonlinear and nonconvex optimization. It unifies both types of unknowns of image orientation, i.e. rotation and translation, as well as object coordinates of the observed homologous points. Initial values for the unknown object coordinates of the homologous points are computed using an average of multiple pairwise spatial intersections. The functional model is given by the well known collinearity equations (Equation (2.1)). The stochastic model is given by the covariance matrix of the observations, which are assumed to be uncorrelated and of equal accuracy. Bundle adjustment is considered to give a maximum likelihood solution of the image orientation problem but it depends on the quality of the initial values.

Because bundle adjustment is a standard photogrammetric procedure, only some important notes about the implementation are given.

Each point, as can be seen in Equation (2.1), leads to two individual observation equations, one for the x- one for the y-coordinate in image space. At the same time each observation equation includes a set of at least 9 unknown parameters, i.e. $\{\omega_i, \phi_i, \kappa_i, t_{i,x}, t_{i,y}, t_{i,z}, P_x^l, P_y^l, P_z^l\}$. As part of this thesis, the interior orientation parameters $\{h_x, h_y, c\}$ as well as any distortion coefficients are assumed to be constant and known if not stated otherwise. In general, the overall redundancy is high. However, it happens that some points or images do not have as many observations as others. In order to maintain a minimum amount of local redundancy and still keeping a high amount of images and points in the optimization, the minimum numbers of observations are set to eight for an image (four points) and six for an object point (three images).

In the first iteration all observations are weighted equally. Then, observations are reweighted in every iteration based on the residuals of the previous iteration $\Delta\hat{\mathbf{p}}_i^l = \left[\Delta\hat{p}_{i,x}^l, \Delta\hat{p}_{i,y}^l\right]^T$, similar to Equation (4.31) [Krarup et al., 1980]:

$$\begin{aligned} \mathbf{W}_i^l &= \begin{bmatrix} w_{i,x}^l & 0 \\ 0 & w_{i,y}^l \end{bmatrix}, \\ w_{i,x}^l &= \frac{1}{1 + \left(\lambda \Delta\hat{p}_{i,x}^l\right)^\nu}, \\ w_{i,y}^l &= \frac{1}{1 + \left(\lambda \Delta\hat{p}_{i,y}^l\right)^\nu}. \end{aligned} \tag{4.53}$$

Again, λ and ν are defined as in (4.31). The following halfweight is used:

$$w_{0.5} = 0.9 \max_{\forall i \in \mathsf{I}, \forall k \in \mathsf{P}} \Delta\hat{\mathbf{p}}. \tag{4.54}$$

After a certain number of iterations, all observations with residuals larger than $w_{0.5}$ are considered to be outliers and excluded from the estimation. If a point or an image remains with less than the minimum number of observations defined above, it is excluded as well. This is repeated for a certain number of times or until $\max_{\forall i \in \mathsf{I}, \forall k \in \mathsf{P}} \Delta\hat{\mathbf{p}}$ reaches a value that is in the trust region of the given problem. Note that by the definition of an outlier given above, the algorithm will always find at least one outlier. Thereafter, a final, unweighted estimation is conducted and iterated until convergence. As in previously presented least-squares optimization, a convergent solution is found when $\frac{(\mathbf{v}^T\mathbf{W}\mathbf{v})^{b-1} - (\mathbf{v}^T\mathbf{W}\mathbf{v})^b}{(\mathbf{v}^T\mathbf{W}\mathbf{v})^{b-1} + (\mathbf{v}^T\mathbf{W}\mathbf{v})^b}$ in iteration b is below a given threshold.

A bundle adjustment finds a (local) optimum only up to a seven parameter similarity transformation. This rank deficiency is related to the datum definitions made in the previous sections: One rotation is set as the identity matrix (Section 4.2.2), and translation and scale are defined as described in Section 4.4.4. There are several ways to deal with this rank deficiency such as introducing control points with known object coordinates or a free network adjustment. In this thesis, the datum is assumed to be known and fixed. Thus, seven parameters are excluded from the estimation and set according to the definitions made in Sections 4.2.2 and 4.4.4.

5. Experiments

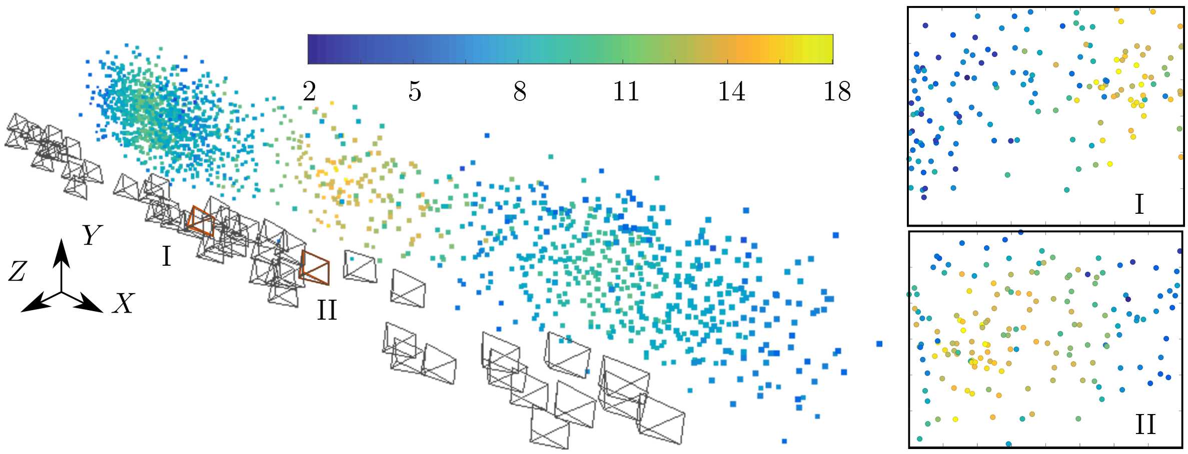
In this chapter, experimental results are presented in order to evaluate the global image orientation approach described in Chapter 4. There are three fundamental criteria, for which the proposed model is examined, corresponding to the objective defined in Section 1.1: accuracy, robustness and applicability to various kinds of image data. First, all the different data used for evaluation are presented in Section 5.1. These data are selected to assess the method in the three directions and can be divided into three categories, into which the following sections are arranged: Experiments on *synthetic data* are described in Section 5.2. These data are suitable to evaluate the accuracy (5.2.1) and robustness (5.2.2) of the method, because the ground truth is known and one has full control over the amount and distribution of noise and outliers. The second category is *benchmark data*, studied in Section 5.3, consisting of real images, for which ground truth information regarding their orientation parameters is given. Therefore, the accuracy of the approach is re-evaluated based on a realistic environment including a comparison to related state-of-the-art methods (5.3.1), which allows to position the method in the broad field of existing approaches. Moreover, a detailed analysis of the convergence of the Lie algebraic averaging will be pursued (5.3.2) as well as an investigation of the elimination of outliers in the homologous points (5.3.3). The last category encompasses *various types of data* and is examined in Section 5.4. Experiments on these data shall reveal the versatility of the approach. Examination includes images from image-hosting websites (5.4.1) and two different sets of self-acquired images from an unmanned aerial vehicle (UAV) (5.4.2).

5.1. Data and implementation

This section gives an overview over the data used for the evaluation of the method that is proposed in this thesis. This data is structured into three different groups: *synthetic data*, *benchmark data* and *various data*, which has different characteristics. On each of these different data, specific experiments are conducted in order to broadly evaluate the proposed approach.

Synthetic data

The synthetic data are based on a linear camera movement in one direction with all cameras pointing in a similar direction and a relatively high overlap as it occurs in UAV photogrammetry or mobile mapping image acquisition, for instance. It consists of 50 images and 2000 object points, distributed as demonstrated in Figure 5.1a. Object points are colored with respect to the number of images



(a) 3D distribution of points and images

(b) projected points in two examplary images

Figure 5.1.: 3D distribution of synthetic image data (5.1a) and point distribution in two exemplary images (5.1b), highlighted in red in 5.1a.

they are observed in, varying between two (blue) and 18 (yellow) observations. The distribution of observations in image space is exemplarily shown in Figure 5.1b. For a reasonable comparison of the results, a metric scale is introduced to the synthetic coordinate system. According to that, images are acquired in a range of approximately 50 m in X-direction and variations of $[-2, 2]$ m in Y- and $[-0.5, 0.5]$ m in Z-direction. Object points are in an average distance of 10 m to the closest camera. This configuration is based on a facade observation from terrestrial and/or UAV images. The interior orientation of the cameras is assumed fixed, representing a wide-angle configuration with $\mathbf{h} = (400, 300)$ px and $c = 600$ px. Thus, the overlap of the images is relatively high. Assuming no noise on the homologous points, 482 relative orientations can be computed which leads to a graph density of approximately 0.39, i.e. 39% of the maximum number of possible edges exist. Higher noise leads to less relative orientations and consequently a lower graph density.

These data are used to allow an exhaustive analysis of the accuracy of the approach. Various noise levels on the observations are realized in order to examine the propagation of uncertainties between the individual optimizations. Moreover, the robustness of the breadth-propagation is evaluated. The control of the rate of outliers in the relative orientations allows to define a breakdown point of the algorithm. Finally, the outlier detection during translation estimation is investigated.

Benchmark data

The benchmark data consist of four different image sets, which were originally published in Strecha et al. [2008]: *fountain-P11*, *Herz-Jesu-P25*, *castle-P19* and *castle-P30*. These image sets comprise between 11 and 30 images taken in an arc-like (*fountain-P11* and *Herz-Jesu-P25*) or circular arrangement (*castle-P19* and *castle-P30*), corrected for radial distortion. Exemplary images of these sequences are shown in Figure 5.2. The two castle-sequences picture the same scene, but the se-

(a) *fountain-P11* (b) *Herz-Jesu-P25* (c) *castle-P19 / castle-P30*

Figure 5.2.: Examplary images of the benchmark datasets [Strecha et al., 2008].

lection and the number of images are different. Images were taken with a *Canon EOS D60* with APS-C sensor (22.5 mm $\times$15.0 mm) at a resolution of 3072 px $\times$2048 px and constant interior orientation parameters. This results in a pixel size of approximately 7.32 μm. For exterior and interior orientation parameters, a ground truth based on laser scanning and a camera calibration exist (cf. Strecha et al. [2008] for more details). This ground truth and calibration information is given in form of projection matrices for every image. Additionally, for every image the Exif-information (Exif $\sim$ exchangeable image file format) is provided, in which a focal length of $c = 20$ mm is stored. Homologous points have been computed using *Micmac* [Pierrot-Deseilligny & Paparoditis, 2006], an open source tool for image orientation and dense matching[1].

In order to compare the estimated global orientation parameters to the ground truth, both coordinate systems have to be aligned. The ground truth coordinate system is moved to a coordinate system centered in the starting image (as defined in Section 4.2.2) and rotated so that $\mathbf{t}^1_{GT} = \mathbf{0}_{3\times1}$ and $\mathbf{R}^1_{GT} = \mathbf{I}_{3\times3}$. The scale adaption is applied to the estimated orientations instead so that the distance of the two farthest images is equal. This allows for a metric comparison to the ground truth.

The benchmark data allow to transfer the findings of the synthetic data to a more realistic environment. The accuracy of the approach is evaluated and compared to various state-of-the-art methods. Moreover, a realistic distribution of outliers in the homologous points allows for a more convincing examination of the outlier detection.

Various data

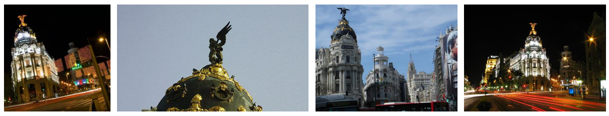
In the last section of this chapter, experiments on various large datasets are presented. These shall underline the variability of the proposed approach. First, publicly available data from image-hosting websites are processed, which show sights and places in Europe and North America (see exemplary

[1] `http://logiciels.ign.fr/?Micmac`

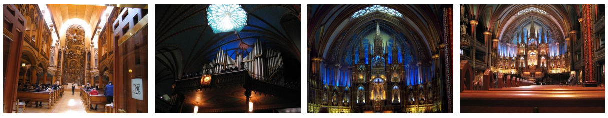
(a) *Alamo*

(b) *Ellis Island*

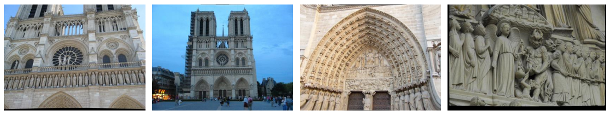
(c) *Gendarmenmarkt*



(d) *Madrid Metropolis*



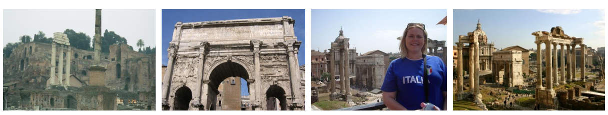
(e) *Montreal Notre Dame*



(f) *Notre Dame de Paris*

(g) *NYC Library*

(h) *Piazza del Popolo*



(i) *Roman Forum*

(j) *Tower of London*

(k) *Vienna Cathedral*

(l) *Yorkminster*

Figure 5.3.: Examplary images of the datasets from image-hosting websites [Wilson & Snavely, 2014].

images in Figure 5.3. Results for these data are published in Wilson & Snavely [2014][2] together with the orientation results of *Bundler*, including the observed homologous points. The *Notre Dame de Paris* dataset was published in the course of the photo-tourism project [Snavely et al., 2006][3]. The number of images in these datasets, which are useful for reconstruction, varies between 230 and 1062 (see also Table 5.6 in Section 5.4.1). These images were not taken with the purpose of photogrammetric reconstruction, which has two major consequences: First, the actual object of interest is not necessarily in focus or even pictured at all, because often persons are the central part of the images. Second, the acquired images have a variety of different camera models with different interior orientation parameters. Approximate values for these parameters are extracted from the Exif-information and are provided with the data. Note that these image sets also include processed images like panoramic images, stitched from several individual images. Using the provided homologous points, a preselection of images has already taken place.

Secondly, self-acquired images from two different UAV-flights are evaluated. These systems be-

[2] `http://www.cs.cornell.edu/projects/1dsfm/`
[3] `http://phototour.cs.washington.edu/datasets/`.

(a) *Leibniztempel* (b) *Wettbergen*

Figure 5.4.: Examplary images of the two UAV flights.

come more and more useful for photogrammetric reconstruction and mapping tasks. To show the variability of the approach, two different kinds of flights are chosen, one around the *Leibniztempel* in Hanover, Germany, in which the images are taken from different perspectives all around the building, the other showing a housing development area in *Wettbergen* in the south of Hanover. These images are taken in a grid, all pointing downwards like in a setting of airborne photogrammetry. Images of the flight around the Leibniztempel are captured with a *Canon Digital IXUS 100 IS*. The whole sequence consists of 184 images with a resolution of 1600×1200 px (pixel size approx. $3.88\,\mu$m) and an initial focal length of 6 mm (from Exif-information). During the flight above the housing development area, 287 images are taken with a *Canon PowerShot S110*, at a resolution of 720×540 px (pixel size approx. 10.28μm) and an initial focal length of 5 mm (from Exif-information). Exemplary images of both flights are shown in Figure 5.4a and 5.4b.

The purpose of these data is to show that the approach is also applicable to images, which were not taken in a controlled environment. Images from the Internet pose two major challenges to the algorithm: many different interior orientations that are only approximately known including highly distorted images and a comparably large number of images. The high number of different interior orientations hinders a self calibration during bundle adjustment, leading to a comparably low geometric accuracy. Images taken from a UAV become more and more important in the scope of reconstruction and mapping applications. The two selected datasets are different regarding the configuration of images, which shall underline the versatility of the proposed approach.

Selection of thresholds and implementation

In the methodology chapter, a number of thresholds and reference values were presented. Each of those has a different task, comprising the control of the number of correspondences, necessary for the computation of a relative orientation ($\tau_{|\mathbf{p}|}$), the convergence of the M-estimation ($\tau_{\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}$),

	4.2	4.2.1	4.2.2		4.3.2	4.4.2	4.4.3	
	$\tau_{\lvert\mathbf{p}\rvert}$	$\tau_{\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}$	τ_α	τ_c	$\tau_{\Delta\mathbf{r}}$	$\sharp pointtracks$	τ_r	τ_s
5.2.1	40	0.01	5°	1.5	0.001	10	10 px	$\mathbf{3med}\left(\mathbf{s}_\mathbf{t}^l\right)$
5.2.2	40	0.01	5°	1.5	0.001	10	2 px	$\mathbf{3med}\left(\mathbf{s}_\mathbf{t}^l\right)$
5.3.1	40	0.01	5°	1.5	0.001	30	2 px	$\mathbf{3med}\left(\mathbf{s}_\mathbf{t}^l\right)$
5.3.2	-	-	5°	1.5	0.001	-	-	-
5.3.3	-	-	-	-	-	30	2 px	$\mathbf{3med}\left(\mathbf{s}_\mathbf{t}^l\right)$
5.4.1	20	0.01	5°	1.5	0.001	20	10 px	$\mathbf{3med}\left(\mathbf{s}_\mathbf{t}^l\right)$
5.4.2	20	0.01	5°	1.5	0.001	20	2 px	$\mathbf{3med}\left(\mathbf{s}_\mathbf{t}^l\right)$

Table 5.1.: Selection of thresholds for different experiments explained in the following sections.

the inlier/outlier classification of the breadth-propagation (τ_α), the ratio between consistent and inconsistent rotation estimates for outlier classification (τ_c), the convergence of the Lie-algebraic averaging ($\tau_{\Delta\mathbf{r}}$), the number of point tracks ($\sharp pointtracks$) and the reprojection error (τ_r) and scale consistency (τ_s) during the estimation of global translations. Table 5.1 gives an overview over these seven thresholds, where they are defined and which values have been selected for the individual experiments that are investigated in the following sections.

The method proposed in this thesis is based on self-implemented *Matlab* code. For the computation of the essential matrix, the toolbox of Stewenius et al. [2006][4] is used. The random sampling framework for the essential matrix estimation is based on the structure and motion toolkit of Philip Torr[5]. The SDP for the rotation estimation is implemented using *YALMIP* [Löfberg, 2004][6], a toolbox to parse linear matrix inequality constraints to various convex optimization solvers. As a solver SDPT3 [Toh et al., 1999], an open SDP-solver for Matlab, is used.

5.2. Synthetic data

In this section, experiments conducted on synthetic data are presented. Because the noise of the homologous points and the number and magnitude of gross errors can be controlled, these data are suitable for the evaluation of the accuracy (5.2.1) and robustness (5.2.2) of the proposed method. All results regarding the accuracy are compared to the ground truth information, which is described in Section 5.1.

5.2.1. Accuracy analysis for varying noise levels

For the evaluation of the accuracy of the proposed global image orientation model, observations are assumed to be affected by different levels of noise. In the given synthetic dataset, object points are projected into the images and modified by normal distributed noise with $\sigma = t \cdot 0.5\,\mathrm{px}$, $t =$

[4] available at `http://vis.uky.edu/~stewe/FIVEPOINT/`.
[5] `http://de.mathworks.com/matlabcentral/fileexchange/4576-structure-and-motion-toolkit-in-matlab`.
[6] available at `http://users.isy.liu.se/johanl/yalmip/`

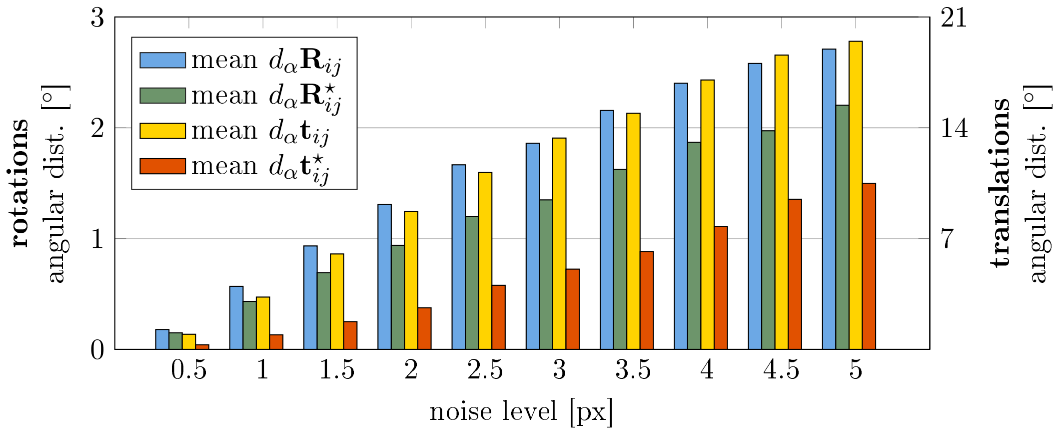


Figure 5.5.: Accuracy of relative orientations before (blue and yellow bars) and after M-estimation (green and red bars) considering different noise levels. Blue and green bars correspond to the left scale, yellow and red bars to the right scale.

$[1, 2, \ldots, 10]$. Then, for each noise level, the standard workflow (see Figure 4.1) is conducted. The following results are averaged from five individual computations.

