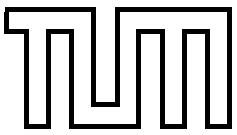
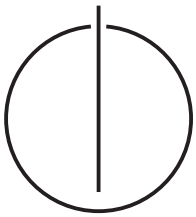


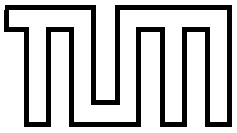
FAKULTÄT FÜR INFORMATIK
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Diplomarbeit in Informatik

Closed-Form Solutions to Computation of Multiple-View Homographies

Pierre Schroeder





FAKULTÄT FÜR INFORMATIK
DER TECHNISCHEN UNIVERSITÄT MÜNCHEN

Diplomarbeit in Informatik

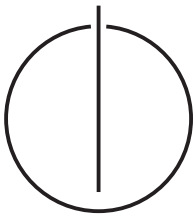
Closed-Form Solutions to Computation of Multiple-View Homographies

—

Lösungen in geschlossener Form zur Berechnung von Mehr-Bild-Homographien

Author: Pierre Schroeder
Supervisor: Prof. Dr. Nassir Navab
Advisors : Prof. Dr. Adrien Bartoli
Pierre Fite-Georgel, M.Sc.

Submission Date: 15th June 2010



Acknowledgements

At this point I'd like to accredit all those people supporting me during my studies and working on this thesis:

First of all, I'd like to express my deepest apologies and strongest acknowledgments to Carole and Svenja for the very hardest sacrifice they made in order to keep me going on chasing my aims in life.

Thankful credits to my family and particularly my parents for termlessly supporting me at any challenge I decided to face and for expressing their concerns – if ever – only as a matter of advice, but hardly ever in terms of restricting me.

Thanks to all my close friends for giving moral uplift to me during hard times and for providing much fun in-between. Thanks to Claudia, Jürgen, Jan, Rafael, and particularly Guy – whose friendship I stressed hard – for backing me in doing the trip related to this thesis and apologies for any inconvenience.

Thanks to Pierre for encouraging me to do the trip related to this work, to Adrien for being an obliging host. Thanks to both of them for their advisory and becoming inspirational friends over the past.

And last but not least, thanks to all those people I met during my journey to Clermont and which I left as a friend, and to those strangers helping me in some unpleasant situations in complete absence of any consideration but pure gratitude.

I assure the single handed composition of this diploma thesis only supported by declared resources

Contents

Abstract

Several fields of applications in computer vision require 2d-homography estimation based on point-correspondences between multiple images, which put into relation the 2d-projective transformations of each image in regard to a global reference coordinate frame. Due to noise in the measured point coordinates and eventual false matches on the point correspondences, it is impossible to find an exact solution for these homographies. Therefore – under certain assumptions about the type of noise – the aim is to find best approximates, notably Maximum Likelihood Estimators (MLE) for those homographies.

Unfortunately the task of determining those MLE involves, amongst other processing steps, the optimization of a non-linear cost function which requires initialization. The quality or accuracy of the initialization also affects the quality and accuracy of the final result of the optimization. Hence, improving the initialization technique is of significant interest when it comes to high-precision mosaicing.

So far, common initialization techniques either apply threading type methods to parallax-free scenarios or they apply batch type methods to sets of images with altering camera centers, although the concept of batch techniques allows the deviation of closed-form solutions. These closed-form solutions mainly differ from the threading type methods by the fact that they discard a whole level of detail from statistical point of view in order to linearize solving for an initial guess instead of heuristically withdrawing known and interpolating missing information.

This thesis introduces some of those closed-form solutions and investigates their behaviour in regard to a simple threading type solution with synthetic experiments and proves practical feasibility with a real example.

Chapter 1

Introduction

The task of 2D mosaicing, which is also called panoramic stitching, consists of aligning multiple images into one bigger image, the mosaic. In the centre of Figure 1.1 a sample mosaic is illustrated.

That may seem to be a quite trivial task because many user applications like Adobe Photoshop and several other, either expert or non-expert, digital image processing programs provide a tool for doing so out-of-the-box. But those tools are intended for a very special scenario, that is to say, for creating mosaics from a sequence or a set of images in which each overlapping region is only visible in two images. This is a special case which needs only two-view homography estimation, which can be realized with standard techniques and relatively low effort as it will be described later on in Section 2.2. Additionally, most of the tools only aim at making the resulting mosaic "look good" for the subjective viewer and do not even require projective correctness, but try to find a simple alignment (e.g. vertical translation) allowing to smoothly blend the transitions between the source images in the mosaic.

There are many other applications however, which have to cope with image sets and sequences containing from highly overlapping to sparely overlapping regions; that is to say, from regions which could be visible in many of the images to regions only visible in two or three images. As a consequence, these applications require a correct projective alignment of the images in order to work well or work at all. This projective alignment is denoted as homography. This work differentiates between *local* homographies – that is to say, homographies aligning one image to another image in the set – and *global* homographies – homographies that align images to the mosaic. In order to create a mosaic thus, it is mandatory to determine the global homographies; in the small example in Figure 1.1 latter are denoted $\mathtt{U}_1, \ldots, \mathtt{U}_6$.

Multiple images of a steady scene (with the additional restriction that either the scene must be planar or the projection centre stays the same) allow to create a mosaic, the resolution of which clearly exceeds the physical one of the capturing sensor [CZ98, Cap04, MFM04] by – in a manner of speaking – filling up the information missing in-between pixels in one image with information known from other images. These applications are generally denoted as super-resolution applications.