Effect on relative orientations

The uncertainty of the homologous points is propagated to the relative orientations. This is depicted in Figure 5.5, which shows the average accuracy of the relative rotations and translations for each noise level. It is further distinguished between relative orientations before (blue and yellow bars) and after M-estimation (green and red bars). Note that the accuracy of the relative translations is also given in the angular distance because their Euclidean length is not part of the parameterization. The angular distance describes the angle between the rotated relative translation direction and the vector between the two respective images in the ground truth. The improvement through the M-estimation is significant for both parameter types, but especially distinctive for the relative translations, for which the accuracy improves up to 70%. The mean accuracy decreases linearly with growing noise and stays below 2.5° angular distance in standard deviation for the relative rotations and up to 10° in standard deviation for the relative translations. This gives evidence that, in general, the estimation of relative rotations is more accurate than the estimation of relative translations.

With increasing noise level, the number of estimated relative orientations and their average number of correspondences also reduce by around 40% and almost 75%, respectively (see Figure 5.6). Besides the increasing noise of the observations, the decreasing number of correspondences contributes to the decreasing quality of the relative orientations.

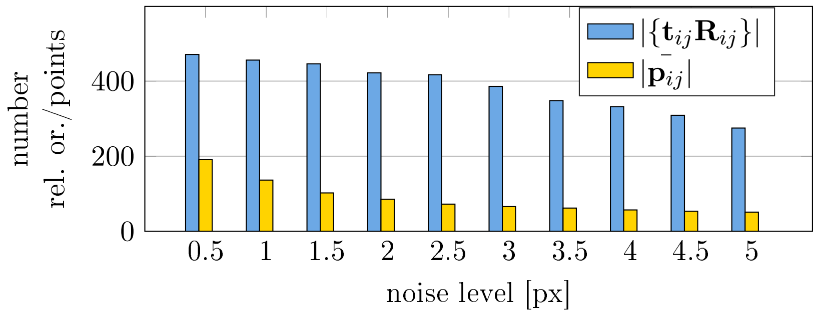


Figure 5.6.: Number of estimated relative orientations and the average number of correspondences per orientation.

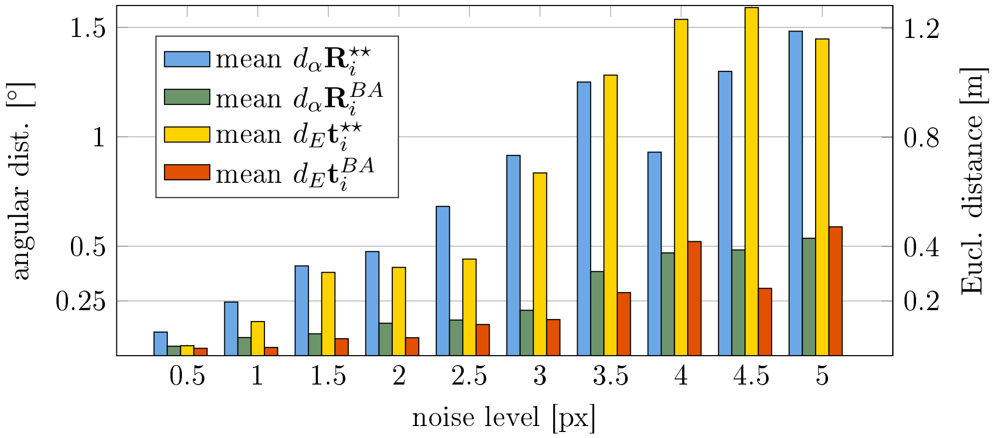
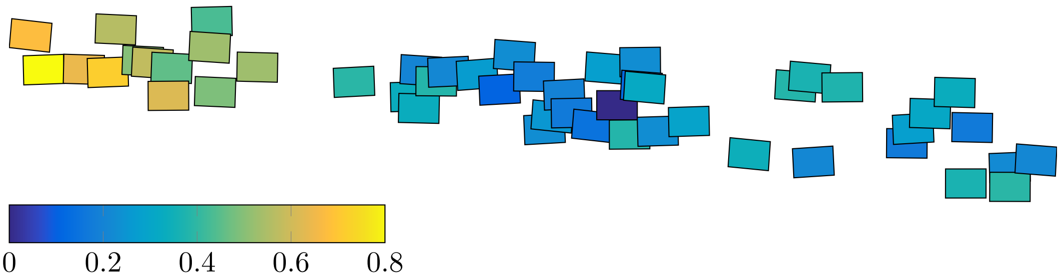


Figure 5.7.: Accuracy of global orientations with respect to ground truth for different noise levels before and after bundle adjustment. Blue and green bars correspond to the left scale, yellow and red bars to the right.

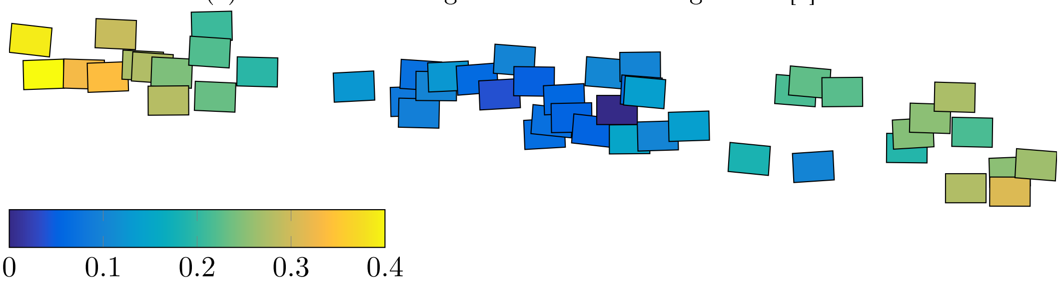
Effect on global orientations

The impact of different noise levels on the quality of the estimated global orientation parameters with respect to the ground truth is depicted in Figure 5.7. As expected, increasing noise negatively affects the accuracy of the global orientation. Rotations before bundle adjustment (blue bars) can be estimated up to 1° accuracy in standard deviation for a noise up to 4 px (except for the value at 3.5 px). The accuracy of the translations (yellow bars) is below 10 cm for 0.5 px noise and increases to approximately 1.2 m in standard deviation for high noise levels. For both types of parameters, rotations and translations, a final bundle adjustment has a significant positive effect on the accuracy (green and red bars). In order to show the spatial distribution of the accuracy, Figures 5.8a and 5.8b demonstrate the angular and Euclidean distances of every single image, respectively. The distances are given for a noise level of 1 px. One can see that the distance to the ground truth depends on the selection of the starting image, i.e. the datum definition (the image in dark blue in the center). To the outer regions of the scene the distances increase.

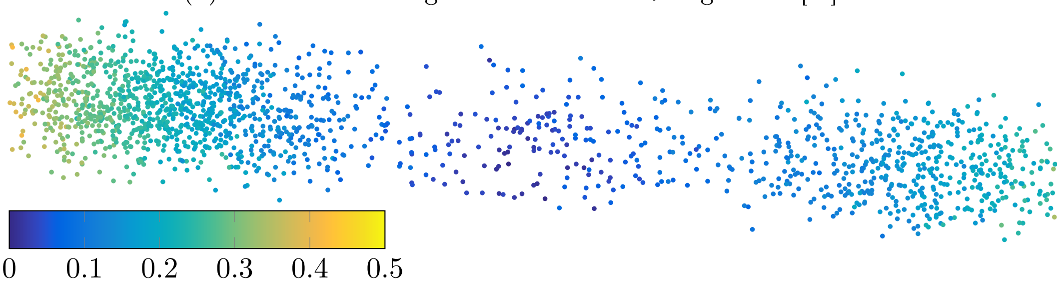
It remains to be shown how the uncertainty propagates to the structure. In Figure 5.8c, the



(a) Deviation of the global rotations $\mathcal{R}^{\star\star}$ given in [°]



(b) Deviation of the global translations $\mathcal{T}^{\star\star}$ given in [m]



(c) Deviation of the object coordinates given in [m]

Figure 5.8.: Accuracy of the global rotations (5.8a) and translations (5.8b) shown for each image and the accuracy of the object coordinates (5.8c). All accuracies are measured as angular and 3D Euclidean distance to ground truth, respectively.

accuracy of the estimated object points with respect to the ground truth is shown. These points are computed via multi-image pairwise spatial intersection using the estimated global image orientation parameters $\mathcal{R}^{\star\star}$ and $\mathcal{T}^{\star\star}$. The average point accuracy is approximately 0.2 m in standard deviation. It reflects the findings for the accuracy of the orientation parameters, i.e. an increasing difference to the ground truth the higher the distance to the starting image. After bundle adjustment, the average accuracy increases by a factor of 5 to approximately 0.04 m in standard deviation.

Dependency between global rotations and translations

The comparison of the accuracy of the relative and the global rotations leads to the conclusion that the quality of the rotations after different stages of the proposed method varies. This is shown in the following based on the three sets of global rotations, $\mathcal{R}$, $\mathcal{R}^{\star}$ and $\mathcal{R}^{\star\star}$. Because global translations depend on the quality of the estimated global rotations, the following experiment shall expose the

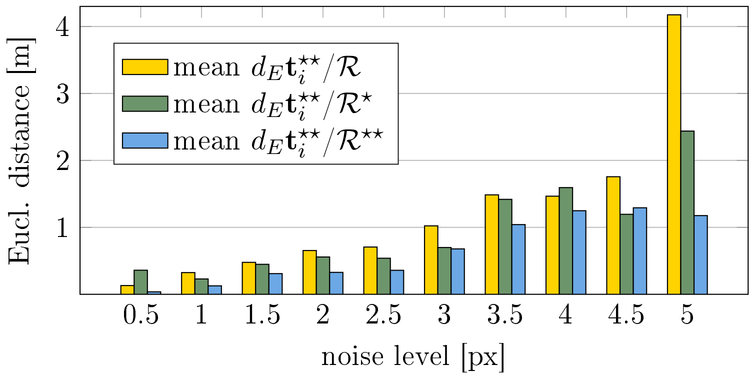


Figure 5.9.: Accuracy of global translations using different global rotations as initialization.

influence rotations have on the accuracy of the translations. Global translations are estimated based on global rotations after breadth-propagation $\mathcal{R}$, SDP estimation $\mathcal{R}^\star$ and after SDP estimation and Lie algebraic averaging $\mathcal{R}^{\star\star}$. The differences to ground truth are visualized in Figure 5.9. Note that the translation accuracy using $\mathcal{R}^{\star\star}$ (blue bars) is also depicted in Figure 5.7 and is shown again for an enhanced comparison. As expected, global translations are estimated most accurately using $\mathcal{R}^{\star\star}$. Except for the case of 5 px noise, the differences, however, are small. An important result of this experiment is that for a low noise level, i.e. 0.5 px, rotations after SDP (green bars) lead to inferior translations than using rotations after breadth-propagation (yellow bars). This was the case in all five individual computations. One reason for this might be a comparably accurate propagation $\mathcal{R}$, because of the small size and high density of the view-graph given such a low noise level. Another reason might be that the semidefinite constraints in (4.21) are too loose. In order to have further evidence for these propositions, an experiment was conducted with zero noise. Translations, estimated using $\mathcal{R}$ and $\mathcal{R}^{\star\star}$, are exact, whereas translations using $\mathcal{R}^\star$ have an average uncertainty of approximately one centimeter, which supports the given reasons. For higher noise levels, translations are often estimated more accurately using rotations after SDP-estimation. This leads to the consequence that the larger the image set, the less dense the view-graph and the lower the expected accuracy of the observed image coordinates the larger might be the benefit of an intermediate SDP-estimation. The finding is taken up again in Sections 5.3.1 and 5.3.2.

In Section 4.3.1 it was mentioned that one can think of solving the problem with a reduced set of semidefinite constraints in order to reduce the computation time and the applicability to larger datasets. In order to show how this affects the outcome, rotations have been estimated in an SDP, applying the convex hull constraint only to those submatrices in $\mathbf{M}$, for which also the respective relative rotation in $\mathbf{M}^0$ is available. In this experiment, the number of constraints is reduced from 1226 to 471. The estimated rotations are equivalent to the ones computed using the full set of constraints, whereas the computation time reduces by approximately 12%. Thus, the reduced problem can be used in practice for the estimation of rotations, although the computation time reduces only slightly.

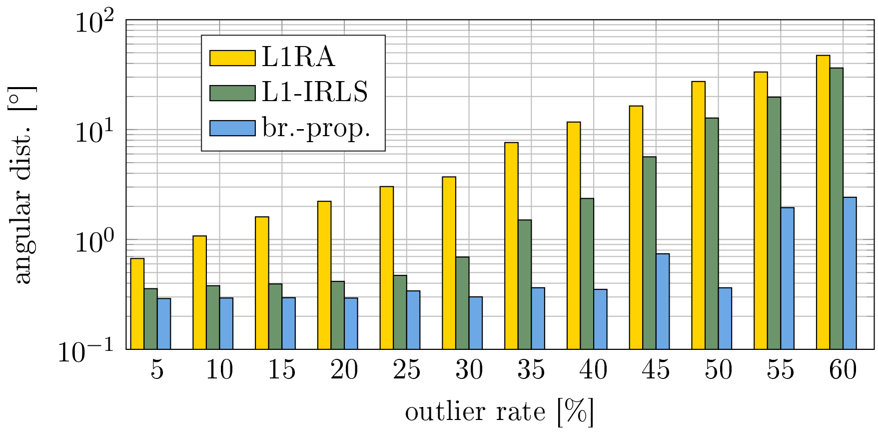


Figure 5.10.: Robustness of the breadth-propagation compared to two state-of-the-art approaches, [Chatterjee & Govindu, 2013], shown as angular distance of global rotations $\mathbf{R}_i^{\star\star}$ with respect to ground truth.

5.2.2. Robustness analysis for varying outlier rates

In this section, the influence of outliers in the image coordinates of the homologous points as well as the relative orientations on the estimated global orientation parameters is analyzed. The homologous points are assumed to be affected by a normally distributed noise with a standard deviation of 0.5 px, leading to a set of noisy relative orientations.

Outliers in the relative rotations

In this experiment, outliers in the relative rotations are generated in order to check the detection ability of the breadth-propagation. An outlier is generated by a multiplication with a random rotation matrix. The random rotation is required to have Euler angles in the interval $\{\omega, \phi, \kappa\} = [15°, 345°]$ in order to have a significant effect on the estimated relative rotations. These outliers, simulating modified relative rotations, are selected randomly in a equally distributed fashion based on a specific outlier rate. Outlier rates are defined lying between 5% and 60%. For every such outlier rate, 100 individual repetitions are conducted, in which the random selection is repeated, so that the average result is independent of the configuration of inliers and outliers.

For a qualitative assessment, the robustness of the proposed method is compared to a state-of-the-art approach. A model often used in recent approaches for the robust estimation of global rotations (e.g. [Wilson & Snavely, 2014; Ozyesil & Singer, 2015; Cui et al., 2015]) is the efficient L1RA (L_1 rotation averaging) and L1-IRLS (L_1 iterative reweighted least squares) approach of Chatterjee & Govindu [2013]. The source code of this model is made available by the authors. As initialization, all three models, breadth-propagation, L1RA and L1-IRLS, have the same relative rotations. In order to qualify the detection results, the angular distance to the ground truth global rotations is computed. Note that after breadth-propagation the SDP and Lie algebraic averaging are performed.

Results are depicted in Figure 5.10; note the logarithmic scale of the y-axis. Up to 40% of outliers, breadth-propagation (blue bars) finds all outliers, which leads to accurate global rotations $\mathbf{R}_i^{\star\star}$, comparable to the results for the respective noise level shown in Figure 5.7. At 45% outliers, in one out of 100 trials, not all outliers have been removed, which affects the average angular distance considerably so that the average angular distance increases by a few degree. L1RA and L1-IRLS lead to inferior results, also at lower outlier rates, which is due to the subsequent rotation estimation as opposed to the outlier detection. Considering L1-IRLS, the accuracy is not affected by outliers up to 25% outlier rate, which gives evidence that all outliers were detected for these rates. From 30% on, the estimation quality quickly decreases, which indicates that the breakdown point of L1-IRLS lies at around 30% outliers. For a mathematical definition of the breakdown point the reader is referred to Hampel et al. [2011]. At high outlier rates large errors are produced using L1RA and L1-IRLS whereas the average angular distance using the proposed method is in the range of a few degree even at 60% outlier rate. Thus, the constraints implemented in the breadth-propagation have the ability to filter defective relative rotations until a high breakdown point of approximately 50% outliers.

For the sake of robustness, this experiment has been conducted additionally with random rotations having Euler angles in the interval $\{\omega, \phi, \kappa\} = [5°, 355°]$, i.e. outliers may correspond only to a small random rotation by 5°. A graphical presentation of these results is omitted, because they are almost equivalent to the ones shown in Figure 5.10.

The experiment described in the previous paragraph only applies to equally distributed outliers. In practice, it may happen that two or more outliers lead to a consistent solution. As is mentioned in Section 5.4.1, breadth-propagation is not necessarily able to cope with these circumstances. An example for the consequences will be shown in Section 5.4.1.

Outliers in the homologous points

In a second experiment, outliers are generated in the image coordinates of the homologous points. The experiment is conducted twice by modifying the image coordinates with a random value so that its distance to the original coordinate is at least 50 px and 10 px using the Euclidean norm. The rate of outliers is varied between 2% and 20% and, for each outlier rate, five individual computations were conducted, of which the average value is taken. It is examined how outliers influence the estimation of relative orientations and global translations and how well the detection as presented in Section 4.4.3 performs. Note that the estimation of global rotations is not investigated at this point, because it is not directly affected by outliers in the image coordinates of homologous points. The thresholds for the reprojection error τ_r and the triplet consistency τ_s are set to 2 px and $\mathbf{med}\left(\mathbf{s}_\mathbf{t}^l\right)$, respectively (see also Table 5.1), which means that, if the tripletwise standard deviation is higher than three times the median of the vector $\mathbf{s}_\mathbf{t}^l$, the constraint is rejected. In Figure 5.11, the average detection rate of outliers during the estimation of global translations is depicted (blue and green bars). The detection rate appears to be independent of the outlier rate and varies between 99% and

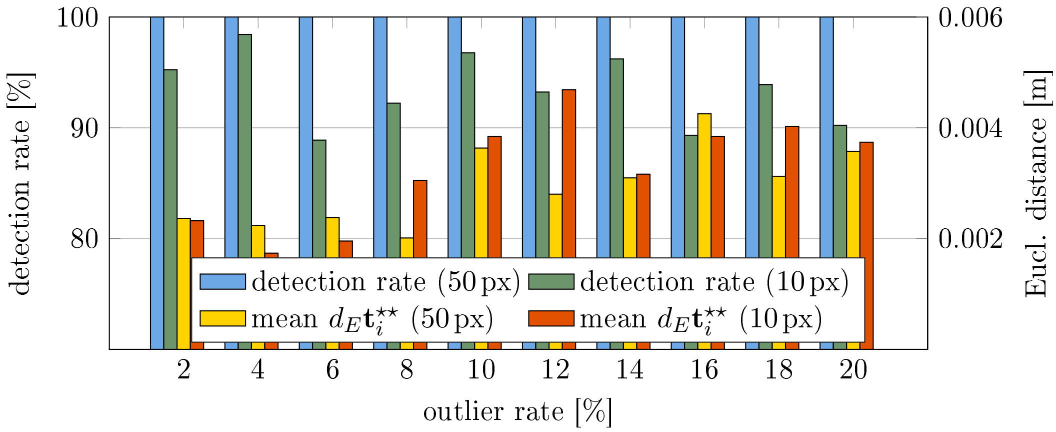


Figure 5.11.: Detection of outliers in the homologous points (blue and green bars, left scale) and effect on the accuracy of global translations (yellow and red bars, right scale).

100% for large outliers (larger than 50 px) and 89% and 98% for small outliers (larger than 10 px). Additionally, the accuracy of the estimated global translations with respect to the ground truth is shown (yellow and red bars). This accuracy does not depend on the size of the outliers, implying that although less outliers are detected in case of small outliers, the remaining ones do not have a significant effect on the result. The detection rate would increase if stricter values for the two thresholds τ_r and τ_s were selected. However, this in general involves a higher false positive rate (inliers that are classified as outliers), which is why the selection of optimal thresholds is a delicate task and remains for future work.

One can see that there is a slight trend showing an increasing Euclidean distance with increasing outlier rate for both types of outliers. A reason for this is that, with increasing number of outliers in the homologous points, less constraints for the global translations are computed, because the respective observations are eliminated. This trend is also ascribed to the declining quality of the relative orientations, which can be seen in Figure 5.12. Although outliers are eliminated within RANSAC and subsequent M-estimation, less correspondences remain to estimate the relative orientations, which has a negative effect on their accuracy. Moreover, the relative orientations with outliers of at least 10 px are slightly less accurate, which indicates that not all outliers have been eliminated by RANSAC and M-estimation. Again, this would require a more careful parameter selection regarding RANSAC and $\tau_{\mathbf{c}_{\{\mathbf{t}_{ij},\mathbf{R}_{ij}\}}}$.

The random design of the outlier generation only by coincidence leads to a configuration where the outlier lies in the same epipolar plane as the original observation, considering neighboring images. Thus, almost all outliers have been detected via the reprojection constraint (4.44). An analysis, also for outliers fulfilling the epipolar constraint, will be addressed in Section 5.3.3.

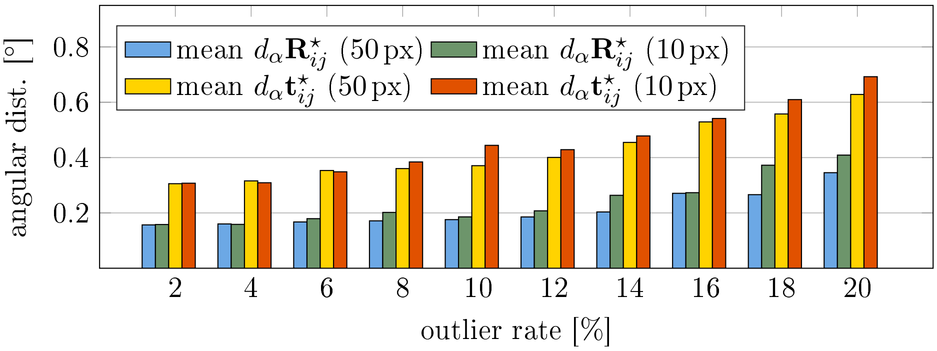


Figure 5.12.: Effect of the outlier rate on the accuracy of the relative orientations.

5.3. Benchmark data

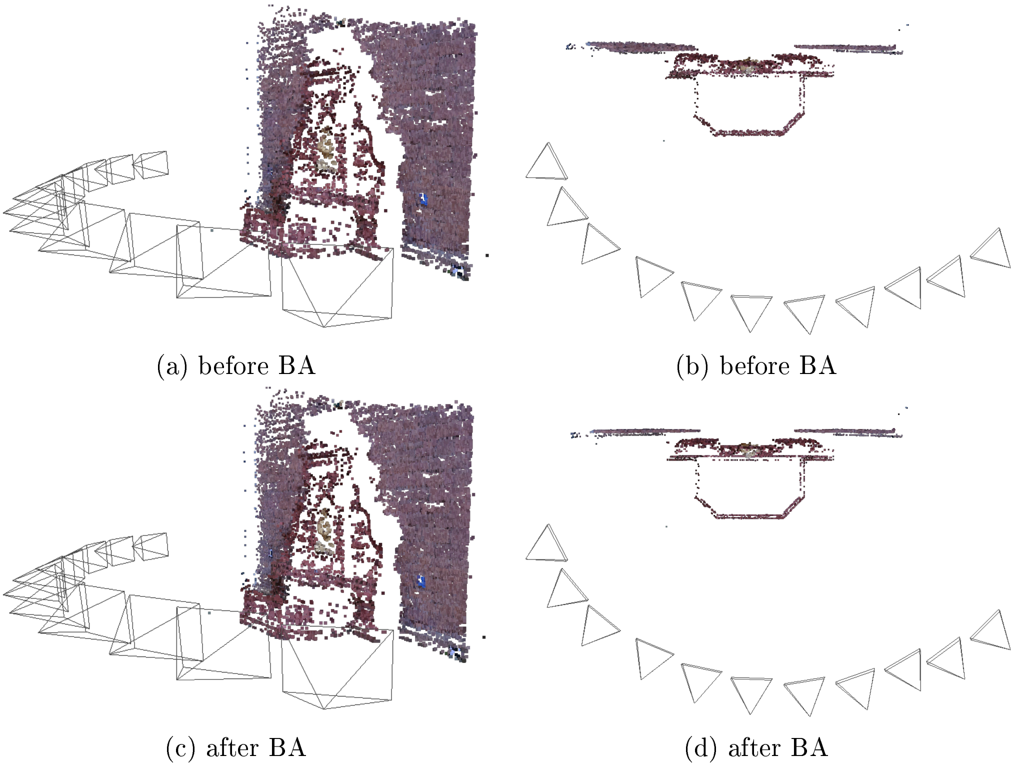
The previous section was dedicated to the analysis of synthetic data. The general behavior of the proposed approach given different noise levels and outlier rates was studied. In this section, experiments using benchmark data are presented. For these data many qualitative analyses were conducted in related work. In order to conduct a comprehensive comparison, two different evaluations are performed. The first one comprises approximate interior orientation from Exif-information. Thus, the focal length is extracted from the Exif-information and the principal point is assumed to lie in the center of the image. In the second evaluation, calibrated values are used for the interior orientation. This differentiation, on the one hand, is reasonable because both configurations are studied in related work. On the other hand, it reveals information about the influence of different interior orientation parameters on the estimated exterior orientations. Additionally, results for the global orientations are shown after a self calibration using the interior orientation from Exif-information as initialization.

In Section 5.3.1 a comparative analysis of the accuracy is presented using various related publications. This is followed by a convergence study of the Lie algebraic averaging in Section 5.3.2. Finally, the outlier elimination during translation estimation is investigated again in Section 5.3.3, giving new insights about outliers inside the epipolar plane and sharpness of discrimination.

5.3.1. Comparative accuracy analysis

Figures 5.13, 5.14 and 5.15 show the reconstruction results of *fountain-P11*, *Herz-Jesu-P25* and *castle-P30*. The reconstruction of *castle-P19* is neglected here because it is the same scene as for *castle-P30*. The top rows show the reconstructions before, the bottom rows the results after the final bundle adjustment[7]. In the right columns, in order to better show the planarity of the reconstructed scenes, the results are shown in an orthographic projection. All results are obtained using approximate interior orientation from Exif-information. In general, it is hard to see any differences. By comparing 5.13b with 5.13d, 5.14b with 5.14d and 5.15b with 5.15d, one can see

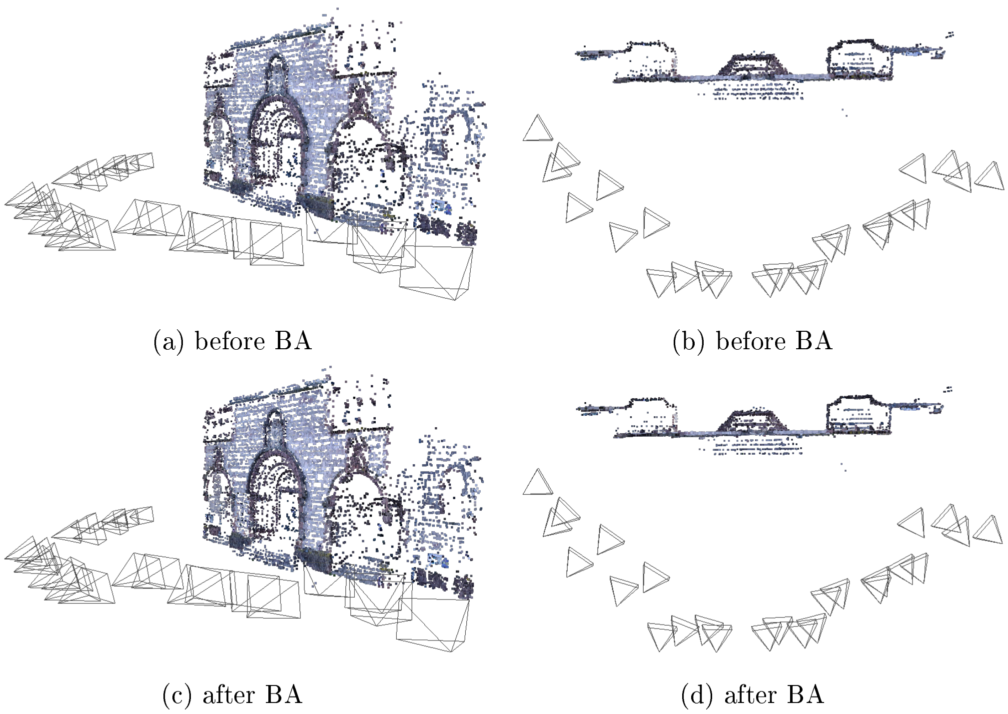
[7]The full model (BP-SDP-Lie and TE-Rob) has been used, see next paragraph for more details



(a) before BA (b) before BA

(c) after BA (d) after BA

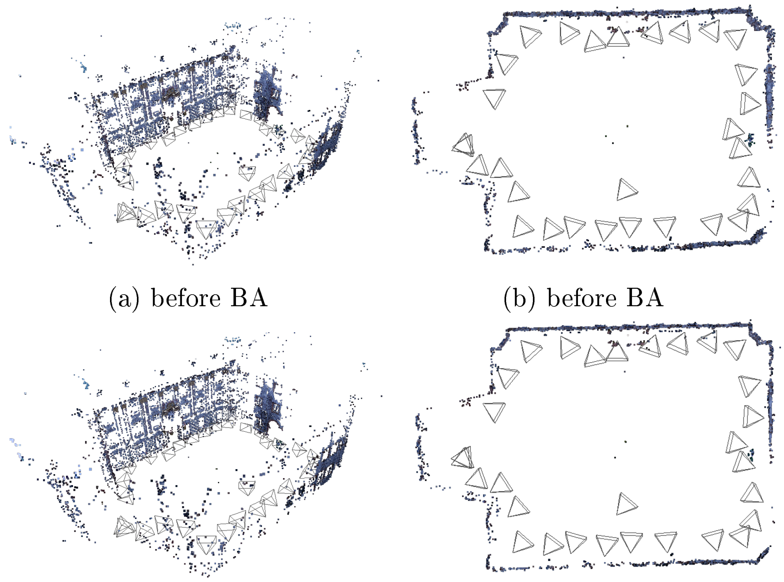
Figure 5.13.: Reconstruction of the *fountain-P11* dataset (see Figure 5.2a), before and after bundle adjustment (BA).



(a) before BA (b) before BA

(c) after BA (d) after BA

Figure 5.14.: Reconstruction of the *Herz-Jesu-P25* dataset (see Figure 5.2b), before and after bundle adjustment (BA).



(a) before BA (b) before BA

(c) after BA (d) after BA

Figure 5.15.: Reconstruction of the *castle-P30* dataset (see Figure 5.2c), before and after bundle adjustment (BA).

that, before bundle adjustment, the point clouds are less planar. Regarding the location and rotation of the images, no differences before and after bundle adjustment are visible.

Comparison of different model instances and to the state-of-the-art

In a first analysis, differences before and after bundle adjustment are unveiled via a numerical analysis regarding ground truth information. In this analysis, different instances of the proposed model are compared in order to qualify the contribution of each single step to the overall result. Regarding rotations, these instances are:

- BP-SDP: Rotations after breadth-propagation and SDP estimation ($\mathcal{R}^\star$).
- BP-Lie: Rotations after breadth-propagation and Lie algebraic averaging. SDP estimation has not been performed.
- BP-SDP-Lie: Rotations after breadth-propagation, SDP estimation and Lie algebraic averaging ($\mathcal{R}^{\star\star}$).
- BP-SDP-Lie-**I**: Rotations after breadth-propagation, SDP estimation and Lie algebraic averaging using $\mathbf{W}_{\Delta\mathbf{r}} = \mathbf{I}$.

Translations are estimated based on $\mathcal{R}^{\star\star}$, but it is distinguished between a robust and a non-robust version:

- TE: Translations without considering outliers.
- TE-Rob: Translations including outlier elimination as described in Section 4.4.3 ($\mathcal{T}^{\star\star}$).

		fountain-P11	*Herz-Jesu-P25*	*castle-P19*	*castle-P30*
before BA	BP-SDP ($\mathcal{R}^\star$)	0.28	0.48	1.36	1.20
	BP-Lie	**0.25**	**0.21**	**0.65**	0.58
	BP-SDP-Lie ($\mathcal{R}^{\star\star}$)	**0.25**	**0.21**	**0.65**	0.58
	BP-SDP-Lie-**I**	**0.25**	**0.21**	0.68	**0.57**
	L1-IRLS [1]	0.29	0.27	1.35	0.72
	[2]	0.52	0.57	-	1.65
	DSE(BA_3) [3]	0.45	0.39	-	0.96
after BA	$\mathcal{R}^{BA}$-SC	**0.03**	**0.02**	**0.08**	**0.04**
	$\mathcal{R}^{BA}$	0.16	0.09	0.43	0.30
	[2]	0.20	0.19	-	0.48
	spct/SDP [4]	0.42	0.35	-	-

Table 5.2.: Mean angular distance of different versions of the proposed model before and after bundle adjustment (BA) using Exif-information for the interior orientation. Results are compared to the robust L1-IRLS model of Chatterjee & Govindu [2013] [1], the global method in Jiang et al. [2013] [2], the DSE(BA_3) in Jiang et al. [2015] [3] and the spectral/SDP method in Arie-Nachimson et al. [2012] [4]. Results of [2], [3] and [4] are taken from the respective publications. Note that no results for *castle-P19* are provided in [2], [3] and [4] and additionally no results for *castle-P30* in [4]. All values are given in [°].

Results are presented in Tables 5.2, 5.3, 5.4 and 5.5. The first two tables contain the angular and metric distances to the ground truth using Exif-information as interior orientation of the cameras. The latter two show these differences regarding calibrated interior orientation. Note that in the first case, bundle adjustment is performed twice, once including self calibration, i.e. an estimation of the interior orientation parameters ($\mathcal{R}^{BA}$-SC and $\mathcal{T}^{BA}$-SC), and once fixing the values from the Exif-information ($\mathcal{R}^{BA}$ and $\mathcal{T}^{BA}$). Bundle adjustment with interior orientation from actual calibration data is performed without self calibration, only.

The accuracies of the proposed method are compared to various state-of-the-art approaches. With exception of the L1-IRLS-method of Chatterjee & Govindu [2013] that is performed on the equivalent set of relative orientations as the proposed method, all values are taken from the respective publications as explained in the captions of the tables. Regarding the results after bundle adjustment, it often is not known, whether the interior orientation is also estimated or not. According to a discussion with Zhaopeng Cui and Nianjuan Jiang, authors of Jiang et al. [2013], Cui et al. [2015] and Jiang et al. [2015], a self calibration is performed when Exif information is used. From VisualSFM it is known that a self calibration is performed, as well. Another factor influencing the results is the estimation of additional distortion coefficients (e.g. tangential distortion), which were not considered in the proposed method.

The angular distances in Table 5.2 reveal that, as soon as the Lie algebraic averaging is performed, the results lie very close together. This especially counts for *fountain-P11* and *Herz-Jesu-P25*, which have a comparably simpler geometry. The improvement from BP-SDP to all other instances, in which Lie-algebraic averaging is performed, is significant. Using initial rotations from the SDP does not

		fountain-P11	*Herz-Jesu-P25*	*castle-P19*	*castle-P30*
before BA	TE	0.038	0.109	0.447	0.980
	TE-Rob ($\mathcal{T}^{\star\star}$)	**0.035**	0.083	**0.428**	**0.950**
	[2]	0.053	0.106	-	1.158
	DSE(BA_3) [3]	0.072	**0.061**	-	1.620
after BA	$\mathcal{T}^{BA}$-SC	**0.007**	**0.013**	**0.081**	**0.044**
	$\mathcal{T}^{BA}$	0.049	0.030	0.685	0.443
	[2]	0.014	0.064	-	0.235
	DSE(BA_3) [3]	0.011	0.056	-	0.200
	spct/SDP [4]	0.027	0.052	-	-
	L_1 [5]	**0.007**	0.026	-	0.167
	1DSfM [6]	0.032	0.065	-	-
	VisualSFM [7]	0.021	0.045	-	0.190

Table 5.3.: Mean Euclidean distance of two different versions of the proposed model before and after bundle adjustment (BA) using Exif-information for the interior orientation. Results are compared to the global method in Jiang et al. [2013] [2] and the DSE(BA_3) in Jiang et al. [2015] [3]. After bundle adjustment there is an additional comparison to Arie-Nachimson et al. [2012] [4], Cui et al. [2015] [5], Wilson & Snavely [2014] [6] and VisualSFM [Wu, 2013] [7]. Results of [2], [3], [4] and [5] are taken from the respective publications, results of [6] and [7] are reported in Cui et al. [2015]. Note that no results for *castle-P19* are provided in the compared publications, for *castle-P30* no results are provided in in [4] and [6]. All values are given in [m].

have a positive effect, thus it does not matter which rotations, $\mathcal{R}$ or $\mathcal{R}^{\star}$, are used to initialize the Lie algebraic averaging, which underlines the findings of Section 5.2.1. This issue will be revisited in Section 5.3.2. For the more complex *castle*-images, the usage of unit weights instead of weights derived from covariance information affects the rotations but positively as well as negatively.

In comparison to the state-of-the-art methods of Chatterjee & Govindu [2013] (L1-IRLS-method [1]), Jiang et al. [2013] ([2]), Jiang et al. [2015] (DSE(BA_3) [3]) and Arie-Nachimson et al. [2012] (spct/SDP [4]), before and after bundle adjustment, the proposed model achieves more accurate results also without performing a self calibration. Except for the L1-IRLS-method ([1]), all results are taken from the respective publications, thus homologous points and relative orientations are different and only the images are equivalent.

The translation error in Table 5.3 shows that, except for *fountain-P11*, the outlier elimination has an effect in the range of a few centimeters. Again, the methods used for comparison consist of the two models by Jiang et al. [Jiang et al., 2013, 2015] ([2]) and (DSE(BA_3) [3]) and Arie-Nachimson et al. [Arie-Nachimson et al., 2012] (spct/SDP [4]), plus Cui et al. [2015] (L_1 [5]), Wilson & Snavely [2014] (1DSfM [6]) and [Wu, 2013] (VisualSFM [7]). The proposed method achieves comparable and partly best results before and after bundle adjustment. Without self calibration, the accuracy of the translations decreases significantly. In case of *fountain-P11* and *Herz-Jesu-P25* the results of $\mathcal{R}^{BA}$-SC are in the region of the accuracy of the ground truth itself, which can be gauged looking at Figure 7 in Strecha et al. [2008]. The *castle-P19* dataset has not been processed in any of the

		fountain-P11	*Herz-Jesu-P25*	*castle-P19*	*castle-P30*
before BA	BP-SDP-Lie ($\mathcal{R}^{\star\star}$)	**0.02**	**0.03**	**0.14**	**0.24**
	L1-IRLS [1]	**0.02**	**0.03**	0.41	0.39
	DSE(BA_3) [3]	**0.02**	0.06	-	0.27
	EIG-SE(3)-Iter [8]	0.03	0.06	1.48	0.47
	EIG-SE(3)-MCB [8]	0.04	0.06	2.46	0.77
	[9]	0.03	0.14	3.69	1.97
after BA	BP-SDP-Lie ($\mathcal{R}^{\star\star}$)	**0.02**	**0.02**	**0.04**	**0.03**
	spct/SDP [4]	**0.02**	0.05	-	-
	EIG-SE(3)-Iter [8]	0.04	0.04	0.06	0.05
	EIG-SE(3)-MCB [8]	0.03	0.04	0.06	0.05
	[9]	0.03	0.04	0.05	0.05

Table 5.4.: Mean angular distance of the proposed model before and after bundle adjustment (BA) using ground truth information for the interior orientation. Results are compared to the robust L1-IRLS model of Chatterjee & Govindu [2013] [1], the DSE(BA_3) method in Jiang et al. [2015] [3], the spectral/SDP method in Arie-Nachimson et al. [2012] [4], two different versions of Arrigoni et al. [2015a] [8] and the model of Ozyesil et al. [2015] [9]. The results of [3], [4] and [8] are taken from the respective publications, the results of [9] are reported in Arrigoni et al. [2015a]. Note that no results for *castle-P19* are provided in [3] and [4] and additionally no results for *castle-P30* in [4]. All values are given in [°].

publications used for comparison.

The fundamental difference when comparing Tables 5.2 and 5.3 with 5.4 and 5.5 is that the accuracy generally improves when using the calibrated interior orientation. In Table 5.4, the angular distance of only the full model (BP-SDP-Lie) is compared to various state-of-the art models. Regarding *fountain-P11* and *Herz-Jesu-P25*, the differences are very small, whereas they are more significant for *castle-P19* and *castle-P30*. In particular, this is meaningful for the comparison with L1-IRLS [Chatterjee & Govindu, 2013], because the relative orientations are identical so that both rotation estimations can be compared without distortion. In comparison with the remaining approaches, especially for *castle-P19*, the differences in the angular distance before bundle adjustment are large not only to the proposed method but also among themselves. This can be attributed to the presence of outliers in the relative orientations and their rigorous elimination in the proposed method. After bundle adjustment all results lie close together.