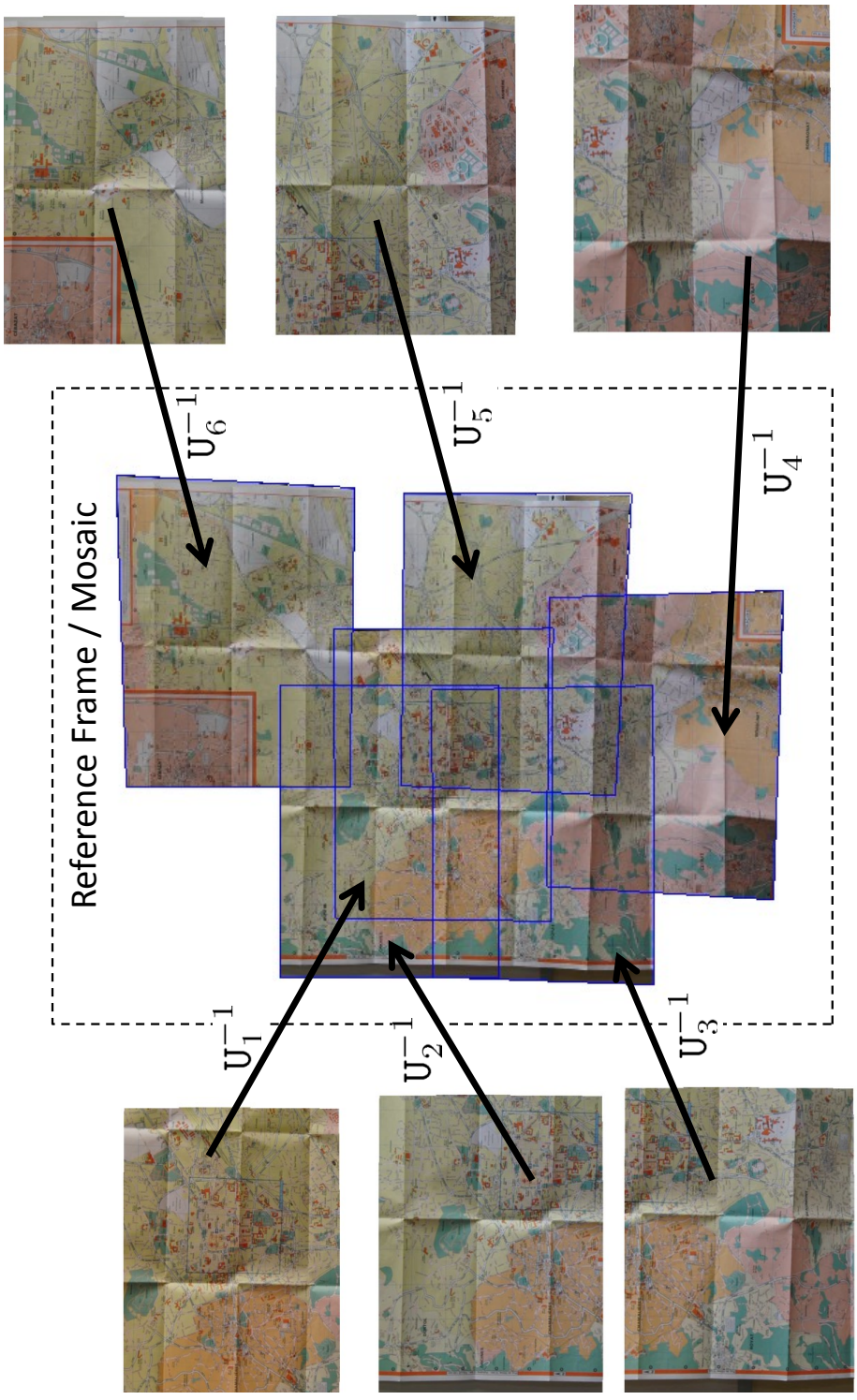


Figure 1.1: *Sample mosaic created from six partly overlapping pictures taken with a tripod-mounted camera and imaging parts of a city map of Clermont. The global homographies* $\mathtt{U}_1^{-1}, \ldots, \mathtt{U}_6^{-1}$ *describe the 2d-projective transformation required to fit the images into the mosaic.*

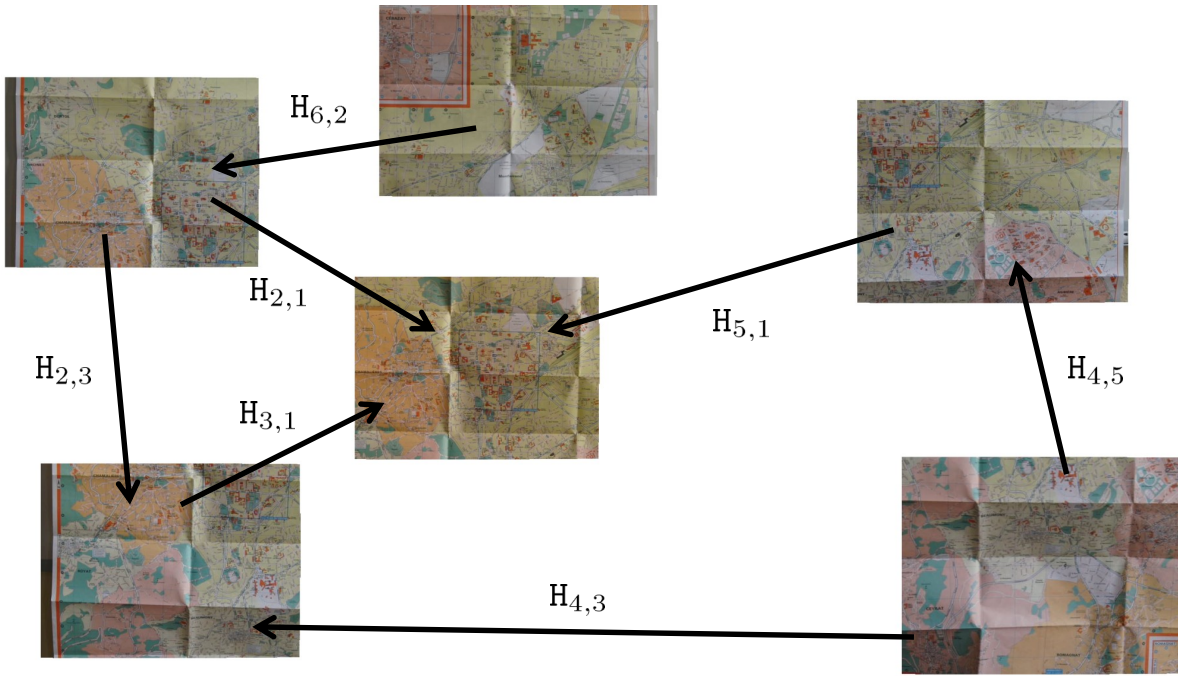


Figure 1.2: *Topology graph of the mosaic in Figure 1.1 illustrating which homographies could be determined in the image-set.*

Another example for an application would be video compression algorithms as for example the MPEG-7 codec [IHA95, IAB+96]. One of its techniques to reduce the size of the video file is to separate the steady background and the moving foreground from each other and keep a single mosaic of the background for a whole sequence and save only the dynamic foreground for each frame.