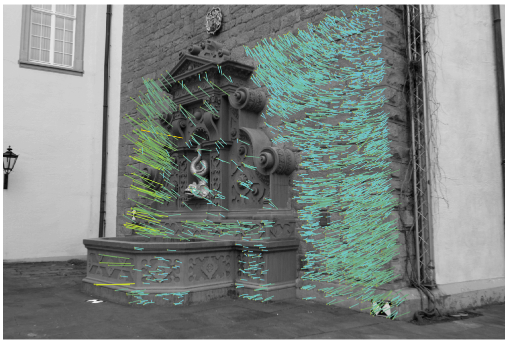
A similar effect is visible regarding global translations (Table 5.5). Before bundle adjustment, the results partly vary considerably. Except for the approach of Wilson & Snavely [2014] (1DSfM [6]), the translations of *fountain-P11* and *Herz-Jesu-P25* are very accurate and in the range of few mm. Cui et al. [2015] discuss that the approach of Wilson & Snavely [2014] is not suitable for this sequential type of image data. The translations of the two castle datasets before bundle adjustment are estimated considerably more accurately with the new approach, which relates to a successful outlier elimination. After bundle adjustment, all results lie very close together.

		fountain-P11	*Herz-Jesu-P25*	*castle-P19*	*castle-P30*
before BA	TE-Rob ($\mathcal{T}^{\star\star}$)	**0.005**	**0.012**	**0.242**	**0.591**
	DSE(BA_3) [3]	0.009	**0.012**	-	1.040
	EIG-SE(3)-Iter [8]	0.236	1.152	4.986	1.974
	EIG-SE(3)-MCB [8]	0.008	0.357	3.967	3.866
after BA	TE-Rob ($\mathcal{T}^{\star\star}$)	0.004	0.009	0.043	0.056
	DSE(BA_3) [3]	**0.003**	0.006	-	0.100
	spct/SDP [4]	0.005	0.008	-	-
	L_1 [5]	**0.003**	**0.005**	-	**0.021**
	1DSfM [6]	0.034	0.036	-	-
	VisualSFM [7]	0.004	0.006	-	0.071
	EIG-SE(3)-Iter [8]	**0.003**	0.022	0.034	0.035
	EIG-SE(3)-MCB [8]	**0.003**	0.008	0.035	0.034
	[10]	**0.003**	**0.005**	**0.026**	0.022

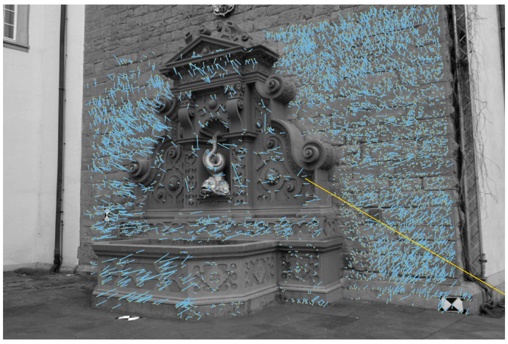
Table 5.5.: Mean Euclidean distance of the proposed model before and after bundle adjustment (BA) using ground truth information for the interior orientation. Results are compared to Jiang et al. [2015] [3], Arie-Nachimson et al. [2012] [4], Cui et al. [2015] [5], Wilson & Snavely [2014] [6], VisualSFM [Wu, 2013] [7], two different versions of Arrigoni et al. [2015a] [8] and to Moulon et al. [2013] [10]. The results of [3], [4], [5] and [8] are taken from the respective publications, the results of [6], [7] and [10] are reported in Cui et al. [2015]. Note that no results for *castle-P19* are provided in [3], [4], [5], [6] and [7] and additionally no results for *castle-P30* in [4] and [6]. All values are given in [m].

Tables 5.2-5.5 show the accuracy of the estimated global orientations with respect to ground truth values. Regarding the precision of the results, i.e. the covariance matrices of the global orientations after bundle adjustment, all values of the four datasets lie in the same order of magnitude (approximately $\pm 0.0005°$ for rotations and $\pm 0.8\,\mathrm{mm}$ for translations in standard deviation).

The accuracy of the global orientations in general depends on many factors like the quality of the relative orientations, which in turn depends on the homologous points. Because these are selected randomly, e.g. via *RANSAC*, the orientation results often vary slightly when computed repeatedly (approximately $\pm 0.01°$ for rotations and $0.005\,\mathrm{m}$ for translations). This hampers a sound comparison to other approaches. Moreover, the bundle adjustment induces additional factors that are hard to control and compare. Often, in the publications used for comparison, there is little information about how parameters are treated, e.g. regarding the estimation of interior orientations parameters or distortion coefficients or the definition of convergence criteria. On the other hand, if one sees the proposed method as a whole, it can be considered a model for an accurate estimation of global orientations, whose results are comparable to and partly better than the state-of-the-art. The comparison with the L1-IRLS method of Chatterjee & Govindu [2013] is more revealing because identical relative rotations have been used. Constantly, the proposed method performs equally well or produces more accurate global rotations than L1-IRLS, which agrees with the analysis in Section 5.2.2 using synthetic data.



(a) I_1 before BA



(b) I_3 before BA

(c) I_1 after BA

(d) I_3 after BA

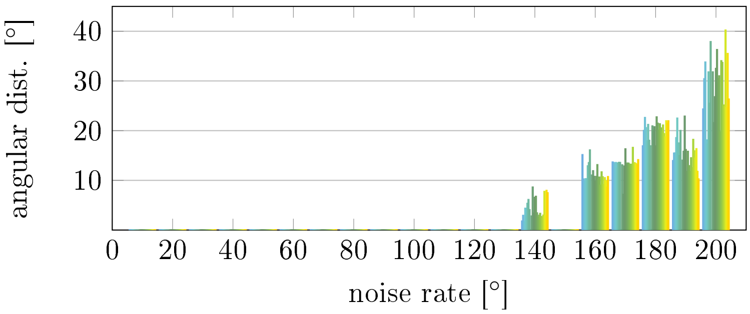
Figure 5.16.: Residuals in I_1 and I_3 of *fountain-P11* before and after bundle adjustment (BA) enlarged by factor 20.

Residuals in image space

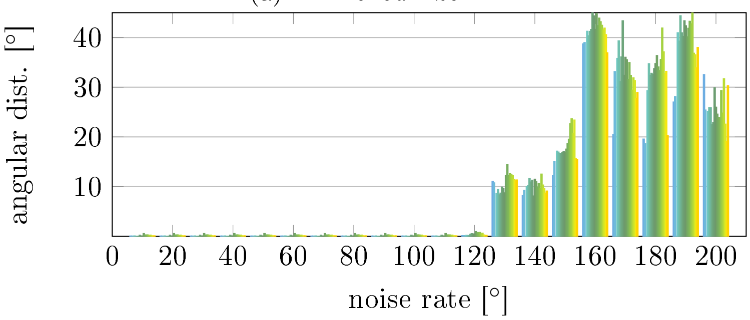
A second analysis compares the residuals in image space before and after bundle adjustment, i.e. the pointwise reprojection error. Both, the size of the errors and remaining systematic effects can thus be revealed. A comparison for two images of the *fountain-P11* dataset is given in Figure 5.16. Note that residuals are shown for all observations of points visible in at least three images, most of these were not used for the estimation of initial translations. Before bundle adjustment, clear systematic effects can be seen for direction and length of the residuals (Figures 5.16a and 5.16b). In I_3 one outlier is also visible. After bundle adjustment the average length of the residual vectors decreased and the directions appear randomly distributed (Figures 5.16c and 5.16d). The outlier in I_3 has been removed. On the one hand, this evaluation shows that the initial solution derived by the proposed approach does not lead to a best and unbiased solution, on the other hand it also reveals that the final bundle adjustment is able to solve these issues, which corresponds to the numerical analysis of the accuracy with respect to ground truth information in the previous part.

5.3.2. Lie algebraic averaging - basin of convergence

In the previous section, the angular distance of different versions of the proposed method was discussed. An important result was that the accuracy of the global rotations after Lie algebraic



(a) *Herz-Jesu-P25*



(b) *castle-P19*

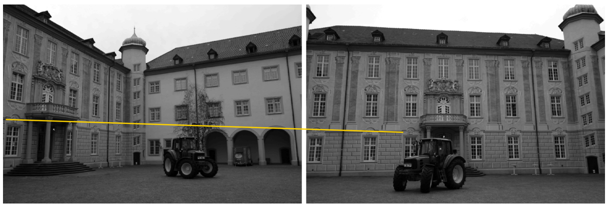
Figure 5.17.: Angular distance of $\mathcal{R}^{\star\star}$ for the *Herz-Jesu-P25* (5.17a) and *castle-P19* dataset (5.17b) depending on corrupted initial global rotations $\mathcal{R}$ (bars).

averaging and before bundle adjustment for all four datasets was insensitive against the type of initialization, i.e. using global rotations after breadth propagation $\mathcal{R}$ or after SDP estimation $\mathcal{R}^{\star}$ did not cause any difference in the results. The question arises how good the initial global rotations must be in order for the Lie algebraic averaging to converge to the global optimum. In other words, what is the basin of convergence of the Lie algebraic averaging? Because the Lie algebraic averaging problem (4.30) is not convex, the term *global optimum* in this sense is defined as the solution that is derived with initial rotations after breadth propagation.

In this experiment, an empirical evaluation is performed. All global rotations derived by breadth-propagation are corrupted by random rotations, similar to the robustness experiment in Section 5.2.2. Random rotations are computed based on random Euler angles, normally distributed with $\sigma = t \cdot 10^{\circ}, t = [1, 2, \ldots, 20]$. For every t, ten independent repetitions are conducted. The average angular distance for each global rotation of the *Herz-Jesu-P25* dataset is depicted in Figure 5.17a. Until 130° noise level the results are not affected and equal to the global optimum. The basin of convergence is thus surprisingly large, which coincides with the findings of Table 5.2, which states that irrespective of the initial rotations, the Lie algebraic averaging produces equivalent results.

The image configuration of the *Herz-Jesu-P25* dataset is relatively simple. In order to show how this might influence the basin of convergence, the same experiment is conducted on the *castle-P19* dataset, which has a circular arrangement of images and, thus, a higher variety in the global rotations. Results are shown in Figure 5.17b. Small differences in the angular distances are already measurable

(a) $\mathbf{p}_2^{3543} - \mathbf{p}_3^{3543}$

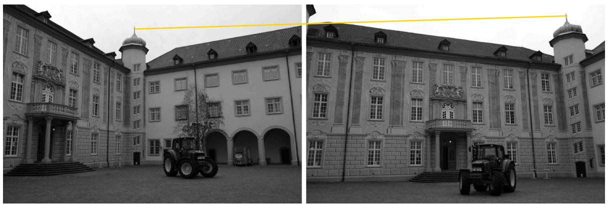


(b) $\mathbf{p}_2^{3543} - \mathbf{p}_6^{3543}$

(c) $\mathbf{p}_2^{3543} - \mathbf{p}_{10}^{3543}$

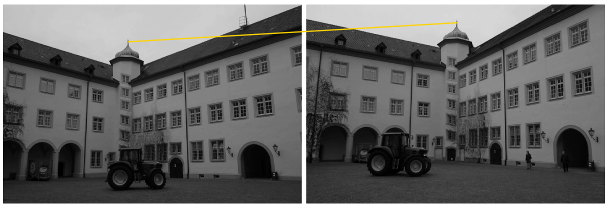
(d) $\mathbf{p}_3^{3543} - \mathbf{p}_{11}^{3543}$

(e) $\mathbf{p}_1^{539} - \mathbf{p}_3^{539}$



(f) $\mathbf{p}_2^{539} - \mathbf{p}_5^{539}$

(g) $\mathbf{p}_4^{539} - \mathbf{p}_{18}^{539}$



(h) $\mathbf{p}_{18}^{539} - \mathbf{p}_{19}^{539}$

Figure 5.18.: Rejected homologous points in the *castle-P19* dataset.

at 120° noise level, which however is hardly visible in the figure. In total, the differences are higher than for the *Herz-Jesu-P25* dataset, but the basin of convergence is still rather large.

In summary, taking the results of Figure 5.9 into account, this investigation shows that the SDP estimation can be skipped in most practical cases. The basin of convergence of the Lie algebraic averaging is relatively large, even for more complex image arrangements. Although SDP often provides a more accurate set of initial global rotations compared to rotations derived by breadth-propagation (cf. Figure 5.9), this characteristic is compensated by the Lie algebraic averaging. Thus, in the following experiments, the SDP-estimation is not performed and the term $\mathcal{R}^{\star\star}$ is also used for the configuration BP-Lie.

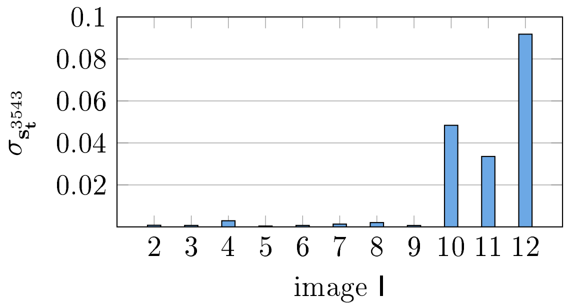
5.3.3. Outlier elimination

In Section 5.3.1, it was shown that the outlier elimination during the estimation of global translations leads to superior results, especially regarding both *castle*-datasets. Although the outliers did not considerably distort the result compared to the estimation without outlier elimination, there was a general improvement in the range of a few centimeters. In this section, two exemplary points, rejected as outliers in the *castle-P19* dataset, are studied in more detail.

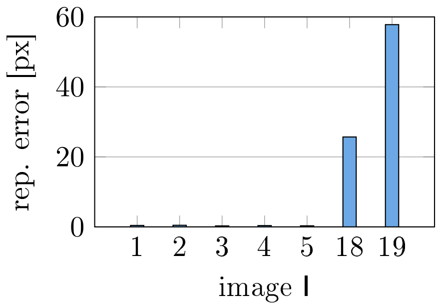
In total, 224 points have been selected, which form 30 point tracks. Two out of these points were considered to be outliers, i.e. they have been measured erroneously in at least one image. In Figure 5.18, exemplary matchings for selected image pairs are shown. Figures 5.18a-5.18d show point 3543, which has inconsistent measurements: in images I_2-I_8 the point is measured at the top left corner of the ground floor window directly left of the entrance. In images I_{10}-I_{12}, this point is measured one window further to the left. In image space, this is an error of more than 400 px. Considering the geometry of the scene and the configuration of the images, it becomes clear that the line between the two window corners, at which point 3543 was measured, is approximately parallel to the plane formed by the projection centers of all images captured. Therefore, taking any image pair, both window corners approximately lie in one plane, i.e. are coplanar, which makes the set of homologous points prone to mismatches, especially considering the repetitive structure at the windows. Consequently, the outlier has not been detected by the reprojection error after local spatial intersection but by the triplet scale constraint. Figure 5.19a shows the median standard deviation computed for each image, considering all respective triplet combinations. For instance, the bar at position 2 shows the median of all standard deviations in $\mathbf{s}_{\mathbf{t}}^{3543}$, which include I_2.

As seen from images I_{10}-I_{12} this standard deviation is considerably higher. The only consistent triplet, formed by at least one of these images, for which a low standard deviation is computed, is the triplet $\{\mathsf{I}_{10}, \mathsf{I}_{11}, \mathsf{I}_{12}\}$. Any combination with an image out of $\{\mathsf{I}_2$-$\mathsf{I}_8\}$ results in a significant difference between the individual scales and thus in a higher $\sigma_{\mathbf{s}^l_{\mathbf{t}_{ijk}}}$. Thus, because the image set $\{\mathsf{I}_2, \ldots, \mathsf{I}_8\}$ is larger than $\{\mathsf{I}_{10}, \mathsf{I}_{11}, \mathsf{I}_{12}\}$, a decision can be taken about in which images the observation is considered to be an inlier and in which it is considered to be an outlier. Hence, constraints for the images I_2-I_8 are kept in the estimation. One could introduce a new point for the image set $\{\mathsf{I}_{10}, \mathsf{I}_{11}, \mathsf{I}_{12}\}$ but because the connection between the two sets is lost, the contribution of these additional constraints is minor.

Point 539 was identified as outlier during spatial intersection. Correct, but inconsistent matches are shown in Figures 5.18e, 5.18f and 5.18h. The crossing between the two inconsistent measurements is the wrong matching depicted in Figure 5.18g, i.e. again two sets of images, $\{\mathsf{I}_1, \mathsf{I}_2, \mathsf{I}_3, \mathsf{I}_4, \mathsf{I}_5\}$ and $\{\mathsf{I}_{18}, \mathsf{I}_{19}\}$, have a different but within themselves correct observation of $\mathbf{P}^{539}$. During pairwise spatial intersection, any combination of elements from both sets leads to a reprojection error higher than the threshold τ_r. The median reprojection error for each image, considering all pairwise combinations is depicted in Figure 5.19b. For the same reason as for $\mathbf{P}^{3543}$, the median is decisive because of the different size of the two image sets $\{\mathsf{I}_1, \ldots, \mathsf{I}_5\}$ and $\{\mathsf{I}_{18}, \mathsf{I}_{19}\}$. In consequence, not all observations are rejected but only those in images I_{18} and I_{19}. It is noteworthy that, when disabling the reprojection error constraint, this inconsistency would have also been detected by the triplet scale constraint later on. This redundancy in the outlier detection is an important side effect. Omitting the reprojection



(a)



(b)

Figure 5.19.: Median standard deviation of $\mathbf{s}_\mathbf{t}^{3543}$ for every corresponding image, computed considering all respective triplets (5.19a) and median reprojection error for the local pairwise spatial reconstruction of point $\mathbf{P}^{539}$, considering all pairwise combinations (5.19b).

error constraint, however, is not advisable because computationally it comes almost for free and the detection of outliers at this early stage saves the computation and evaluation of the triplet constraints.

5.4. Various types of data

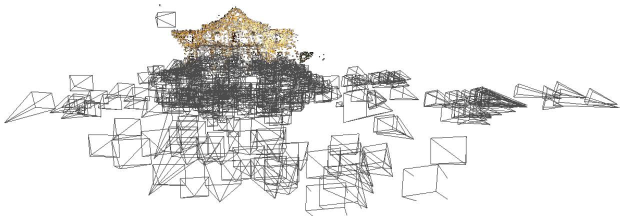
In the experiments shown in the previous sections, the used image sets were small, consisting of at maximum 30 and 50 images in the benchmark and synthetic data, respectively. The datasets investigated in this section are significantly larger and shall demonstrate the versatility of the proposed approach. The first part of this section addresses datasets from image-hosting websites. In the second part, self-acquired images from two flights with a UAV are examined.

5.4.1. Images from image-hosting websites

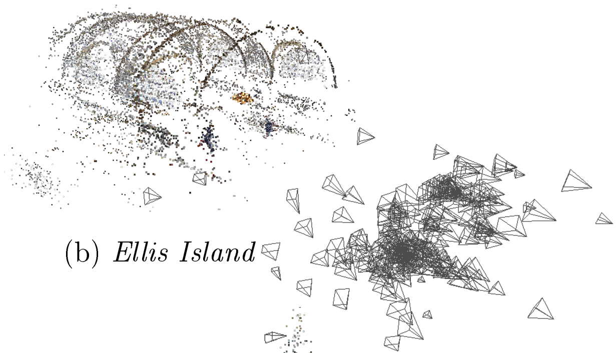
The images investigated in this section are captured with different camera models and interior orientations. Initial values for these parameters from Exif-information come together with the images [Wilson & Snavely, 2014]. Also homologous points are provided and used here in order to derive information about overlapping images and possible pairwise combinations. It has to be noted that these homologous points are not free of outliers. Thus, the preliminaries are similar to the evaluation of the benchmark data reported in Section 5.3 with the difference that the homologous points do not stem from *Micmac*.

For the estimation of rotations the BP-Lie model is used (cf. Section 5.3.1). Translations are estimated as in the previous sections (TE-Rob) using slightly higher thresholds (see Table 5.1). During bundle adjustment, the interior orientation parameters are not estimated and stay fixed.