The technique of separating both, foreground and background, from each other, leads immediately to compositing in movie making and television applications. The ability to separate fore- and background in video sequences can avoid the usage of green-screen (or blue-screen) technology in several situations. Latter commonly suffer from the difficulty to uniformly illuminate the background screen behind the foreground scene, so that the screen is correctly recognized. For example, the weather forecast on television – probably the sample application for compositing in all-day-life per se – wouldn't require any chroma-screen technology anymore, as the situation is a fortunate constrained one (because the camera centers don't have to move).

Furthermore, camera self-calibration techniques require multiple view homography estimation [Har97, MC02]. These do not require to literally create mosaics, but as this work specifically aims on homography estimation, camera calibration may be mentioned as application example too.

Whereas there exists an exact solution for homography estimation in the theoretical and ideal case, in reality and under presence of noise, the best estimation one can get is the Maximum Likelihood Estimators (MLE) for the global homographies. Unfortunately, the estimation of those requires the optimization of a non-linear cost, and by consequence, this

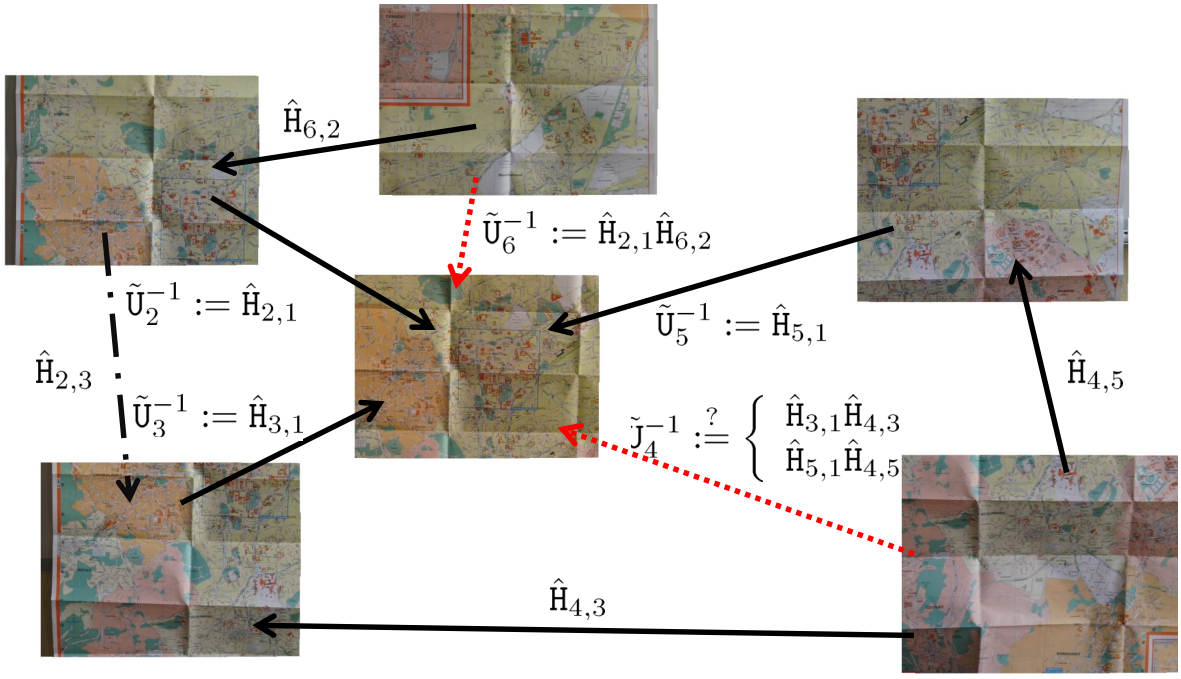


Figure 1.3: *Graph from Figure 1.2 modified to show how one could initialize the global homographies by using the MLE of the known local ones. The drawbacks are that* $\tilde{U}_6^{-1}$ *would have to be interpolated by* $\hat{H}_{2,1}$ *and* $\hat{H}_{6,2}$ *and the estimation error of both would be accumulated by* $\tilde{U}_6^{-1}$*; and* $\tilde{U}_4^{-1}$ *raises the question whether to use the detour over image* 3 *or image* 5*, or both if scaling issues could be handled. Furthermore, known information contained in* $\hat{H}_{2,3}$ *would be discarded.*

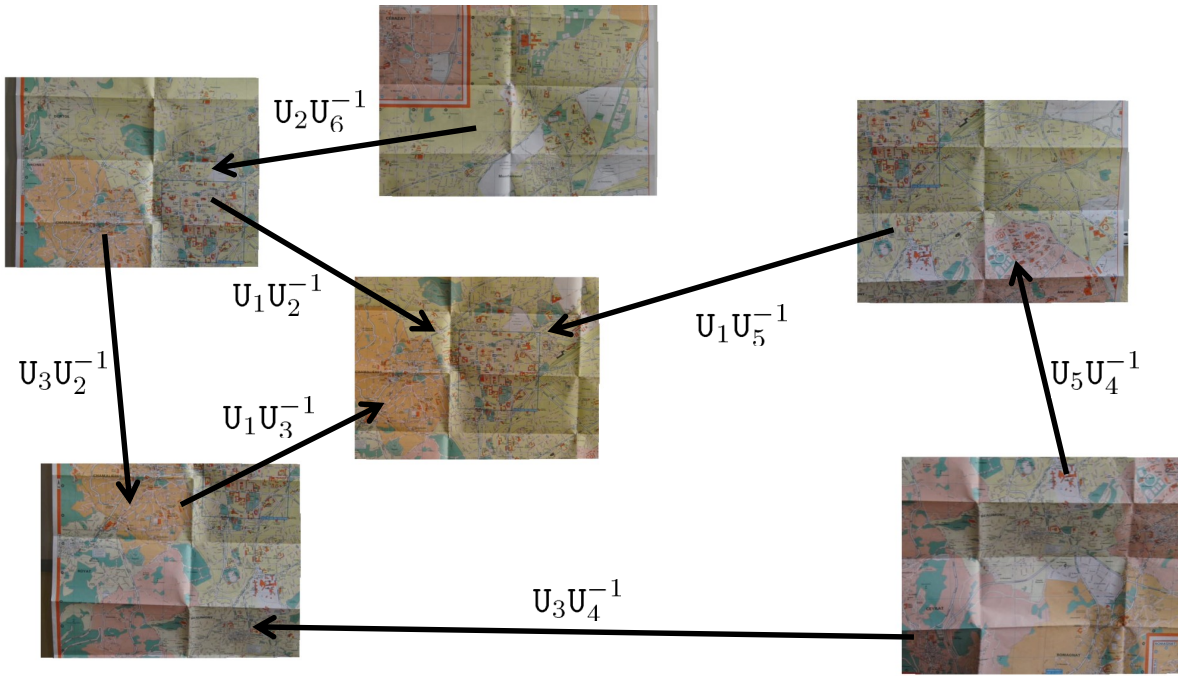


Figure 1.4: *Graph from Figure 1.2 modified to visualize in which way the global homographies* $\mathtt{U}_1^{-1}, \ldots, \mathtt{U}_6^{-1}$ *are contained (up to scale) in the determinable homographies.*

optimization, called Bundle Adjustment (BA), requires a reasonable initialization, because the quality of the estimated homographies depends on the quality of the initialization, as it will be explained later in Section 2.1. As the immediate determination of initial guesses for the global homographies based on point correspondences would most likely consist of an extremely complex system (if feasable at all) and seemingly hasn't been found so far, it is very common to base the initial guessing on the known local homographies. Latter can be estimated by using standard methods of reasonable complexity and resource usage. Already a multiple-view robust matching system would result in a considerably complex system and hence the robust matching is done with pairwise image-to-image matching and yields the local homographies anyways. Figure 1.2 illustrates the local homographies $\mathtt{H}_{i,j}$ aligning the images in the set pairwise wherever possible according to the sample mosaic from Figure 1.1.

A straight forward solution to initializing the BA with the local homographies is the one denoted and used by Capel in [Cap04]. In a video stream, first the homographies of consequent frames are computed and, based on a topology estimation relying on those immediate homographies, in a second step, all the local homographies for image pairs with common overlapping region are estimated regardless their distance in the time-line. A reference image is chosen (often simply the first one) and if a local homography to that reference image is known, it is used as initialization for the BA. An example for that case is given in illustration 1.3 where the initializers $\tilde{\mathtt{U}}_2^{-1}$, $\tilde{\mathtt{U}}_3^{-1}$, and $\tilde{\mathtt{U}}_5^{-1}$ are assigned the local homographies MLE $\hat{\mathtt{H}}_{2,1}$, $\hat{\mathtt{H}}_{3,1}$, and $\hat{\mathtt{H}}_{5,1}$ respectively.

For those frames for which no local homography to the reference image is known, the

shortest path through – or, in other words – the shortest composition of local homographies leading to the reference image is taken as initialization. In the example in Figure 1.3 this is illustrated with the help of $\tilde{\mathtt{U}}_6^{-1}$. Enchaining alignments this way is theoretically completely valid but has, in reality, the drawback that the estimation errors in the used local homographies, arising from noisy point coordinates and false matches in the point correspondences, are accumulated by the initializer [SHK98].

Furthermore, it might be possible to find multiple shortest paths as for example for $\tilde{\mathtt{U}}_4^{-1}$ which could be approximated either by composing $\hat{\mathtt{H}}_{4,5}$ and $\hat{\mathtt{H}}_{5,1}$, or by composing $\hat{\mathtt{H}}_{4,3}$ and $\hat{\mathtt{H}}_{3,1}$ and one has to decide which path to choose. Actually, thinking about that situation, one even has to figure out what could be a reasonable criterion for "shortest path", because a path composing three really good subsequent alignments could be significantly better than the composition of two inaccurate homographies.

As will be pointed out later in Section 3.1, some work has been done in the past, trying to solve the problem as a searching problem by considering images as nodes in a graph and homographies as edges, connecting the images among one another (illustrated in Figure 1.2). The arising solutions will be referred to as "threading solutions" herein. They try to find reasonable weightings for the edges in the connection graph, which reflects the quality or reliability of the local homographies they represent [MFM04]; they try to find reasonable criteria for weighting paths according to their edge weightings, as their sum or product would most likely not make too much sense [KCM00]; and they try to alter the standard graph searching algorithms, so that the path or edge weightings are being validated by alternative paths in order to detect outliers [VLW08, VLMW07, BD08]. But finally, the threading solutions will always stay somewhat heuristic ones and to some extent ignore available information; in the example in Figure 1.3 a simple solution as the one used in [Cap04] would probably not have considered the local homography $\hat{\mathtt{H}}_{2,3}$ as source of information although it contains valuable information because the local homographies already contain the global alignments (up to scale and in terms of a common arbitrary reference frame) as Figure 1.4 illustrates.

Thus, it could probably be of benefit to have a non-heuristic method for finding an initialization which does not propagate errors throughout solutions as threading type methods do, can handle missing data, and which uses a maximum of the redundantly available information. In contrast to the threading type solutions the class of batch type solutions approaches the issue as desired. Generally spoken, those solutions try to somehow reformulate a deviant cost function which is either convex or linearly solvable and, the evaluation of which is, at best, significantly cheaper than the real cost function. This can be achieved either by optimizing an approximation of the original cost, the minimum of which is at the same location as is the minimum of the original cost in the noise free case, as does for example the Direct Linear Transform for the initialization of two-view homography estimation as will be seen in Section 2.2.2; or by altering the level of detail of input data from statistical point of view, in this case for example trying at first to find global homographies which best fit the estimated local homographies instead of trying to find global homographies which best fit the point correspondences immediately. For the constrained case in which all the images in the set have the same centre of projection though, it seems that there hasn't been done any work so far which follows a batch type strategy.

After pointing out more detailed the required background in Chapter 2 and giving an overview over related work regarding this topic in Chapter 3, closed-form solutions following the batch type strategy will be introduced in Chapter 4. Then they are evaluated and compared against a standard threading solution with the help of synthetic experiments and a real example as described in Chapter 5 and 6 respectively. Finally, Chapter 7 contains a conclusion together with an outlook about potentially future work.

Chapter 2

Theoretical Background

In this chapter the theoretical background for homography estimation, as used herein, will be elucidated and the notation used in this document will be pointed out in order to avoid misunderstandings arising from notation issues. At this point, the reader is portended, that the contents of Sections 2.2 and 2.3 can be found in [HZ04], but these sections summarize the essentials for this work and are included, on the one hand, for the sake of completeness and, on the other hand, for improving the readability of this document.