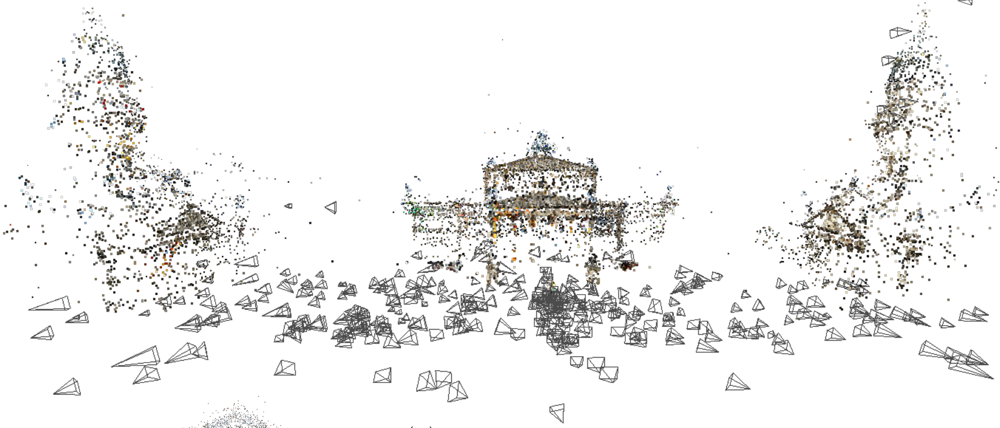
Figure 5.20 shows exemplary views of the reconstruction results after bundle adjustment. The focus of this section is not an evaluation of accuracy but to show the applicability to various types



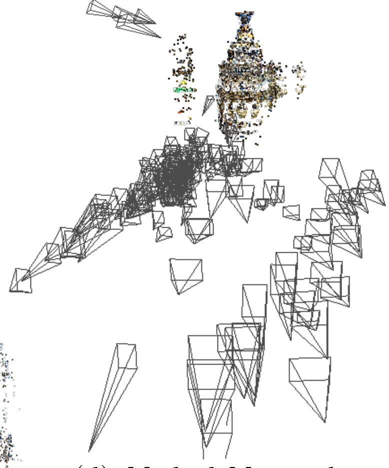
(a) *Alamo*



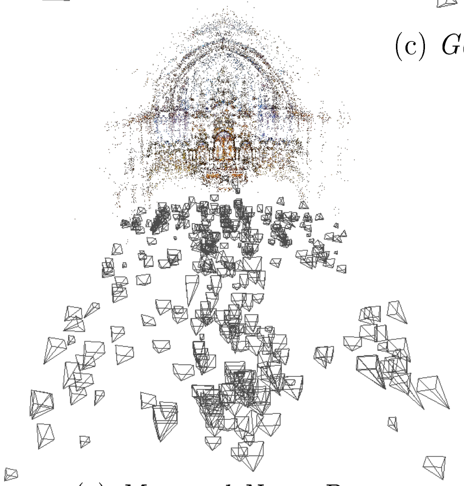
(b) *Ellis Island*



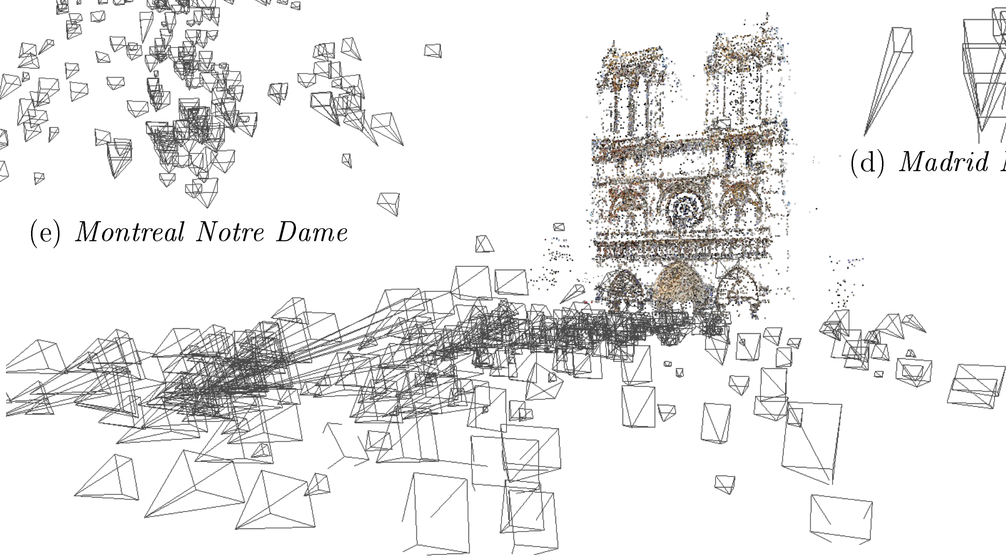
(c) *Gendarmenmarkt*



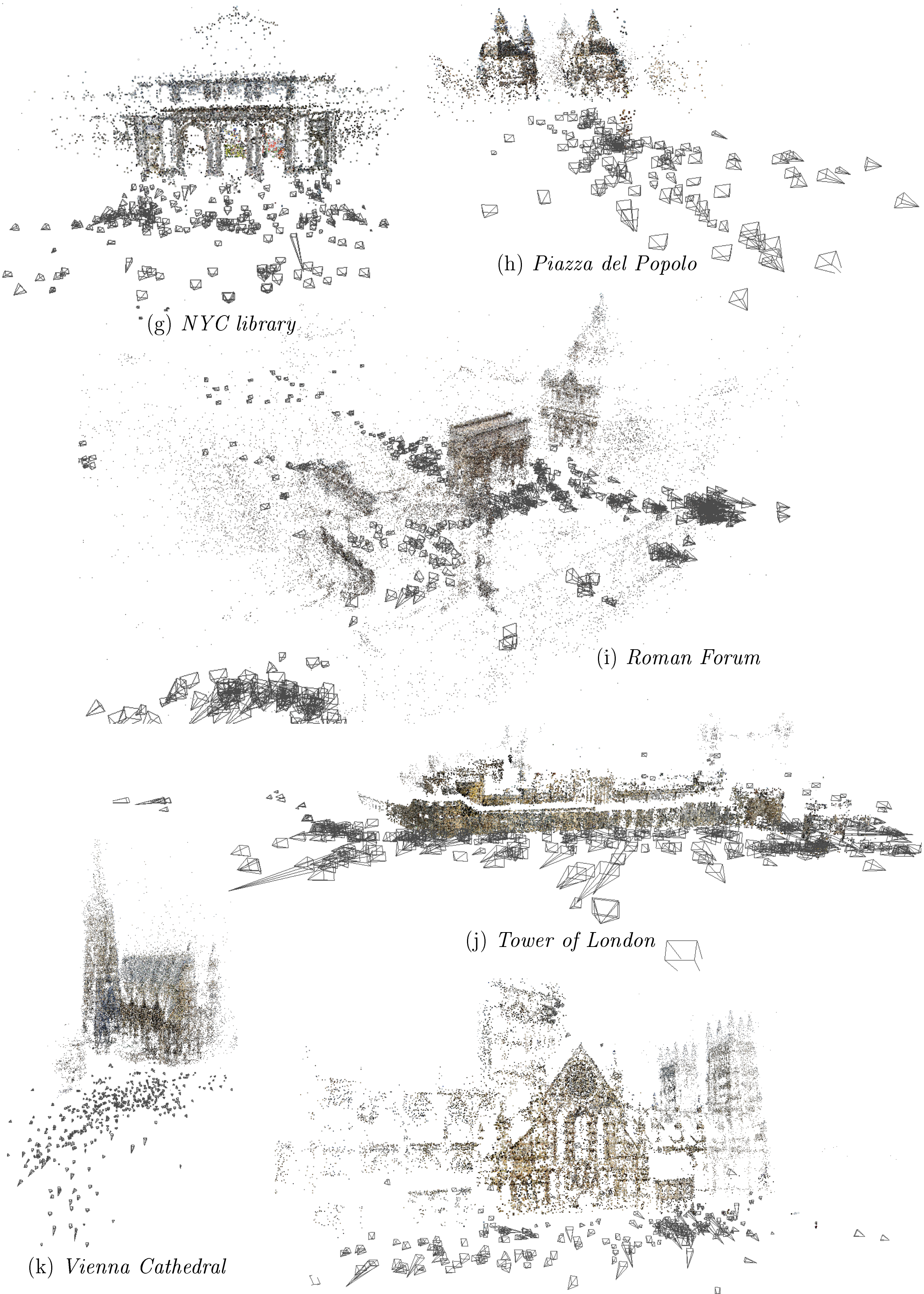
(d) *Madrid Metropolis*



(e) *Montreal Notre Dame*



(f) *Notre Dame de Paris*



(g) *NYC library*

(h) *Piazza del Popolo*

(i) *Roman Forum*

(j) *Tower of London*

(k) *Vienna Cathedral*

(l) *Yorkminster*

Figure 5.20.: Reconstruction of various image datasets from [Wilson & Snavely, 2014] using the proposed method.

	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{I}^{\star\star}\rvert$	$\lvert\{\mathbf{t}_{ij}^0, \mathbf{R}_{ij}^0\}\rvert$	$\lvert\{\mathbf{R}_{ij}^\star, \mathbf{t}_{ij}^\star\}\rvert$
Alamo	571	526	72513	59165
Ellis Island	230	216	11213	9638
Madrid Metropolis	366	310	11480	9092
Montreal Notre Dame	459	423	38057	36541
Notre Dame de Paris	553	538	43782	33372
NYC Library	359	256	13064	10962
Piazza del Popolo	327	301	14405	12752
Roman Forum	1062	789	40179	29652
Tower of London	469	395	15527	13409
Vienna Cathedral	864	628	73159	51759
Yorkminster	418	338	17553	13376
Gendarmenmarkt	598	429	8662	7706

Table 5.6.: Important numbers of the results of the images of [Wilson & Snavely, 2014] using the proposed method.

of data. Because there is no ground truth information, no comparative but only a visual analysis is made. All reconstructions in Figure 5.20 show a reasonable representation of the respective object of interest. Some statistical information about these datasets is provided in Table 5.6: The columns $|\mathcal{I}|$ and $|\mathcal{I}^{\star\star}|$ give the number of images in the largest connected component and after orientation estimation ($\{\mathcal{T}^{\star\star}, \mathcal{R}^{\star\star}\}$, i.e. before bundle adjustment), respectively. In the next two columns the number of relative orientations before and after breadth-propagation ($|\{\mathbf{t}_{ij}^0, \mathbf{R}_{ij}^0\}|$ and $|\{\mathbf{R}_{ij}^\star, \mathbf{t}_{ij}^\star\}|$) are shown. The difference of these two columns indicates the number of relative orientations excluded during breadth propagation.

It can be seen that in all datasets the number of images decreases after applying the proposed approach, because for some of them no valid translation constraint could be established. While this is a usual effect considering similar evaluations in related works (e.g. [Wilson & Snavely, 2014; Cui et al., 2015; Cui & Tan, 2015]), it reveals an important drawback of the proposed model. If, for a subset of images, all points in the given point tracks are considered to be outliers, there is no opportunity to apply further constraints for the respective images. Therefore, theses images are excluded from estimation. An extension of the proposed model would capture these cases and try to find additional points to replace those that have been classified as errors. In this way the number of reconstructed images could be increased.

The number of relative orientations varies between 8662 for the *Gendarmenmarkt* and more than 70000 for the *Alamo* and *Vienna Cathedral* datasets, although the number of images does not reflect this large difference. The reason is that, in the *Gendarmenmarkt* scene (Figure 5.20c), the variety of viewing directions is larger, whereas almost all images in the *Alamo* and *Vienna Cathedral* scenes (Figures 5.20a and 5.20k) point in the same direction.

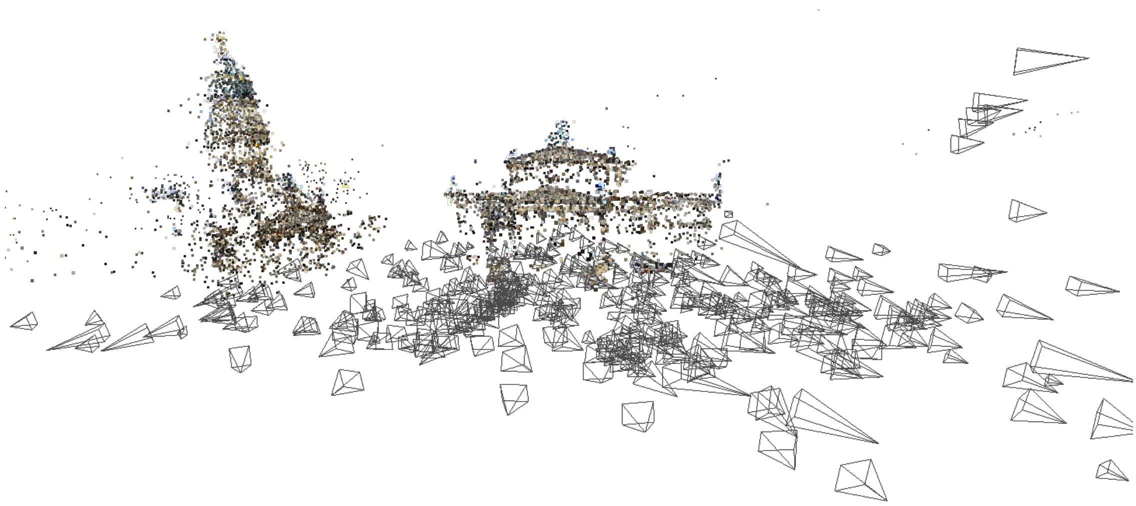
The *Gendarmenmarkt* images caused problems in Wilson & Snavely [2014] because the scene is almost symmetric, i.e. two buildings on opposites sides of a plaza look highly identical, which led to wrong reconstructions. A similar effect also occurs using the model proposed in this thesis when changing the starting image. Figure 5.21a shows an erroneous reconstruction using the starting image pictured in Figure 5.21c. Apparently, regarding pairwise relative orientations, the building on the right side of this image was identified to be equal to the almost identically looking building on the opposite side of the plaza, which led to only one reconstructed building. Thus, starting from this image, leads to an erroneous estimation of global rotations in the very beginning of the sequential breadth-propagation. A considerable amount of pairwise relative orientations supports these wrong estimates, hence, the correct relative orientations are considered as outliers. The starting image in 5.21b was used for the correct reconstruction in Figure 5.20c. Thus, the selection of the starting image *does* affect the final image orientation. In order to show, how large the risk for an erroneous estimation is, a systematic evaluation is conducted, in which every image is used as starting image. Then the differences in the global rotations between each individual estimation and the correct estimation are checked. Approximately 75% of the images led towards a correct estimation. Because of its symmetry, this scene is peculiar, a comparable effect could not be observed in any of the remaining image sets.

The reconstructions in Figure 5.20 suggest that the proposed method provides accurate image orientation parameters so that a bundle adjustment converges to a correct solution. It remains to actually show the quality of these initial orientations. Figure 5.22 shows a comparison of the orientation before (5.22a) and after bundle adjustment (5.22b) of the *Yorkminster* dataset (cf. Figure 5.20(l)) viewed from above in an orthographic projection. One can see that the flatness of the object points increases significantly after bundle adjustment, which reflects the structure of the pictured church. However, the pattern of the images hardly changes, which stresses the quality of the image orientations before bundle adjustment. The high scatter of the points before bundle adjustment can be explained by a rather unfavorable base to distance ratio leading to many glancing intersections. Moreover, no drift is visible, which would reveal itself e.g. by a bending facade plane.

In summary, the orientation of large image sets, which are considered to be ambitious because only a rough guess of the interior orientation is available, is possible using the proposed method. High quality initial parameters for a final bundle adjustment are provided. Note that the run time of the proposed method is not investigated. The implementation is not optimized regarding efficiency, which is why a qualified run time analysis is not possible.

5.4.2. UAV image sequences

In this section, two image sequences acquired from a UAV are evaluated. As for the benchmark data, homologous points are computed using *Micmac*. Reconstruction results of the *Leibniztempel* sequence are depicted in Figure 5.23, both before (5.23a) and after bundle adjustment (5.23b). The challenge of this dataset lies in the high variation of perspectives. Images are taken from all directions and with varying distance to the object. Both reconstructions look similar, differences can only be seen in some image locations, e.g. those highlighted by the red circles. These can be

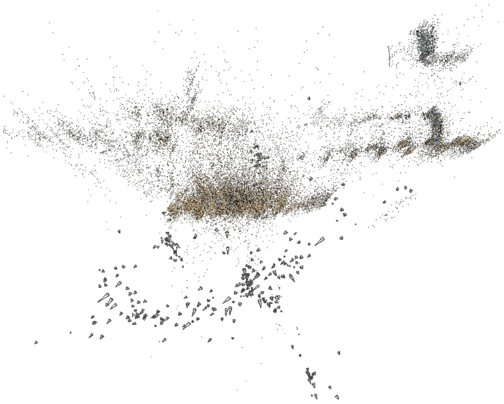


(a) Erroneous reconstruction of *Gendarmenmarkt*

(b) Correct starting image

(c) Problematic starting image

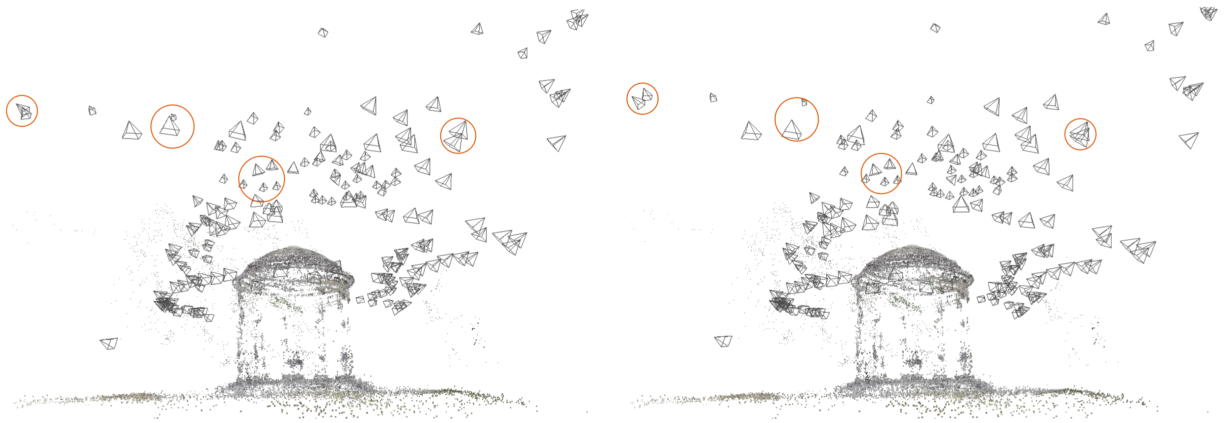
Figure 5.21.: Erroneous reconstruction of *Gendarmenmarkt* and two starting images that lead to different results.



(a)

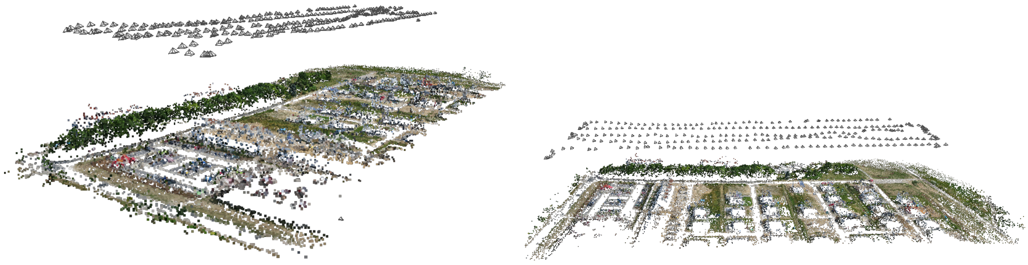
(b)

Figure 5.22.: Reconstruction of the *Yorkminster* dataset before (5.22a) and after bundle adjustment (5.22b).



(a) before BA (b) after BA

Figure 5.23.: Reconstruction of the *Leibniztempel* dataset (see Figure 5.4a) before and after bundle adjustment (BA).



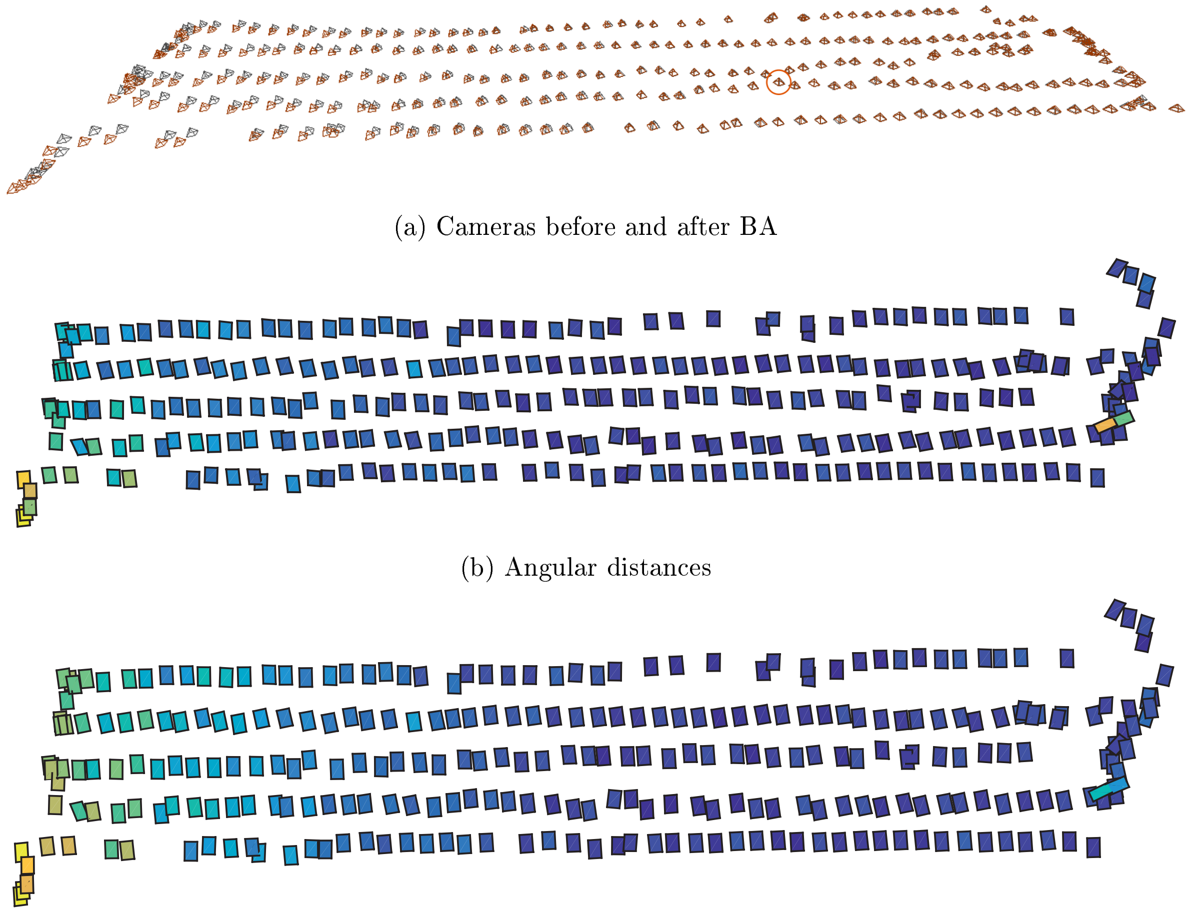
(a) (b)

Figure 5.24.: Reconstruction of the *Wettbergen* dataset (see Figure 5.4b).

considered small, moreover, they do not have a visible effect on the reconstruction results.

In Figure 5.24, the reconstruction of the second UAV-images, the flight above the housing area in *Wettbergen* is shown. As for the *Leibniztempel* sequence, the reconstruction looks as expected. This time, only the results after bundle adjustment are depicted from two different perspectives because no differences can be grasped by visual inspection of individual reconstructions. Therefore, a direct comparison between both orientation estimations is provided in Figure 5.25a, which shows the estimated camera locations before (black) and after bundle adjustment (red). The starting image is highlighted by a red circle. It can be seen that the difference between both estimations is largest on the left side and smallest close to the center of the view.

A visualization of the differences separately for rotations and translations before and after bundle adjustment is provided in Figures 5.25b and 5.25c, respectively. These patterns are similar and support the plotted camera locations in Figure 5.25a. The maximal distances are approximately 4.2m and 1.6°. The distribution of the differences is as expected, because the largest corrections occur generally at the borders of the scene. At these regions the number of overlapping images is smaller than in the central part. Taking the density of homologous points into account, which is visible in Figure 5.24, it becomes clearer, why the distances are larger on the left side. The point



(a) Cameras before and after BA

(b) Angular distances

(c) Euclidean distances

Figure 5.25.: Camera location of the *Wettbergen* dataset before (black) and after bundle adjustment (BA) (red) (5.25a), distribution of angular distances, varying between 0 and 1.6° (5.25b) and distribution of Euclidean distances, varying between 0 and 4.2m (5.25c).

density is considerably sparser at the bottom left corner of Figure 5.25. Another reason is the selection of the starting image, which is situated in the right third.

These two datasets stand for a variety of different aerial but also terrestrial image sets, which are taken in a similar manner. Solving the two UAV-datasets did not pose any problems to the proposed model.

5.5. Synthesis - experiments

In this section, the most noticeable findings of the conducted experiments, selected by each step of the workflow, are summarized.

Preprocessing

Regarding the *preprocessing*, the M-estimation of relative orientations has a significant positive effect on the accuracy. Especially relative translations benefit from an up 70% higher accuracy regarding their standard deviation. This, of course, reflects on the quality of the global orientation parameters. The effectiveness of the breadth-propagation was proven on synthetic data. It was shown to be more robust than the current state-of-the-art rotation estimation algorithm of Chatterjee & Govindu [2013]. Up to 50% of outliers are tolerated until the quality of the rotations decreases significantly. The global rotations stemming from the propagation are sufficient to serve as initialization for the nonconvex Lie algebraic rotation averaging. Another important finding is that the selection of a starting image for the breadth-propagation may have a crucial effect on the resulting image orientations.

Estimation of rotations

The most important finding concerning the estimation of global rotations is that it produces highest accuracy compared to several state-of-the-art methods. Based on benchmark data, four different versions of the proposed algorithm are compared. The Lie algebraic averaging has the largest positive effect on the results. It could not be proven that a weighting using the covariance matrices of the relative rotations leads to a superior result, but the statistical cost is smaller. In general, the estimated rotations are of very high quality, which is shown by a comparison to the state-of-the-art. Another major finding concerns the basin of convergence of the Lie algebraic averaging. It has been shown that the result to a large extend does not depend on the initial global rotations, which, in principle, allows the usage of rotations after breadth-propagation and makes the SDP unnecessary in practice.

Estimation of translations

The estimation of translations is highly supported by the new outlier detection and elimination. It was shown on real images that the imposed criteria allow a sharp discrimination between inliers and outliers. This also counts for outliers that coincide with the pairwise epipolar geometry. The accuracy before bundle adjustment compared to ground truth constitutes the state-of-the-art. One drawback is the selection of point tracks before the detection of outliers. It may happen that for some images no constraints are established after outlier elimination or that the overlapping criterion is not fulfilled anymore, leading to a set of images, for which a translation estimation is impossible.

Overall global image orientation model

In summary, the proposed method produces very accurate results on all investigated datasets. It does not achieve the same level of precision and accuracy that is derived with a bundle adjustment but it provides accurate initialization. The better this initialization the faster the bundle adjustment converges and the lower the risk that it is caught in a local optimum. This behavior has been shown on a variety of different data, including synthetic data, benchmark sequences and more ambitious image sets from image-hosting websites and UAV flights.

6. Conclusions and outlook

This chapter concludes this thesis. The main objective is revisited and set into context to the proposed model and the results of its evaluation. Moreover, important challenges for future research are outlined.

The main objective of this thesis was to *present a novel global image orientation approach that achieves accurate orientation parameters, is robust against outliers in relative orientations and homologous points and is applicable to various types of data.* This objective was approached by combining several individual optimizations regarding the estimation of relative orientations, global rotations and translations. During each of these steps, outliers in the homologous points and relative orientations are detected and eliminated. Prior information about relative orientations is encoded as covariance matrices and used for a weighting in the estimation of global orientations.

The *preprocessing* serves the computation of an accurate foundation of relative orientations. It was shown that the M-estimation has a significant positive effect on their accuracy and that further outliers in the homologous points can be found. The new breadth-propagation algorithm showed to be highly effective and allows high outlier rates in the relative rotations. It provides a higher robustness than the comparable state-of-the-art rotation averaging algorithm of Chatterjee & Govindu [2013].

The *estimation of rotations* was designed to combine a convex SDP [Saunderson et al., 2014] and a subsequent iterative Lie algebraic averaging [Govindu, 2004]. It could be shown that the basin of convergence of the Lie algebraic averaging is in general relatively large, which allows to skip the SDP. The estimated global rotations showed to be of high accuracy, both using synthetic data and real images, with the angular accuracy being often better than using any other method that served for a comparison.

The *estimation of translations* builds on the linear method of Cui et al. [2015] and is extended by an effective two-step outlier detection and elimination. Suitable point tracks are selected, not only based on their length, but also on fulfilling a sensible distribution in image space. Outliers in the homologous points, whereas coinciding with the epipolar plane or not, are detected reliably, which allows an efficient L_2 estimation of global translation parameters.

Global image orientation models in general have the ability to overcome the heuristic and inefficient nature of a sequential estimation of initial image orientation parameters. Only a few thresholds have to be defined in order to control the estimation. In this work, these thresholds are set without considering a more in depth analysis on their influence. Future research could deal with this issue and try to learn optimal values that meet the average image data. For thresholds like the minimal number of correspondences for the estimation of a relative orientation or the number of point tracks, this learning might be cumbersome. A decision for a minimal number of correspondences leads to a compromise between a smaller number of reliable relative orientations and a larger number of perhaps less reliable relative orientations. For the number of point tracks, a higher number ensures a higher redundancy and thus statistically superior results but also affects the efficiency. Consequently, a selection of these parameters also depends on the computational resources and the desired outcome.

The main benefit of global image orientation approaches might at the same time become a major drawback: its one for all and all for one strategy. For instance, if the linear constraints during the estimation of global translations contain an outlier or lead to a singular matrix of linear constraints, e.g. due to a violated overlapping criterion, the estimated translations of all images suffer. Thus, it is crucial that, before each optimization problem is solved, those issues are taken care of. In the case of global rotations, this is achieved by breadth-propagation, which guarantees a set of consistent relative rotations. In case of global translations, the overlapping criterion of the point tracks and the outlier detection using pairwise spatial intersection and triplet scale constraints serve the constitution of valid constraints. However, the scheme of the proposed approach in this regard is error prone: Because outliers are detected and eliminated *after* point tracks have been selected, the coverage of images by overlapping point tracks might be questioned. This issue requires a subsequent evaluation of the imposed constraints, which eventually leads to a rejection of image subsets from estimation. A strategy to approach this problem consists in unifying the point track selection and outlier detection, so that compensating points are selected adaptively, if an erroneous observation was detected. In this way an outlier-free set of constraints respecting the overlapping requirement can be established. An effective implementation of this strategy remains for future research.

While an adaptive selection of points would allow for some images to remain available for the estimation of global translations, the system of linear equations might become large for well connected view-graphs, if all $\binom{n_{\mathbf{P}^l}}{3}$ triplet combinations are considered. Although the system of linear equations is sparse, which allows an efficient decomposition, for large image sets, it is sometimes useful to reduce the number of constraints. One way of doing this is to compute a minimum spanning tree for the corresponding sub-view-graph of every point in every point track and only take triplets from all pairwise combinations of edges. For the computation of the MST the covariance matrices of the relative translations could be used. This would reduce the number of triplet combinations to at least $n_{\mathbf{P}^l} - 2$ and at maximum $\left(n_{\mathbf{P}^l}\left(n_{\mathbf{P}^l} - 1\right)\right)/2$, which is considerably smaller than $\binom{n_{\mathbf{P}^l}}{3}$. The question how this reduction affects the estimation quality, in particular in the presence of outliers, is open for future research.

Another important issue was shown for the *Gendarmenmarkt* dataset. Although the selected starting image led to a correct result, a discrimination of a suitable starting image based on the minimal maximal distance in the view-graph often lacks sharpness. Thus, given the fact that the starting image may have a tremendous effect on the result and that a decision between possible candidates is rarely unique, the reliability of the proposed method is questionable. This could be highly improved by performing a sequence of individual breadth-propagations each using a different starting image. Because these do not depend on each other, this procedure can be parallelized and only leads to a minor additional computational effort. The individual sets of global rotations can then be averaged and situations, in which breadth-propagation leads to a considerably different result, can be identified.

Besides these important extensions, there are several additional possibilities to further enhance the estimation quality. For instance, breadth-propagation could also be assisted by using the covariance information of the relative rotations. If a clear decision cannot be taken, the individual covariance matrices could be consulted to infer a statistically more promising solution. In order to diminish the number of erroneous relative orientations, a linear estimation of the trifocal tensor could be considered [Ressl, 2000]. A derivation of pairwise relative orientations from the trifocal tensor in general is more robust. Moreover, various triplet combinations could be considered in order to derive several individual estimations of each pairwise relative orientation, which could then be averaged. This higher robustness come at the price of higher computational demand. Given the high robustness of the breadth-propagation, the actual benefit has to be evaluated.

In summary, the main objective of this thesis is met successfully. The results of the proposed global image orientation method provide a sound basis for a final bundle adjustment, delivering optimal parameters. Therefore, this method is a useful tool for the orientation of various types of images or image sequences. In order to enhance the estimation of translations, the two-step point track computation and outlier detection is to be replaced by a unified version that covers violated overlapping criteria and failures due to outliers. The dependency on the starting image can be covered by a parallel evaluation of multiple view-graphs so that the risk of degenerate solutions is minimized. Finally, attention should be focused on the estimation of suitable thresholds, which apply to the most common image sets.

Bibliography

Agarwal S, Snavely N, Simon I, Seitz S, Szeliski R (2009) Building Rome in a day. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*: 72 –79.

Arie-Nachimson M, Kovalsky SZ, Kemelmacher-Shlizerman I, Singer A, Basri R (2012) Global motion estimation from point matches. In: *Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*: 81–88.

Arrigoni F, Magri L, Rossi B, Fragneto P, Fusiello A (2014a) Robust absolute rotation estimation via low-rank and sparse matrix decomposition. In: *Proceedings of the International Conference on 3D Vision (3DV)*. 1: 491–498.

Arrigoni F, Rossib B, Malapellec F, Fragnetob P, Fusiello A (2014b) Robust Global Motion Estimation with Matrix Completion. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 40 (5): 63–70.

Arrigoni F, Fusiello A, Rossi B (2015a) Spectral Motion Synchronization in SE (3). *arXiv preprint arXiv:1506.08765*.

Arrigoni F, Rossi B, Fusiello A (2015b) Robust and Efficient Camera Motion Synchronization via Matrix Decomposition. In: *Proceedings of the International Conference on Image Analysis and Processing (ICIAP)* (pp. 444–455). Springer.

Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110 (3): 346–359.

Boyd S, Vandenberghe L (2004) *Convex Optimization*. Cambridge University Press.

Brown D (1971) Close-Range Camera Calibration. *Photogrammetric Engineering*, 37 (8): 855–866.

Brown M, Hua G, Winder S (2011) Discriminative learning of local image descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33 (1): 43–57.

Candès EJ, Tao T (2010) The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56 (5): 2053–2080.

Carlone L, Rosen DM, Calafiore G, Leonard JJ, Dellaert F (2015) Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*: 125–132.

Chatterjee A, Govindu VM (2013) Efficient and Robust Large-Scale Rotation Averaging. In: *Proceedings of the IEEE International Conf. on Computer Vision (ICCV)*: 521–528.

Chen L, Rottensteiner F, Heipke C (2016) INVARIANT DESCRIPTOR LEARNING USING A SIAMESE CONVOLUTIONAL NEURAL NETWORK. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 3 (3): 11–18.

Chen P, Suter D (2004) Recovering the missing components in a large noisy low-rank matrix: Application to SFM. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26 (8): 1051–1063.

Clausen J (1999) *Branch and bound algorithms-principles and examples*. Department of Computer Science, University of Copenhagen, Technical report.

Crandall D, Owens A, Snavely N, Huttenlocher D (2011) Discrete-continuous optimization for large-scale structure from motion. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 3001–3008.

Cremers D, Pock T, Kolev K, Chambolle A (2011) Convex Relaxation Techniques for Segmentation, Stereo and Multiview Reconstruction. In: *Markov Random Fields for Vision and Image Processing* (pp. 185–200). MIT Press.

Cui Z, Jiang N, Tang C, Tan P (2015) Linear Global Translation Estimation with Feature Tracks. In: Xianghua Xie MWJ, Tam GKL (eds) *Proceedings of the British Machine Vision Conference (BMVC)*: 46.1–46.13.

Cui Z, Tan P (2015) Global Structure-from-Motion by Similarity Averaging. In: *Proceedings of the IEEE International Conference on Computer Vision*: 864–872.

Davison AJ, Reid ID, Molton ND, Stasse O (2007) MonoSLAM: Real-Time Single Camera SLAM.

IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 29: 1052–1067.

Dennis Jr JE, Schnabel RB (1996) *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16. Society for Industrial and Applied Mathematics (SIAM).

Enqvist O, Jiang F, Kahl F (2011a) A brute-force algorithm for reconstructing a scene from two projections. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2961–2968.

Enqvist O, Kahl F, Olsson C (2011b) Non-sequential structure from motion. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*: 264–271.

Faugeras O (1992) What can be seen in three dimensions with an uncalibrated stereo rig? In: *Proceedings of the European Conference on Computer Vision (ECCV)*: 563–578.

Faugeras O (1993) *Three-dimensional computer vision: a geometric viewpoint*. MIT press.

Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the Association for Computing Machinery (CACM)*, 24 (6): 381–395.

Fitzgibbon AW, Zisserman A (1998) Automatic camera recovery for closed or open image sequences. In: *Proceedings of the European Conf. on Computer Vision (ECCV)* (pp. 311–326). Springer.

Floyd RW (1962) Algorithm 97: shortest path. *Communications of the Association for Computing Machinery (CACM)*, 5 (6): 345.

Förstner W (1986) A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry and Remote Sensing*, 26 (3): 150–166.

Förstner W, Wrobel B (2004) *Manual of Photogrammetry*, chapter 2, Mathematical concepts in photogrammetry, (pp. 15–180). American Society for Photogrammetry and Remote Sensing.

Förstner W, Wrobel B, Paderes F, Craig R, Fraser C, Dolloff J (2004) *Manual of Photogrammetry*, chapter 11, Analytical photogrammetric operations, (pp. 763–948). American Society for Photogrammetry and Remote Sensing.

Fusiello A, Crosilla F (2015) Solving bundle block adjustment by generalized anisotropic Procrustes analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 102: 209–221.

Gherardi R, Farenzena M, Fusiello A (2010) Improving the efficiency of hierarchical structure-and-motion. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1594–1600.

Gilmore R (1974) Baker-Campbell-Hausdorff formulas. *Journal of Mathematical Physics*, 15 (12): 2090–2092.

Gilmore R (2008) *Lie groups, physics, and geometry: an introduction for physicists, engineers and chemists*. Cambridge University Press.

Gould R (1988) *Graph Theory*. The Benjamin/Cummings Publishing Company.

Govindu VM (2001) Combining two-view constraints for motion estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2: II–218.

Govindu VM (2004) Lie-algebraic averaging for globally consistent motion estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1: I–684.

Govindu VM (2006) Robustness in motion averaging. In: *Computer Vision–ACCV 2006* (pp. 457–466). Springer.

Grisetti G, Kummerle R, Stachniss C, Burgard W (2010) A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2 (4): 31–43.

Grün A, Huang TS (2013) *Calibration and orientation of cameras in computer vision*, volume 34. Springer Science & Business Media.

Hampel FR (1968) *Contributions to the theory of robust estimation*. PhD thesis, University of California Berkeley.

Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (2011) *Robust statistics: the approach based on influence functions*, volume 114. John Wiley & Sons.

Harris C, Stephens M (1988) A combined corner and edge detector. In: Taylor CJ (ed) *Proceedings of the Alvey Vision Conference*. 15: 23.1–23.6.

Hartley R (1992) Estimation of relative camera positions for uncalibrated cameras. In: *Proceedings of the European Conference on Computer Vision (ECCV)*: 579–587.

Hartley R (1997) Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision (IJCV)*, 22 (2): 125–140.

Hartley R, Sturm P (1997) Triangulation. *Computer Vision and Image Understanding*, 68 (2): 146–157.

Hartley R, Aftab K, Trumpf J (2011) L1 rotation averaging using the Weiszfeld algorithm. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1: 3041–3048.

Hartley R, Zisserman A (2003) *Multiple view geometry in computer vision.* Cambridge university press.

Hartley R, Schaffalitzky F (2004) L_∞ minimization in geometric reconstruction problems. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1: I–504–I–509 Vol.1.

Hartley R, Kahl F (2007) Global optimization through searching rotation space and optimal estimation of the essential matrix. In: *Proceedings of the IEEE International Conf. on Computer Vision (ICCV)*: 1–8.

Hartley R, Trumpf J, Dai Y, Li H (2013) Rotation averaging. *International Journal of Computer Vision (IJCV)*, 103 (3): 267–305.

Havlena M, Torii A, Knopp J, Pajdla T (2009) Randomized structure from motion based on atomic 3D models from camera triplets. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2874–2881.

Horn BK (1990) Relative orientation. *International Journal of Computer Vision (IJCV)*, 4 (1): 59–78.

Horowitz MB, Matni N, Burdick JW (2014) Convex Relaxations of SE(2) and SE(3) for Visual Pose Estimation. In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).* abs/1401.3700: 1148–115.

Huber P et al. (1964) Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1): 73–101.

Jiang N, Cui Z, Tan P (2013) A Global Linear Method for Camera Pose Registration. In: *Proceedings of the IEEE International Conf. on Computer Vision (ICCV)*: 481–488.

Jiang N, Lin WY, Do MN, Lu J (2015) Direct Structure Estimation for 3D Reconstruction. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2655–2663.

Kahl F (2005) Multiple view geometry and the L_∞-norm. In: *Proceedings of the IEEE International Conf. on Computer Vision (ICCV).* 2: 1002–1009.

Ke Q, Kanade T (2007) Quasiconvex optimization for robust geometric reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29 (10): 1834–1847.

Konolige K, Agrawal M (2008) FrameSLAM: From bundle adjustment to real-time visual mapping. *IEEE Transactions on Robotics*, 24 (5): 1066–1077.

Krarup T, Juhl J, Kubik K (1980) Götterdämmerung over least squares adjustment. *International Archives of Photogrammetry*, 23: 369–378.

Kraus K (1993) *Photogrammetry: Volume 1, Fundamentals and Standard Processes*, volume 1. Dümmler Verlag, 4 edition.

Kraus K (1997) *Photogrammetry: Volume 2, Advanced Methods and Applications*, volume 2. Dümmler Verlag, 4 edition.

Läbe T, Förstner W (2006) Automatic relative orientation of images. In: *Proceedings of the 5th Turkish-German Joint Geodetic Days.* 29: 31.

Land AH, Doig AG (1960) An automatic method of solving discrete programming problems. *Econometrica: Journal of the Econometric Society*, : 497–520.

Levi N, Werman M (2003) The viewing graph. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1: I–518.

Löfberg J (2004) YALMIP: A toolbox for modeling and optimization in MATLAB. In: *Proceedings of the IEEE International Symposium on Computer Aided Control Systems Design*: 284–289.

Lowe D (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60 (2): 91–110.

Luong QT, Faugeras OD (1996) The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision (IJCV)*, 17 (1): 43–75.

Martinec D, Pajdla T (2007) Robust rotation and translation estimation in multiview reconstruction. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1–8.

Marzan GT, Karara HM (1975) A computer program for direct linear transformation solution of the collinearity condition, and some applications of it. In: *Proceedings of the symposium on close-range photogrammetric systems*: 420–476.

Mayer H (2014) Efficient hierarchical triplet merging for camera pose estimation. In: *Proceedings of the German Conference on Pattern Recognition (GCPR)*: 399–409.

Meidow J (2012) Effcient Multiple Loop Adjustment for Computer Vision Tasks. *Photogrammetrie-Fernerkundung-Geoinformation*, 2012 (5): 501–510.

Moore EF (1959) *The shortest path through a maze.* Bell Telephone System.

Moulon P, Monasse P, Marlet R et al. (2013) Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion. In: *Proceedings of the IEEE International Conf. on Computer Vision (ICCV)*

Mugnier CJ, Förstner W, Wrobel B, Paderes F, Riadh M (2004) *Manual of Photogrammetry*, chapter 3, The mathematics of photogrammetry, (pp. 181–316). American Society for Photogrammetry and Remote Sensing.

Ni K, Dellaert F (2012) HyperSfM. In: *Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*: 144–151.

Nistér D (2000) Reconstruction from uncalibrated sequences with a hierarchy of trifocal tensors. In: *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 649–663). Springer.