Notation All throughout this document a consistent math notation will be used which visualizes to some extent of which object or type of object a certain token stands for. Furthermore it will be supposed, that this notation sufficiently indicates the nature of the instances it describes, and for some newly introduced instances the type of object won't explicitly be pointed out.

Ordinary scalars and functions are written in italics: x, y, $f(x)$.

Vectors and points will be noted with bold letters. 3D points and vectors with capitals: $\mathbf{Q}$, and homogeneous 2D coordinates with lower-case letters: $\mathbf{q}$.

Matrices will be written with typewriter captials: $\mathtt{M}$.

Images are referred to as $\mathcal{I}_i$, with i the identifier of the image.

Homographies are written as 3×3 matrices $\mathtt{H}_{i,j}$, for local homographies between image $\mathcal{I}_i$ and $\mathcal{I}_j$; and $\mathtt{U}_i$, for global homographies between the reference frame and $\mathcal{I}_i$.

Block matrices are noted with calligraphic capitals: $\mathcal{M}$.

Cost functions are noted with a calligraphic capital "C", usually with an index: $\mathcal{C}_{\mathtt{X}}$.

Intervals of natural numbers without 0 as $\{1, \ldots, n\}$ are referred to with the short notation $[n]$.

Estimators and initializers are indicated by $\hat{\cdot}$ for Maximum Likelihood Estimators and $\tilde{\cdot}$ for initializers for optimization algorithms.

2.1 Linear vs. Non-Linear Optimization Problems

Optimization problems intend to find an exact or the most likely parameterization of one or multiple objects or object-states which are part of a model which reflects a more or less complex system. Depending on the complexity of the system, the knowledge about it, or available computing resources, the model is based on assumptions and approximations and puts the object-states into mathematical relations. Those relations allow the deviation of cost functions which reflect the probability of the model being in a certain state. By consequence, one can figure out the most likely object-states of unknown ones regarding observations made on other object-states by finding a global extremum (maximum or minimum depending on the formulation) of the cost.

Depending on the complexity and on the nature of the problem statement, the optimization problem will be of a particular class of optimization problems. Each of those classes has its own particular properties which will cause either advantages or disadvantages for solving a problem belonging to that class and will either allow to solve the problem with less effort or require explicit handling of awkward issues in order to get a sub-optimal solution at all. In the following, only those two classes required for this work - in particular linear and non-linear optimization problems - are mentioned:

Linear Optimization Problems are the most obliging ones. Problems of this class can precisely be solved in a straight forward manner with relatively low computational effort.

Non-Linear Optimization Problems have some unfavourable properties causing some difficulties which can partly be suppressed if handled with care, but which cannot be avoided to full extent. Problems belonging to this class cannot be solved analytically and consequently have to be solved by iteratively searching the extremum in a try-and-error manner. Whereas this property is already undesirable by the fact that an exhaustive search on the domain space becomes unfeasible already for low-dimensional domains, it is additionally unfavourable because for large scale problems the evaluation of the cost may be relatively time consuming. Hence the number of iterations should be kept as low as possible.

As it is very common that non-linear optimization problems do not converge to any solution if the algorithm starts searching at a point sufficiently far away from the solution, it is mandatory to have a means for determining an initial approximation for the algorithm in order to make it converge to a solution at all [NW99].

Another issue results from the several local minima a non-linear cost function may possess: Optimization algorithms will iterate into a local minimum next to the initialization point although it is very unlikely that the found minimum is the global minimum of the cost. That is to say, it is worth finding an initialization as close as possible to the real solution and not only finding an initialization which makes the optimization algorithm converge.

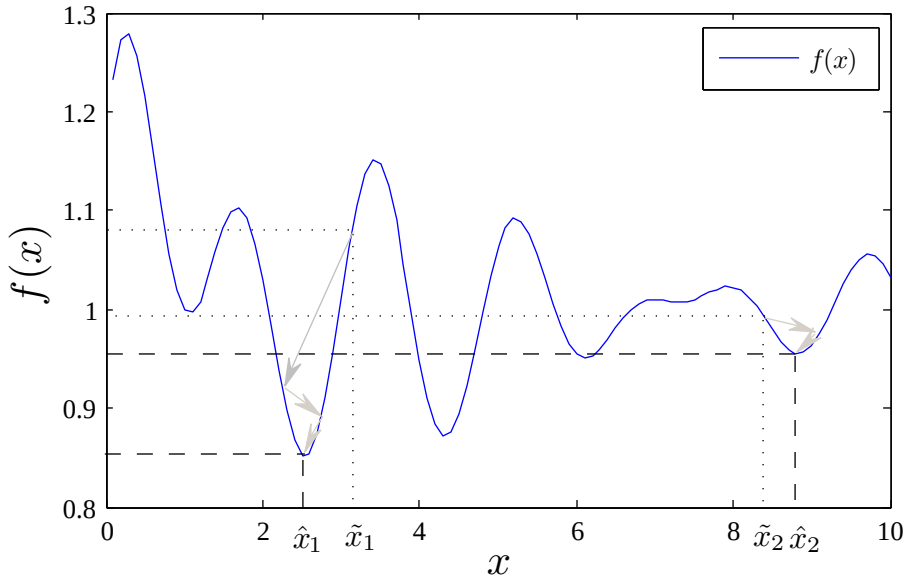


Figure 2.1: *An arbitrary one-dimensional non-linear cost function $f(x)$ is given and is to be minimized. Many local minima of the cost function will cause the optimization algorithm to run into a local minimum ($\hat{x}_1$ reps. $\hat{x}_2$) close to the initialization point ($\tilde{x}_1$ reps. $\tilde{x}_2$).*

2.2 Two-View Homographies

Two-view homographies describe the geometric relation between point coordinates in the coordinate frame of a projective plane and the point coordinates in the coordinate frame of another projective plane. Hence, they allow to warp one image so that it aligns appropriately with another image which partly pictures the same scene. This mapping is required in order to align two images up to a single mosaic.