Nistér D (2004) An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 26 (6): 756–770.

Olson E (2009) Recognizing places using spectrally clustered local matches. *Robotics and Autonomous Systems*, 57 (12): 1157–1172.

Ozyesil O, Singer A (2015) Robust Camera Location Estimation by Convex Programming. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 2674–2683.

Ozyesil O, Singer A, Basri R (2015) Stable camera motion estimation using convex programming. *SIAM Journal on Imaging Sciences*, 8 (2): 1220–1262.

Pierrot-Deseilligny M, Paparoditis N (2006) A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 36 (1/W41).

Pierrot-Deseilligny M, Clery I (2011) Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 38 (5/W16): 269–276.

Pollefeys M (2007) *Handbook of mathematical models in computer vision*, chapter 24, (pp. 393–398). Springer Science & Business Media.

Pollefeys M, Van Gool L, Vergauwen M, Verbiest F, Cornelis K, Tops J, Koch R (2004) Visual modeling with a hand-held camera. *International Journal of Computer Vision (IJCV)*, 59 (3): 207–232.

Reich M, Heipke C (2014) A global Approach for Image Orientation using Lie algebraic Rotation Averaging and Convex L_∞ Minimization. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 40 (3): 265–272.

Reich M, Heipke C (2016) Convex image orientation from relative orientations. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 3 (3): 107–114.

Remondino F, El-Hakim S (2006) Image-based 3D Modelling: A Review. *The Photogrammetric Record*, 21 (115): 269–291.

Ressl C (2000) An Introduction To The Relative Orientation Using The Trifocal Tensor. In: *International Archives of Photogrammetry and Remote Sensing.* Amsterdam, 33 (B3): 769–776.

Rottensteiner F (2001) *Semi-automatic extraction of buildings based on hybrid adjustment using 3D surface models and management of building data in a TIS.* PhD thesis, Technische Universität Wien.

Sanyal R, Sottile F, Sturmfels B (2011) Orbitopes. *Mathematika*, 57 (02): 275–314.

Sattinger DH, Weaver OL (2013) *Lie groups and algebras with applications to physics, geometry, and mechanics*, volume 61. Springer Science & Business Media.

Saunderson J, Parrilo P, Willsky A (2014) Semidefinite relaxations for optimization problems over rotation matrices. In: *Proceedings of the IEEE Conference on Decision and Control (CDC)*: 160–166.

Saunderson J, Parrilo PA, Willsky AS (2015) Semidefinite Descriptions of the Convex Hull of Rotation Matrices. *SIAM Journal on Optimization*, 25 (3): 1314–1343.

Schwidefsky K, Ackermann F (1976) *Photogrammetrie.* BG Teubner.

Servatius B, Whiteley W (1999) Constraining plane configurations in computer-aided design: Combinatorics of directions and lengths. *SIAM Journal on Discrete Mathematics*, 12 (1): 136–153.

Shashua A (1995) Algebraic functions for recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17 (8): 779–789.

Sim K, Hartley R (2006) Removing Outliers Using the L_∞ Norm. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1: 485–494.

Simonyan K, Vedaldi A, Zisserman A (2012) Descriptor learning using convex optimisation. In: *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 243–256). Springer.

Sinha SN, Steedly D, Szeliski R (2010) A multi-stage linear approach to structure from motion. In: *Proceedings of the Trends and Topics in Computer Vision, ECCV workshops* (pp. 267–281). Springer.

Snavely N, Seitz SM, Szeliski R (2006) Photo Tourism: Exploring Photo Collections in 3D. In: *ACM Transactions on Graphics.* 25 (3): 835–846.

Snavely N, Seitz SM, Szeliski R (2008) Modeling the world from internet photo collections. *International Journal of Computer Vision (IJCV)*, 80 (2): 189–210.

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

Strecha C, von Hansen W, Van Gool L, Fua P, Thoennessen U (2008) On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1–8.

Thormählen T, Broszio H, Weissenfeld A (2004) Keyframe selection for camera motion and structure estimation from multiple views. In: *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 523–535). Springer.

Toh KC, Todd MJ, Tütüncü RH (1999) SDPT3a MATLAB software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11 (1-4): 545–581.

Torr PH, Zisserman A, Maybank SJ (1998) Robust detection of degenerate configurations while estimating the fundamental matrix. *Computer Vision and Image Understanding*, 71 (3): 312–333.

Torr PH, Zisserman A (2000) MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78 (1): 138–156.

Torr PH, (2002) Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision (IJCV)*, 50 (1): 35–61.

Triggs B, McLauchlan P, Hartley R, Fitzgibbon A (2000) Bundle adjustment–a modern synthesis. In: *Vision Algorithms: Theory and Practice, LNCS.* Springer Verlag, 1883: 298–375.

Trzcinski T, Christoudias M, Lepetit V (2015) Learning image descriptors with boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37 (3): 597–610.

Vandenberghe L, Boyd S (1996) Semidefinite programming. *SIAM review*, 38 (1): 49–95.

Warshall S (1962) A theorem on boolean matrices. *Journal of the Association for Computing Machinery (JACM)*, 9 (1): 11–12.

Weiszfeld E (1937) Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal*, 43 (2): 355–386.

Weiszfeld E, Plastria F (2009) On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research*, 167 (1): 7–41.

Wilson K, Snavely N (2014) Robust Global Translations with 1DSfM. In: *Proceedings of the European Conf. on Computer Vision (ECCV)* (pp. 61–75). Springer.

Wright M (2005) The interior-point revolution in optimization: history, recent developments, and lasting consequences. *Bulletin of the American Mathematical Society*, 42 (1): 39–56.

Wu C (2013) Towards linear-time incremental structure from motion. In: *Proceedings of the International Conference on 3D Vision (3DV)*: 127–134.

Yang J, Li H, Jia Y (2014) Optimal essential matrix estimation via inlier-set maximization. In: *Proceedings of the European Conf. on Computer Vision (ECCV)* (pp. 111–126). Springer.

Zach C, Klopschitz M, Pollefeys M (2010) Disambiguating visual relations using loop constraints. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 1426–1433.

List of Figures and Tables

Figures: All non-referenced figures have been composed by the author himself.

List of Figures

List of Tables

Appendices

A. Convex relaxation of $SO(3)$

In Section 4.3.1, a convex optimization problem was formulated to estimate global rotations from relative estimates. It is based on a parameterization of the convex hull of the rotation manifold $SO(3)$ as an LMI, which is derived in this appendix. The derivation is based on Horowitz et al. [2014] and Sanyal et al. [2011].

A common representation of rotations is given by unit quaternions $\mathbf{q}$, as described in Section 2.1.3. The mapping function between rotation matrices and quaternions is defined in (2.11). Written in the elements of the quaternions, this mapping is given by:

$$\mathbf{R} = \begin{bmatrix} s^2 + v_x^2 - v_y^2 - v_z^2 & 2v_xv_y - 2sv_z & 2v_xv_z + 2sv_y \\ 2v_yv_x + 2sv_z & s^2 - v_x^2 + v_y^2 - v_z^2 & 2v_yv_z - 2s_vx \\ 2v_xv_z - 2sv_y & 2v_yv_z + 2sv_x & s^2 - v_x^2 - v_y^2 + v_z^2 \end{bmatrix}. \tag{A.1}$$

Using the Gramian matrix

$$\mathbf{Q} = \mathbf{q}\mathbf{q}^T = \begin{bmatrix} s^2 & sv_x & sv_y & sv_z \\ v_xs & v_x^2 & v_xv_y & v_xv_z \\ v_ys & v_yv_x & v_y^2 & v_yv_z \\ v_zs & v_zv_x & v_zv_y & v_z^2 \end{bmatrix}, \tag{A.2}$$

(A.1) changes to:

$$\mathbf{R} = \begin{bmatrix} Q_{11} + Q_{22} - Q_{33} - Q_{44} & 2Q_{23} - 2Q_{14} & 2Q_{13} + 2Q_{24} \\ 2Q_{23} + 2Q_{14} & Q_{11} - Q_{22} + Q_{33} - Q_{44} & 2Q_{34} - 2Q_{12} \\ 2Q_{24} - 2Q_{13} & 2Q_{12} + 2Q_{34} & Q_{11} - Q_{22} - Q_{33} + Q_{44} \end{bmatrix}. \tag{A.3}$$

Matrix $\mathbf{Q}$ is positive semidefinite and of rank 1, because it is a Gramian matrix, and has a trace of 1, because it is composed of unit quaternions. Thus, any rotation in $SO(3)$ can be written in terms of the matrix $\mathbf{Q}$. The parameterization of $SO(3)$ given in (A.3) is known as the *Cayley transform*. According to Sanyal et al. [2011], this linear map $\mathbb{R}^{4\times4} \to \mathbb{R}^{3\times3}$ commutes with taking the convex hull.

The symmetric matrices $\mathbf{Q} \in \mathbb{R}^{4\times 4}$ with trace equal to one form a nine-dimensional affine space. Because of the symmetry, $\mathbf{Q}$ only has ten distinct elements, the trace constraint reduces the number of degrees of freedom by one. This nine-dimensional space is isomorphic to the space $\mathbb{R}^{3\times 3}$ under the linear map in (A.3).

By implication, if one writes the matrix $\mathbf{Q}$ by elements of $\mathbf{R}$, the formulation of the convex hull in Section 4.3.1 is derived. The reason for doing this reveals itself by taking the geometrical interpretation into account. Matrix $\mathbf{Q}$ is formed by a multiplication of unit quaternions, thus vectors in $\mathbb{R}^4$ of length one, which form the unit sphere $S^3 \subset \mathbb{R}^4$. Recalling the derivation in Section 4.3.1, the convex hull of $SO(2)$ is built by a relaxation of S^1, including all values inside the unit disk. In a similar fashion, the convex hull of $SO(3)$ is derived. First, a set of linear equations is constructed from (A.3):

$$\begin{aligned}
R_{11} &= Q_{11} + Q_{22} - Q_{33} - Q_{44} \\
R_{12} &= 2Q_{23} - 2Q_{14} \\
R_{13} &= 2Q_{13} + 2Q_{24} \\
&\dots \\
R_{33} &= Q_{11} - Q_{22} - Q_{33} + Q_{44}.
\end{aligned} \tag{A.4}$$

It comprises nine equations (one for each element in $\mathbf{R}$) and the ten unknowns $\{Q_{11}, \dots Q_{44}\}$, counting only the upper triangular part. Using the constraint that $Q_{11} + Q_{22} + Q_{33} + Q_{44} = 1$, the system becomes determinable. Now, matrix $\mathbf{Q}$ is written in terms of elements of $\mathbf{R}$, dropping the trace constraint and only requiring the matrix to be positive semidefinite, and the formulation of (4.20) is derived:

$$\begin{bmatrix}
1 + R_{11} + R_{22} + R_{33} & R_{32} - R_{23} & R_{13} - R_{31} & R_{21} - R_{12} \\
R_{32} - R_{23} & 1 + R_{11} - R_{22} - R_{33} & R_{21} + R_{12} & R_{13} + R_{31} \\
R_{13} - R_{31} & R_{21} + R_{12} & 1 - R_{11} + R_{22} - R_{33} & R_{32} + R_{23} \\
R_{21} - R_{12} & R_{13} + R_{31} & R_{32} + R_{23} & 1 - R_{11} - R_{22} + R_{33}
\end{bmatrix} \succeq 0. \tag{A.5}$$

B. Single rotation averaging

The goal of single rotation averaging is to find a rotation $\bar{\mathbf{R}}$ that is closest to an arbitrary number of noisy rotation estimates in the set R with respect to some norm. This comprises computing the mean in the rotation manifold $SO(3)$ (cf. Section 2.2). The optimization problem is formulated using the angular distance function (cf. Section 2.1.4).

$$\begin{aligned} &\text{minimize} && d\left(\bar{\mathbf{R}}\right) = \sum_{i=1}^{|\mathsf{R}|} d_\alpha \left(\mathbf{R}_i^T \bar{\mathbf{R}}\right)^p \\ &\text{subject to} && \bar{\mathbf{R}} \in SO(3). \end{aligned} \tag{B.1}$$

Using the logarithm and exponential map, this L_p-mean can be iteratively estimated in the tangent space and projected back to $SO(3)$ as proposed in Hartley et al. [2011, 2013]. In Reich & Heipke [2014], an extension is presented to compute a robust Huber-cost average [Huber et al., 1964]. In the course of this thesis a robust single rotation averaging is not necessary since, at any time, only two rotation estimates are averaged, which are weighted based on the covariance matrix of the relative rotations. A pseudo code for the averaging algorithm is shown in algorithm 1 for the case of quaternions.

Algorithm 1 Single rotation averaging

1: **initialization:** $\bar{\mathbf{q}} = \mathbf{q}_i$, weights $\mathbf{w}$, $\Omega = 1$, $\epsilon = 1e^{-5}$
2: **while** $\Omega > \epsilon$ **do**
3: compute conjugate mean $\bar{\mathbf{q}}^c$
4: multiply with conjugate mean $\mathbf{q}_i^c = \bar{\mathbf{q}}^c * \mathbf{q}_i$
5: get angular part $\mathbf{q}_i^\alpha = \mathbf{q}_i^c\,(2:4)$
6: compute weighted mean $\bar{\mathbf{q}}^\alpha = 1/|\mathsf{R}| \sum w_i \mathbf{q}_i^\alpha$
7: compute update $\bar{\mathbf{q}} = \bar{\mathbf{q}} * [\|\bar{\mathbf{q}}^\alpha\|; \bar{\mathbf{q}}^\alpha]$
8: $\Omega = \left(\bar{\mathbf{q}}^{\alpha,T} \bar{\mathbf{q}}^\alpha\right) / |\mathsf{R}|$

This algorithm can be easily adapted to meet an M-estimator. In this case the weights $\mathbf{w}$ are computed in every iteration based on the distance between the mean and the respective rotation estimate. Note that steps 5 and 7 of algorithm 1 comprise the analogues of the logarithm and exponential maps, respectively.

C. Linearization of $SO(3)$

In Section 2.2.1, it is shown that the mappings between the Lie group $SO(3)$ and its Lie algebra $\mathfrak{so}(3)$ is given by the exponential and the logarithm map. In the following, first the linearization of $SO(3)$ is derived analytically which leads to a valid basis for the Lie algebra $\mathfrak{so}(3)$. Then, a proof for the exponential and logarithmic map is given.

A well known parameterization of a rotation matrix is given by the Euler angles $\{\omega, \varphi, \kappa\}$ (in the order $X_\omega \to Y_\varphi \to Z_\kappa$):

$$\mathbf{R} = \begin{bmatrix} \cos\varphi\cos\kappa & -\cos\varphi\sin\kappa & \sin\varphi \\ \sin\omega\sin\varphi\cos\kappa + \cos\omega\sin\kappa & -\sin\omega\sin\varphi\sin\kappa + \cos\omega\cos\kappa & -\sin\omega\cos\varphi \\ -\cos\omega\sin\varphi\cos\kappa + \sin\omega\sin\kappa & \cos\omega\sin\varphi\sin\kappa + \sin\omega\cos\kappa & \cos\omega\cos\varphi \end{bmatrix} \tag{C.1}$$

For the linearization, the partial derivatives of (C.1) are needed.

$$\frac{\partial \mathbf{R}}{\partial \omega} = \begin{bmatrix} 0 & 0 & 0 \\ \cos\omega\sin\varphi\cos\kappa - \sin\omega\sin\kappa & -\cos\omega\sin\varphi\sin\kappa - \sin\omega\cos\kappa & -\cos\omega\cos\varphi \\ \sin\omega\sin\varphi\cos\kappa + \cos\omega\sin\kappa & -\sin\omega\sin\varphi\sin\kappa + \cos\omega\cos\kappa & -\sin\omega\cos\varphi \end{bmatrix}, \tag{C.2}$$

$$\frac{\partial \mathbf{R}}{\partial \varphi} = \begin{bmatrix} -\cos\kappa\sin\varphi & \sin\varphi\sin\kappa & \cos\varphi \\ \sin\omega\cos\varphi\cos\kappa & -\sin\omega\cos\varphi\sin\kappa & \sin\omega\sin\varphi \\ -\cos\omega\cos\varphi\cos\kappa & \cos\omega\cos\varphi\sin\kappa & -\cos\omega\sin\varphi \end{bmatrix}, \tag{C.3}$$

$$\frac{\partial \mathbf{R}}{\partial \kappa} = \begin{bmatrix} -\sin\kappa\cos\varphi & -\cos\varphi\cos\kappa & 0 \\ -\sin\omega\sin\varphi\sin\kappa + \cos\omega\cos\kappa & -\sin\omega\sin\varphi\cos\kappa - \cos\omega\sin\kappa & 0 \\ \cos\omega\sin\varphi\sin\kappa + \sin\omega\cos\kappa & \cos\omega\sin\varphi\cos\kappa - \sin\omega\sin\kappa & 0 \end{bmatrix}. \tag{C.4}$$

A linearization at the identity, $(\omega, \varphi, \kappa) = (0, 0, 0)$, reveals the 3×3 basis vectors in the Lie algebra:

$$\mathbf{L}_\omega = \left.\frac{\partial \mathbf{R}}{\partial \omega}\right|_{(\omega,\varphi,\kappa)=(0,0,0)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{C.5}$$

$$\mathbf{L}_\varphi = \left.\frac{\partial \mathbf{R}}{\partial \varphi}\right|_{(\omega,\varphi,\kappa)=(0,0,0)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \tag{C.6}$$

$$\mathbf{L}_\kappa = \left.\frac{\partial \mathbf{R}}{\partial \kappa}\right|_{(\omega,\varphi,\kappa)=(0,0,0)} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{C.7}$$

Using the Lie bracket $[A, B] = AB - BA$, a commutator associated to every Lie algebra, one can see the relation to the basis vectors in $\mathbb{R}^3$ under the cross product:

$$[\mathbf{L}_\omega, \mathbf{L}_\varphi] = \mathbf{L}_\kappa, \quad [\mathbf{L}_\kappa, \mathbf{L}_\omega] = \mathbf{L}_\varphi, \quad [\mathbf{L}_\varphi, \mathbf{L}_\kappa] = \mathbf{L}_\omega. \tag{C.8}$$

From the three basis vectors $(\mathbf{L}_\omega, \mathbf{L}_\varphi, \mathbf{L}_\kappa)$ every element in the Lie algebra can be constructed by a linear combination with an Euler vector $\mathbf{r}$ (i.e. the axis-angle representation, see Section 2.1.3):

$$\mathbf{r} * \mathbf{L} = r_x \mathbf{L}_\omega + r_y \mathbf{L}_\varphi + r_z \mathbf{L}_\kappa = \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix} \in \mathfrak{so}(3). \tag{C.9}$$

C.1. Proof of the exponential map

The exponential map $\exp(\cdot) : \mathfrak{so}(3) \rightarrow SO(3)$ is given in Equation (2.9) in Section 2.1.3. In the following, this mapping is proven.

Proof. In order to prove Equation (2.9), let us assume that some point $\mathbf{P}_k$ in $\mathbb{R}^3$ is rotated by a rotation $\mathbf{R}$, ${}^r\mathbf{P}_k = \mathbf{R}\mathbf{P}_k$. The rotated point ${}^r\mathbf{P}_k$ will lie in a plane perpendicular to the rotation axis $\bar{\mathbf{r}}$. This plane is described by two basis vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. $\mathbf{v}_1$ is defined to be orthogonal to the plane of $\bar{\mathbf{r}}$ and $\mathbf{P}_k$, thus $\mathbf{v}_1 = [\bar{\mathbf{r}}]_\times \mathbf{P}_k$, whereas $\mathbf{v}_2$ is orthogonal to $\bar{\mathbf{r}}$ and $\mathbf{v}_1$, $\mathbf{v}_2 = [\mathbf{v}_1]_\times \bar{\mathbf{r}}$. Regarding the offset of the plane to the origin of the coordinate system, the track of ${}^r\mathbf{P}_k$ in space can be described by:

$$\begin{aligned} {}^r\mathbf{P}_k &= \mathbf{P}_k - \mathbf{v}_2 + \sin(\alpha)\mathbf{v}_1 + \cos(\alpha)\mathbf{v}_2 \\ &= \mathbf{P}_k + \sin(\alpha)\mathbf{v}_1 + (\cos(\alpha) - 1)\mathbf{v}_2 \end{aligned} \tag{C.10}$$

Applying the basis vectors leads to

$$\begin{aligned} {}^r\mathbf{P}_k &= \mathbf{P}_k + \sin(\alpha)\,[\bar{\mathbf{r}}]_\times \mathbf{P}_k + (\cos(\alpha) - 1)\left[[\bar{\mathbf{r}}]_\times \mathbf{P}_k\right]_\times \bar{\mathbf{r}} \\ &= \mathbf{P}_k + \sin(\alpha)\,[\bar{\mathbf{r}}]_\times \mathbf{P}_k + (1 - \cos(\alpha))\,[\bar{\mathbf{r}}]_\times [\bar{\mathbf{r}}]_\times \mathbf{P}_k \\ &= \left(\mathbf{I}_{3\times3} + \sin(\alpha)\,[\bar{\mathbf{r}}]_\times + (1 - \cos(\alpha))\,[\bar{\mathbf{r}}]_\times^2\right)\mathbf{P}_k \end{aligned} \tag{C.11}$$

Now, the left term on the right side of Equation (C.11) can be substituted with the rotation matrix $\mathbf{R}$ which leads to Equation (2.9) and finishes the proof. □

C.2. Proof of the logarithm map

The logarithm map $\log(\cdot) : SO(3) \rightarrow \mathfrak{so}(3)$ is given in Equation (2.10) in Section 2.1.3. The proof of this map affiliates to the proof of the exponential map and is presented in the following.