As this thesis treats the special case in which the camera centre stays the same, this constraint might be applied to the model used herein. The restriction that the camera centre does not vary from image to image allows dropping the translational componend of the camera motion completely and drastically simplifies the model. Under these circumstances a projective camera projects a point $\mathbf{Q} \in \mathbb{R}^3$ in space to a point $\mathbf{q}^i \in \mathbb{P}^2$ (hence, usage of homogeneous coordinates are implicated) in the image plane of image $\mathcal{I}_i$ as:

$$\mathbf{q}^i \sim \mathtt{K}\mathtt{R}_i\mathbf{Q}, \tag{2.1}$$

with $\mathtt{K}$ the matrix of intrinsic parameters and $\mathtt{R}_i$ the rotation matrix specifying the camera orientation. (As the camera centre will not change, we assume that the origin coincides with the centre of projection.)

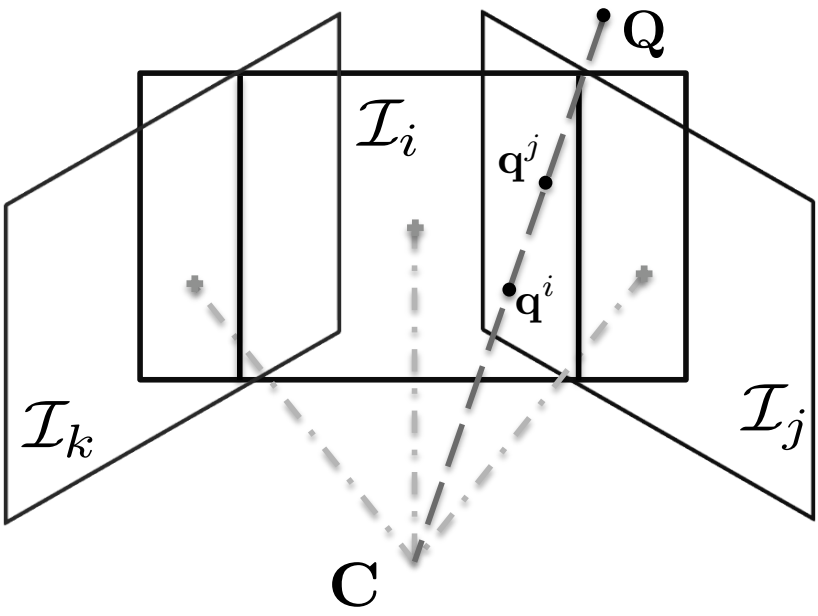


Figure 2.2: *Sweep of images with a static center of projection* $\mathbf{C}$ *(no parallax): the real 3D point* $\mathbf{Q}$ *projects to* $\mathbf{q}^i$ *in image* $\mathcal{I}_i$ *and its corresponding point* $\mathbf{q}^j$ *in* $\mathcal{I}_j$*, but is not visible in* $\mathcal{I}_k$*.*

With the projection being reversed[1]

$$(2.1) \Leftrightarrow \qquad \mathbf{Q} \sim \mathtt{R}_i^\mathsf{T}\mathtt{K}^{-1}\mathbf{q}^i, \tag{2.2}$$

it follows that a point $\mathbf{q}^i$ in image $\mathcal{I}_i$ is related to its corresponding point $\mathbf{q}^j$ in image $\mathcal{I}_j$ by:

$$(2.2) \text{ in } (2.1): \qquad \mathbf{q}^j \sim \mathtt{K}\mathtt{R}_j\mathtt{R}_i^\mathsf{T}\mathtt{K}^{-1}\mathbf{q}^i. \tag{2.3}$$

This is illustrated in Figure 2.2. Thus, an inter-frame homography is given by

$$\mathtt{H}_{i,j} \sim \mathtt{P}_j\mathtt{P}_i^{-1}, \ \forall k : \mathtt{P}_k \sim \mathtt{K}\mathtt{R}_k \tag{2.4}$$

from which we derive the *consistency relationship*

$$\mathtt{H}_{i,j} \sim \mathtt{P}_j\mathtt{P}_k^{-1}\mathtt{P}_k\mathtt{P}_i^{-1} \sim \mathtt{H}_{k,j}\mathtt{H}_{i,k}. \tag{2.5}$$

In theory, for m points visible in $\mathcal{I}_i$ and their m corresponding points visible in $\mathcal{I}_j$ it follows

$$\forall z \in [m] : \mathbf{q}_z^j \sim \mathtt{H}_{i,j}\mathbf{q}_z^i. \tag{2.6}$$

(Of course, if $\mathcal{I}_i$ and $\mathcal{I}_j$ overlap, there is an infinity of corresponding points, but the focus herein is laid onto practical applications, and so only a subset of points which can be identified as key points will be considered.)

[1]Remember that Equation 2.2 does not have to return the original point in 3D coordinates, but it does absolutely suffice to return a point which – thinking of it as a vector – points to the same direction as the original 3D point.

2.2.1 Maximum Likelihood Estimator

In reality however, the points' coordinates are not exact and it will be impossible to comply with Equation 2.6 whilst solving for $\mathtt{H}_{i,j}$. Obviously, it is desirable to find the Maximum Likelihood Estimator (MLE) $\hat{\mathtt{H}}_{i,j}$ instead. Assuming that the points in both images suffer from Gaussian noise, the MLE $\hat{\mathtt{H}}_{i,j}$ minimizes the reprojection error

$$\mathcal{C}_{\mathsf{RP2}} = \sum_{z=1}^{m} d^2(\mathbf{q}_z^i, \hat{\mathbf{q}}_z^i) + d^2(\mathbf{q}_z^j, \hat{\mathtt{H}}_{i,j}\hat{\mathbf{q}}_z^i) \tag{2.7}$$

as Hartley and Zisserman do explain in [HZ04]. d is the Euclidean distance of two points of their non-homogeneous coordinates

$$d\left(\begin{bmatrix} x \\ y \\ w \end{bmatrix}, \begin{bmatrix} x' \\ y' \\ w' \end{bmatrix}\right) = \sqrt{\left(\frac{x}{w} - \frac{x'}{w'}\right)^2 + \left(\frac{y}{w} - \frac{y'}{w'}\right)^2}. \tag{2.8}$$

However, $\mathcal{C}_{\mathsf{RP2}}$ is a non-linear cost and finding it's minimum thus represents an optimization problem. As a consequence, minimizing it in terms of $\hat{\mathbf{q}}_1^i, \ldots, \hat{\mathbf{q}}_m^i$, and $\hat{\mathtt{H}}_{i,j}$ requires a reasonable initialization as explained in the preceding Section 2.1.

2.2.2 Direct Linear Transformation (DLT)

In order to determine initialization parameters for optimizing $\mathcal{C}_{\mathsf{RP2}}$, Hartley and Zisserman suggest to use their Direct Linear Transformation (DLT) algorithm which approximates the non-linear cost with a linear one. Although that modified cost won't allow to find the MLE, it can be optimized very efficiently and is numerically quite stable, if handled with care.

As the non-linearity of $\mathcal{C}_{\mathsf{RP2}}$ arises from the homogeneous coordinates which imply the equality only up to scale in 2.6, the DLT aims at dropping that scaling factor by only focusing on the direction of the vectors represented by the homogeneous points. If that direction of two points $\mathbf{q}_z'$ and $\mathtt{H}\mathbf{q}_z$ is the same, their cross product is $\mathbf{0}$. If $\mathbf{q}_z' = [x_z' \; y_z' \; w_z']^\mathsf{T}$ and $\mathbf{h}_1^\mathsf{T}, \ldots, \mathbf{h}_3^\mathsf{T}$ denote the first to third row of $\mathtt{H}$, then

$$\forall z \in [m]: \quad \begin{bmatrix} \mathbf{0}^\mathsf{T} & -w_z'\mathbf{x}_z^\mathsf{T} & y_z'\mathbf{x}_z^\mathsf{T} \\ w_z'\mathbf{x}_z^\mathsf{T} & \mathbf{0}^\mathsf{T} & -x_z'\mathbf{x}_z^\mathsf{T} \\ -y_z'\mathbf{x}_z^\mathsf{T} & x_z'\mathbf{x}_z^\mathsf{T} & \mathbf{0}^\mathsf{T} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix} = \mathbf{0} \tag{2.9}$$

is equivalent to $\mathbf{q}_z' \times \mathtt{H}\mathbf{q}_z = \mathbf{0}$. Since only two of the three equations represented by Equation 2.9 are linearly independent, the third one can be dropped, which yields

$$\forall z \in [m]: \quad \underbrace{\begin{bmatrix} \mathbf{0}^\mathsf{T} & -w_z'\mathbf{x}_z^\mathsf{T} & y_z'\mathbf{x}_z^\mathsf{T} \\ w_z'\mathbf{x}_z^\mathsf{T} & \mathbf{0}^\mathsf{T} & -x_z'\mathbf{x}_z^\mathsf{T} \end{bmatrix}}_{\mathtt{A}_z} \underbrace{\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix}}_{\mathbf{h}} = \mathbf{0} \tag{2.10}$$

$$\Leftrightarrow \underbrace{\begin{bmatrix} \mathtt{A}_1 \\ \vdots \\ \mathtt{A}_m \end{bmatrix}}_{\mathtt{A}} \mathbf{h} = \mathbf{0}. \tag{2.11}$$

In order to solve Equation 2.11 for a non-trivial solution (which is obviously what the target of the DLT is about), a set of at least four point correspondences is required [HZ04] in which there are no three co-linear ones. If there are more point correspondences available, the system is overdetermined and under presence of noise, no exact solution can be found. Therefore an approximated solution

$$\underset{\hat{\mathbf{h}}|\hat{\mathbf{h}}^\top\hat{\mathbf{h}}=1}{\arg\min}\|\mathtt{A}\hat{\mathbf{h}}\|^2 \tag{2.12}$$

for Equation 2.11 is used, which can quickly be found by performing an SVD on $\mathtt{A}$ (see Appendix B).

As the point coordinates in real images will in general yield values ranging from 10^0 to 10^5 for the elements of $\mathtt{A}$ [HZ04] the problem would be bad conditioned without any further treatment. Therefore it is absolutely mandatory to normalize the point coordinates image-wise, so that the centroid of the points is at the origin and the mean distance to the centroid is $\sqrt{2}$.

2.3 Multiple-View Homographies

When it comes to aligning more than two images from multiple views, Equation 2.6 should, in the noise free case, hold for each pair of images at the same time. Similar to the real two-view case, the projected points certainly suffer from noise and again it is desirable to find the MLE which do best reflect the system's state.

2.3.1 Maximum Likelihood Estimators

Even though an estimator $\hat{\mathtt{H}}_{i,j}$ might be the MLE while exclusively taking into account $\mathcal{I}_i$ and $\mathcal{I}_j$ it does not meet the requirements to be the MLE for $\mathtt{H}_{i,j}$ as soon as any other image overlapping with $\mathcal{I}_i$ and $\mathcal{I}_j$ is taken into account.

Whereas three homographies $\mathtt{H}_{i,j}$, $\mathtt{H}_{i,k}$, and $\mathtt{H}_{k,j}$ should theoretically hold the relation

$$\mathtt{H}_{i,j} \sim \mathtt{H}_{k,j}\mathtt{H}_{i,k} \tag{2.13}$$

their estimators won't do so for sure and – as a consequence – a point $\mathbf{q}_z^i$ which is transformed to the coordinate frame of $\mathcal{I}_k$, $\mathcal{I}_j$, and back to that one of $\mathcal{I}_i$ will get a drift compared to the initial point.

$$\mathbf{q}_z^i \not\sim \hat{\mathtt{H}}_{j,i}\hat{\mathtt{H}}_{k,j}\hat{\mathtt{H}}_{i,k}\mathbf{q}_z^i \tag{2.14}$$

and subsequently applying that transformation multiple times will increase the drift more and more.

In order to get rid of this loop closing problem and to find more appropriate Maximum Likelihood Estimators, the homographies $\mathtt{U}_1, \ldots, \mathtt{U}_n$ are introduced. $\mathtt{U}_i$ puts image $\mathcal{I}_i$ into relation with a supplementary and arbitrary coordinate frame. For a set of n images this

arbitrary coordinate frame is consistently the same throughout the involved images, so that Equation 2.6 can be extended to

$$\forall i, j \in [n], z \in [m] : \ \mathbf{q}_z^j \sim \mathtt{U}_j \mathtt{U}_i^{-1} \mathbf{q}_z^i. \tag{2.15}$$

Assuming still that the measured points in each image suffer from Gaussian noise, the Maximum Likelihood Estimators $\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n$ can now be found by minimizing the reprojection error

$$\mathcal{C}_{\mathsf{BA}}(\hat{\mathbf{q}}_1, \ldots, \hat{\mathbf{q}}_m, \hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \sum_{z=1}^{m} \sum_{i=1}^{n} \delta_{z,i} d^2(\mathbf{q}_z^i, \hat{\mathtt{U}}_i \hat{\mathbf{q}}_z) \tag{2.16}$$

with

$$\delta_{z,i} = \begin{cases} 1 & \text{if } \mathbf{q}_z^i \text{ exists} \\ 0 & \text{otherwise.} \end{cases} \tag{2.17}$$

The process of optimizing this cost is called Bundle Adjustment (BA).