Proof. Let us start from the equation of the exponential map (see Equation (2.9)). The squared skew-symmetric matrix $[\bar{\mathbf{r}}]_\times^2$ can be simplified:

$$[\bar{\mathbf{r}}]_\times^2 = \begin{bmatrix} -\bar{r_y}^2 - \bar{r_z}^2 & \bar{r_x}\bar{r_y} & \bar{r_x}\bar{r_z} \\ \bar{r_x}\bar{r_y} & -\bar{r_x}^2 - \bar{r_z}^2 & \bar{r_y}\bar{r_z} \\ \bar{r_x}\bar{r_z} & \bar{r_y}\bar{r_z} & -\bar{r_x}^2 - \bar{r_y}^2 \end{bmatrix} = \begin{bmatrix} \bar{r_x}\bar{r_x} - 1 & \bar{r_x}\bar{r_y} & \bar{r_x}\bar{r_z} \\ \bar{r_x}\bar{r_y} & \bar{r_y}\bar{r_y} - 1 & \bar{r_y}\bar{r_z} \\ \bar{r_x}\bar{r_z} & \bar{r_y}\bar{r_z} & \bar{r_z}\bar{r_z} - 1 \end{bmatrix} \tag{C.12}$$

Hence Equation (2.9) can be written as

$$\mathbf{R} = \cos(\alpha)\mathbf{I}_{3\times3} + \sin(\alpha)\,[\bar{\mathbf{r}}]_\times + (1-\cos(\alpha))\left([\bar{\mathbf{r}}]_\times^2 + \mathbf{I}_{3\times3}\right) \tag{C.13}$$

$$= \begin{bmatrix} \cos(\alpha) & 0 & 0 \\ 0 & \cos(\alpha) & 0 \\ 0 & 0 & \cos(\alpha) \end{bmatrix} + \sin(\alpha) \begin{bmatrix} 0 & -\bar{r_z} & \bar{r_y} \\ \bar{r_z} & 0 & -\bar{r_x} \\ -\bar{r_y} & \bar{r_x} & 0 \end{bmatrix} + (1-\cos(\alpha)) \begin{bmatrix} \bar{r_x}\bar{r_x} & \bar{r_x}\bar{r_y} & \bar{r_x}\bar{r_z} \\ \bar{r_x}\bar{r_y} & \bar{r_y}\bar{r_y} & \bar{r_y}\bar{r_z} \\ \bar{r_x}\bar{r_z} & \bar{r_y}\bar{r_z} & \bar{r_z}\bar{r_z} \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\alpha) + \bar{r_x}\bar{r_x}(1-\cos(\alpha)) & -\sin(\alpha)\bar{r_z} + \bar{r_x}\bar{r_y}(1-\cos(\alpha)) & \sin(\alpha)\bar{r_y} + \bar{r_x}\bar{r_z}(1-\cos(\alpha)) \\ \sin(\alpha)\bar{r_z} + \bar{r_x}\bar{r_y}(1-\cos(\alpha)) & \cos(\alpha) + \bar{r_y}\bar{r_y}(1-\cos(\alpha)) & -\sin(\alpha)\bar{r_x} + \bar{r_y}\bar{r_z}(1-\cos(\alpha)) \\ -\sin(\alpha)\bar{r_y} + \bar{r_x}\bar{r_z}(1-\cos(\alpha)) & \sin(\alpha)\bar{r_x} + \bar{r_y}\bar{r_z}(1-\cos(\alpha)) & \cos(\alpha) + \bar{r_z}\bar{r_z}(1-\cos(\alpha)) \end{bmatrix} \tag{C.14}$$

Regarding the symmetry in the second terms of the off-diagonal elements, the matrix $\mathbf{W}$ from Equation (2.9) can now be written as:

$$\mathbf{W} = \frac{\mathbf{R} - \mathbf{R}^T}{2} = \begin{bmatrix} 0 & -\sin(\alpha)\bar{r_z} & \sin(\alpha)\bar{r_y} \\ \sin(\alpha)\bar{r_z} & 0 & -\sin(\alpha)\bar{r_x} \\ -\sin(\alpha)\bar{r_y} & \sin(\alpha)\bar{r_x} & 0 \end{bmatrix} \tag{C.15}$$

Regarding first that the matrix 2-norm is equal to the maximum eigenvalue and second that every 3×3 skew-symmetric matrix comes with a pair of equivalent eigenvalues and a zero-eigenvalue, it can be shown that $\|\mathbf{W}\|_2 = \|\,[W_{32}, W_{13}, W_{21}]\,\|_2 = \sin(\alpha)$ (which is equivalent to the vector-norm of the elements of the upper or lower triangular part of $\mathbf{W}$). From this, Equation (2.9) follows directly for the case $\mathbf{R} \neq \mathbf{I}_{3\times3}$. □

C.3. Logarithm map for $\alpha = \pi$

In case the angle of rotation is equal to π, the logarithm map as defined in Equation (2.10) is not determined because it involves a division by 0. The matrix $\mathbf{W} = \mathbf{0}_{3\times3}$ can neither be used to determine the angle nor the axis. However, the angle can be computed from $\alpha = \arccos\left(\frac{\mathrm{tr}(\mathbf{R})-1}{2}\right)$ (see proof in C.4). For $\alpha = \pi$ the exponential map (Equation (2.9)) reduces to:

$$\mathbf{R} = \mathbf{I}_{3\times3} + 2\,[\bar{\mathbf{r}}]_\times^2 \tag{C.16}$$

From this, using (C.12), one can derive $[\bar{\mathbf{r}}]_\times^2 = \bar{\mathbf{r}} \otimes \bar{\mathbf{r}} - \mathbf{I}_{3\times3}$ and thus (C.16) changes to:

$$\mathbf{R} = \mathbf{I}_{3\times3} + 2\left(\bar{\mathbf{r}} \otimes \bar{\mathbf{r}} - \mathbf{I}_{3\times3}\right) = 2\bar{\mathbf{r}} \otimes \bar{\mathbf{r}} - \mathbf{I}_{3\times3} \tag{C.17}$$

$$\bar{\mathbf{r}} \otimes \bar{\mathbf{r}} = 1/2\left(\mathbf{R} + \mathbf{I}_{3\times3}\right). \tag{C.18}$$

The elements of $\bar{\mathbf{r}}$ can be derived taking the square root of the main diagonal elements. The sign ambiguity of the square root can be solved considering the signs of the off-diagonal entries (up to an overall scale because $\mathbf{r} = -\mathbf{r} \Leftrightarrow \|\mathbf{r}\| = \pi$).

C.4. Proof for $\alpha = \arccos\left(\frac{\mathrm{tr}(\mathbf{R})-1}{2}\right)$

In Section C.3, the computation of the axis-angle representation from a rotation matrix with rotation angle $\alpha = \pi$ is presented. In the following, it is shown that the angle $\alpha = \arccos\left(\frac{\mathrm{tr}(\mathbf{R})-1}{2}\right)$ is equivalent to the representation used in the logarithm map $\alpha = \arcsin\|\mathbf{W}\|_2$.

Proof.

$$\arccos\left(\frac{\mathrm{tr}\left(\mathbf{R}\right)-1}{2}\right) = \arcsin\left\|\frac{\mathbf{R}-\mathbf{R}^T}{2}\right\|_2 \tag{C.19}$$

$$\Leftrightarrow \quad 2\sin\left(\arccos\left(\frac{\mathrm{tr}\left(\mathbf{R}\right)-1}{2}\right)\right) = \left\|\mathbf{R}-\mathbf{R}^T\right\|_2$$

$$\Leftrightarrow \quad 2\sin\left(\arccos\left(\frac{R_{11}+R_{22}+R_{33}-1}{2}\right)\right) = \left\|\begin{bmatrix} 0 & R_{12}-R_{21} & R_{13}-R_{31} \\ R_{21}-R_{12} & 0 & R_{23}-R_{32} \\ R_{31}-R_{13} & R_{32}-R_{23} & 0 \end{bmatrix}\right\|_2 .$$

Using the fact that the matrix 2-norm is equal to the vectorial 2-norm of the off-diagonal elements it continues:

$$\left(2\sin\left(\arccos\frac{R_{11}+R_{22}+R_{33}-1}{2}\right)\right)^2 = \left(R_{32}-R_{23}\right)^2 + \left(R_{13}-R_{31}\right)^2 + \left(R_{21}-R_{12}\right)^2$$

$$\Leftrightarrow \quad 4 - 4\left(R_{11}/2 + R_{22}/2 + R_{33}/2 - 1/2\right)^2 - \left(R_{32}-R_{23}\right)^2 - \left(R_{13}-R_{31}\right)^2 - \left(R_{21}-R_{12}\right)^2 = 0$$

$$\Leftrightarrow \quad R_{11}^2 + R_{12}^2 + R_{13}^2 + R_{21}^2 + R_{22}^2 + R_{23}^2 + R_{31}^2 + R_{32}^2 + R_{33}^2 + \ldots \tag{C.20}$$

$$2\left(R_{11}R_{22} + R_{11}R_{33} + R_{22}R_{33} - R_{12}R_{21} - R_{13}R_{31} - R_{23}R_{32}\right) - \ldots \tag{C.21}$$

$$2\left(R_{11} + R_{22} + R_{33}\right)\ldots \tag{C.22}$$

$$= 3.$$

Equation (C.20) is the squared Frobenius norm of $\mathbf{R}$, $\|\mathbf{R}\|_F^2 = 3$. Thus, (C.21) and (C.22) must be equal. Ignoring the constant factor, (C.22) is just the trace of $\mathbf{R}$. Using the orthogonality of $\mathbf{R}$ (i.e. the three columns $[\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3]$ form a basis) (C.21) can be reformulated as the trace of $\mathbf{R}$ exploiting the vector cross product:

$$R_{11}R_{22} - R_{12}R_{21} = R_{33}, \quad R_{11}R_{33} - R_{13}R_{31} = R_{22}, \quad R_{22}R_{33} - R_{23}R_{32} = R_{11}. \tag{C.23}$$

This ends the proof. □

D. Computation of a quaternion from a rotation matrix

The computation of a quaternion from a given rotation matrix is an important step for two reasons. Firstly, calculus with quaternions instead of rotation matrices is often more efficient and, secondly, the mapping between the manifold and the Lie algebra is easier. The algorithm proposed here is not closed analytical as the inversion given in Equation (2.11) and involves two steps. The first one comprises the estimation of the absolute values of the quaternion using the fact, that the norm of the quaternion is equal to one. Starting from Equation (2.11) the rotation matrix can be written in terms of the quaternion elements:

$$\mathbf{R} = \begin{bmatrix} 1-2(v_y^2+v_z^2) & 2(v_xv_y - v_zs) & 2(v_xv_z + v_ys) \\ 2(v_xv_y + v_zs) & 1-2(v_x^2+v_z^2) & 2(v_yv_z - v_xs) \\ 2(v_xv_z - v_ys) & 2(v_yv_z + v_xs) & 1-2(v_x^2+v_y^2) \end{bmatrix}. \tag{D.1}$$

Now, the absolute value for each element in $\mathbf{q}$ is retrieved via permutation of the main diagonal entries of (D.1):

$$s = \frac{\sqrt{1+R_{11}+R_{22}+R_{33}}}{2} = \frac{\sqrt{4-4v_x^2-4v_y^2-4v_z^2}}{2} \tag{D.2}$$

$$v_x = \frac{\sqrt{1+R_{11}-R_{22}-R_{33}}}{2} = \frac{\sqrt{4v_x^2}}{2} \tag{D.3}$$

$$v_y = \frac{\sqrt{1-R_{11}+R_{22}-R_{33}}}{2} = \frac{\sqrt{4v_y^2}}{2} \tag{D.4}$$

$$v_z = \frac{\sqrt{1-R_{11}-R_{22}+R_{33}}}{2} = \frac{\sqrt{4v_z^2}}{2}. \tag{D.5}$$

Note that for a rotation by π, the term inside the square root of Equation (D.2) is equal to zero. In case of rounding issues, the square root might have an imaginary solution, which have to be taken care of.

Due to the sign ambiguity of the square root (i.e. the result of Equations (D.2)-(D.5) can either be positive or negative), a subsequent step is necessary. Because $\mathbf{q}$ and $-\mathbf{q}$ both represent the same rotation, it is possible to fix one element and determine the signs of the remaining elements with respect to that one. Without loss of generality the largest value $\max([s, v_x, v_y, v_z])$ is selected as is proposed in Horn [1990]. Let us assume $s = \max([s, v_x, v_y, v_z])$, then the signs of the remaining

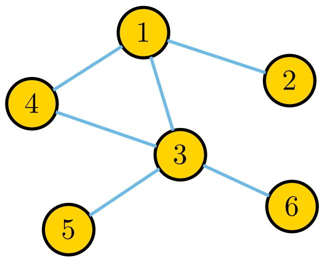


Figure E.1.: Example graph for the Floyd-Warshall algorithm.

elements are retrieved as:

$$v_x = \frac{R_{32} - R_{23}}{4s} = \frac{2(v_y v_z + v_x s) - 2(v_y v_z - v_x s)}{4s} \tag{D.6}$$

$$v_y = \frac{R_{13} - R_{31}}{4s} = \frac{2(v_x v_z + v_y s) - 2(v_x v_z - v_y s)}{4s} \tag{D.7}$$

$$v_z = \frac{R_{21} - R_{12}}{4s} = \frac{2(v_x v_y + v_z s) - 2(v_x v_y - v_z s)}{4s}. \tag{D.8}$$

This works likewise in case a different element is fixed. In summary, this also shows that the mapping from the rotation matrix to a quaternion is a one-to-two mapping (whereas the inverse mapping is one-to-one).

E. Starting vertex selection using the Floyd-Warshall algorithm

The starting vertex for the breadth-propagation algorithm is the vertex with the minimum distance to all other vertices. It is easily found in the distance matrix $\mathbf{D}$ that can be derived from the adjacency matrix $\mathbf{A}$ using the Floyd-Warshall algorithm which is given in pseudo-code in algorithm 2. In the following, this algorithm is explained based on a small example.

Algorithm 2 Floyd-Warshall algorithm

initialization: Adjacency matrix $\mathbf{A}$, distance matrix $\mathbf{D}^0 = \infty_{n \times n}$, iterator $k = 1$
set $\mathbf{A}'^0 = \mathbf{A}$
while $\mathbf{D}$ contains ∞ **do**
  $\{r, c\} \leftarrow \{\mathbf{A}' > 0 \cap \mathbf{D} = \infty\}$
  $\mathbf{D}^k_{r,c} \leftarrow k$
  $\mathbf{A}'^k = \mathbf{A}'^{k-1}\mathbf{A}$
  $k = k + 1$

Consider the example of a graph shown in Figure E.1. The symmetric adjacency matrix looks

$\mathbf{d}^1$	$\mathcal{V}_1$	$\mathcal{V}_2$	$\mathcal{V}_3$	$\mathcal{V}_4$	$\mathcal{V}_5$	$\mathcal{V}_6$
$\mathcal{V}_1$	∞	1	1	1	∞	∞
$\mathcal{V}_2$	1	∞	∞	∞	∞	∞
$\mathcal{V}_3$	1	∞	∞	1	1	1
$\mathcal{V}_4$	1	∞	1	∞	∞	∞
$\mathcal{V}_5$	∞	∞	1	∞	∞	∞
$\mathcal{V}_6$	∞	∞	1	∞	∞	∞

(a) $\mathbf{D}^1$

$\mathbf{d}^2$	$\mathcal{V}_1$	$\mathcal{V}_2$	$\mathcal{V}_3$	$\mathcal{V}_4$	$\mathcal{V}_5$	$\mathcal{V}_6$
$\mathcal{V}_1$	2	1	1	1	2	2
$\mathcal{V}_2$	1	2	2	2	∞	∞
$\mathcal{V}_3$	1	2	2	1	1	1
$\mathcal{V}_4$	1	2	1	2	2	2
$\mathcal{V}_5$	2	∞	1	2	2	2
$\mathcal{V}_6$	2	∞	1	2	2	2

(b) $\mathbf{D}^2$

$\mathbf{d}^3$	$\mathcal{V}_1$	$\mathcal{V}_2$	$\mathcal{V}_3$	$\mathcal{V}_4$	$\mathcal{V}_5$	$\mathcal{V}_6$
$\mathcal{V}_1$	2	1	1	1	2	2
$\mathcal{V}_2$	1	2	2	2	3	3
$\mathcal{V}_3$	1	2	2	1	1	1
$\mathcal{V}_4$	1	2	1	2	2	2
$\mathcal{V}_5$	2	3	1	2	2	2
$\mathcal{V}_6$	2	3	1	2	2	2

(c) $\mathbf{D}^3$

Table E.1.: Distance matrix $\mathbf{D}$ after first (E.1a), second (E.1b) and third iteration (E.1a).

like:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{E.1}$$

Then the distance matrix $\mathbf{D}$ is updated in every iteration. All the instances of $\mathbf{D}$ are shown in table E.1. In the first iteration, all indices that are nonzero in $\mathbf{A}$ are set to 1 in $\mathbf{D}$. After three iterations, the whole distance matrix $\mathbf{D}$ is computed. The matrices $\mathbf{A}'$ look like:

$$\mathbf{A}'^1 = \mathbf{A}'\mathbf{A} = \begin{bmatrix} 3 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 4 & 1 & 0 & 0 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad \mathbf{A}'^2 = \mathbf{A}'^1\mathbf{A} = \begin{bmatrix} 2 & 3 & 6 & 4 & 1 & 1 \\ 3 & 0 & 1 & 1 & 1 & 1 \\ 6 & 1 & 2 & 5 & 4 & 4 \\ 4 & 1 & 5 & 2 & 1 & 1 \\ 1 & 1 & 4 & 1 & 0 & 0 \\ 1 & 1 & 4 & 1 & 0 & 0 \end{bmatrix}. \tag{E.2}$$

Now, the rows or columns with the smallest maximal distance build the solution space to the optimization problem (4.8) in Section 4.2.2. Thus, $\{\mathcal{V}_1, \mathcal{V}_3, \mathcal{V}_4\}$ all have a maximum distance of 2. For this example $\mathcal{V}_3$ is taken as starting vertex because it has the maximum degree $|\mathcal{V}_3| = 4$.

F. Solving a linear homogeneous system of equations with SVD

The least-squares solution of a system of linear homogeneous equations like $\mathbf{Ct} = \mathbf{0} + \mathbf{v}$ with residual-vector $\mathbf{v}$ is found with a SVD. An optimization problem like this has to be solved for instance for the estimation of global translations, as defined in (4.49) in Section 4.4. This problem includes the constraint $\|\mathbf{t}\| = 1 \Leftrightarrow \mathbf{t}^T\mathbf{t} = 1$. In order to show that a least-squares-optimal solution is found through a SVD, the decomposition is plugged into the least-squares objective function: minimize $\mathbf{v}^T\mathbf{v}$:

$$\mathbf{v}^T\mathbf{v} = \mathbf{t}^T\mathbf{C}\mathbf{C}\mathbf{t} \tag{F.1}$$

$$\Leftrightarrow \quad = \mathbf{t}^T\mathbf{V}\mathbf{S}^T\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T\mathbf{t} \tag{F.2}$$

$$\Leftrightarrow \quad = \mathbf{t}^T\mathbf{V}\mathbf{S}^T\mathbf{S}\mathbf{V}^T\mathbf{t}, \text{ because } \mathbf{U}^T\mathbf{U} = \mathbf{I}. \tag{F.3}$$

Now, it is substituted $\mathbf{z} = \mathbf{V}^T\mathbf{t}$ and (F.3) reduces to:

$$\mathbf{v}^T\mathbf{v} = \mathbf{z}^T\mathbf{S}^T\mathbf{S}\mathbf{z}. \tag{F.4}$$

Applying the substitution to the constraint $\mathbf{t}^T\mathbf{t} = 1$ leads to:

$$\mathbf{t}^T\mathbf{t} = \mathbf{z}^T\mathbf{V}\mathbf{V}^T\mathbf{z} = \mathbf{z}^T\mathbf{z}, \text{ because } \mathbf{V}\mathbf{V}^T = \mathbf{I}. \tag{F.5}$$

Thus the problem can be written as:

$$\begin{aligned} &\underset{\mathbf{z}}{\text{minimize}} && \mathbf{z}^T\mathbf{S}^T\mathbf{S}\mathbf{z} \\ &\text{subject to} && \mathbf{z}^T\mathbf{z} = 1 \;. \end{aligned} \tag{F.6}$$

Because $\mathbf{S}$ is a diagonal matrix with singular values of $\mathbf{C}$ in decreasing order on the main diagonal, the optimal solution of problem (F.6) is given as $\mathbf{z}^\star = [0, 0, \ldots, 1]^T$. Redoing the substitution, the optimal solution for $\mathbf{t}$ is $\mathbf{t}^\star = \mathbf{Vz}$, thus the rightmost column of $\mathbf{V}$.

One has to take care of the rank defect of matrix $\mathbf{C}$. The desired solution in case of a rank defect of k would be the $n - k$th column of $\mathbf{V}$.

Index