The Bundle Adjustment obviously needs to be initialized and, as mentioned in Section 2.1, the result will depend on the quality of this initialization.

2.3.2 Initialization

In Chapter 1 it has already been pointed out, that when it comes to cope with exactly aforementioned required initialization, several questions arise instantaneously. Although it is easily possible to work around those issues in order to make the BA "just work", there are very good reasons for assuming that one could deduce initializations of superior quality by carefully developing the initialization technique.

The main issue for initializing the BA is the one of finding initializers for the global homographies $\mathtt{U}_1, \ldots, \mathtt{U}_n$ and as long as the ultimate initialization technique – which reasonably approximates global homographies directly based on the point correspondences and removing false matches – is still to be found, the initialization techniques are based on the MLE of the local inter-frame homographies, the determination of which are proven concepts. In the past, several publications have been made, proposing different techniques and trying to figure out what would be reasonable criteria for deciding which local homographies to choose as initialization for the global ones, and how to combine them to interpolate missing homographies or how to extract an initialization out of the local homographies. The techniques resulting from the investigation of the initialization problem will be presented in the following Chapter 3.

Chapter 3

Related Work

In the topic of multiple-view homography estimation (not only limited to the rotation-only constrained model which is handled in this report), the work which has been done so far can be classified into two main strategies. Both aim on extracting the best initial guesses for bundle adjustments in order to get better results out of the final optimization.

On the one hand, there are those approaches which treat the problem as a searching problem and focus on different heuristics about how error propagation may be parameterized in order to find a shortest path in the image graph and which will be referred to as threading type strategies during this report. Preceding work which has been done on this field will be mentioned in Section 3.1.

On the other hand, the methods which will be denoted as batch type methods. These methods intend to formulate a reasonable cost function which is to be optimized in order to find an appropriate solution (resp. likely state of the model). The cost functions are usually formulated with the aim of being linearly solvable or to be at least convex in order to avoid the circumstances which one encounters while optimizing non-linear functions. Some batch type approaches studied in the past are summarized in Section 3.2.

3.1 Threading Type Methods

Threading type solutions are probably the first type of solutions one might think of when one faces the problem of recovering initialization parameters for a bundle adjustment. As a consequence some research has been made in the past, which will be shortly described in the following.

In order to create multiple view mosaics Capel focuses mainly on removing false 2-view matches by validating matches in multiple views in [Cap04]. The initialization of the BA though is done by choosing a reference image and enchaining those homographies, the images of which they put into relation have maximum overlapping surface.

In [KCM00] Kang et al. compute the offsets of features projected to a reference mosaic. These offsets are then used to build a weighted frame-graph which reflects the geometric

offset and correlation between the frames. When the graph has been built, global homographies can be extracted from it by searching an optimal (to some extent shortest) path or tree. In order to determine the distance of a whole path, the geometric offsets of the features themselves are accumulated through the path. Similar to the model considered herein, they restrained their work on the parallax-free case.

A slightly different approach was followed by Vergés-Llahí et al. in [VLMW07] in order to extract an initialization from a Camera Dependency Graph (CDG) in the context of Structure from Motion (SfM). The method was later reused by Bajramovic and Denzler in [BD08] which focused on the selection criterion of relative poses for multi camera calibration. They follow the idea of determining the reliability of edges in the CDG and selecting only the most reliable subgraph for a global initialization, where the reliability of an edge depends on how much it can be approved by circuitous (2-edge) paths known in the CDG. Vergés-Llahí and Wada call the problem in [VLW08] the one of "finding the shortest triangle connected subgraph". As they state, the approach significantly improves the results of the respective application. Although their reliability model (obviously) assumes different camera centres, it could certainly be applied to static camera centre scenarios too.

For the application of high resolution video mosaicing, Marzotto et al. propose a relatively simple edge-weighting criterion which favours edges between images with higher "degree of overlap" as they denote it in [MFM04]. They state that their method can even reduce the number of iterations needed, compared to other common methods.

3.2 Batch Type Methods

But also in the field of batch type methods much significant work has been done. Most of the solutions arising from that work represent not so trivial procedures but most of them aim at being the most general and flexible as possible. Thus, they include i.a. camera translation and/or multiple planes in their model or even provides Structure from Motion (SfM) techniques.

Sturm proposes a means to estimate globally consistent poses by factorizing local plane and camera poses appropriately [Stu00]. The solution addresses quite generally the problem of pose estimation in human created environments (as sufficient planar structures are required). Although the used factorization technique cannot handle missing data implicitly, interpolating that data from known data, provides still very accurate results. The solution handles the issue of missing depth of planes in views by arbitrarily initializing them (in the paper it is done so with 1) and iteratively converging to the final depth. A drawback of this method is that the convergence to a solution is not necessarily guaranteed.

In [MC02], Malis and Cipolla formulate a solution which iteratively improves the estimates of various parameters (intrinsic camera parameters, collineations) in a batch and enforces the multi-view constraints imposed on their model. Latter is held quite general too and handles multiple planes, views, and altering camera intrinsics including radial distortion. Although they solve the problem iteratively, the work aims on solving quite similar least-squares optimization problems as the other batch type solutions mentioned in this section.

Govindu's Lie-algebraic averaging framework [Gov04] provides a solution which also iteratively converges to an approximated registration. The motion model comprises rotation and translation and the averaging over known data is constrained according to Special Orthogonal Group $\mathbb{SO}_3$ and the Special Euclidean Group $\mathbb{SE}_3$ respectively. The globally consistent estimation of motion ignores point correspondences and will fit a global motion model to the known two-view motions. Depth issues are addressed by approximating depths with the previously developed approach from [Gov01] and which will converge to a final solution. The technique copes with unknown two-view motions implicitly without the need for interpolation of missing data.

Chapter 4

Closed-Form Solutions

For some unknown reason the research area coping with the initialization of the BA treats either the parallax free case with threading type methods or approaches the full motion (rotation and translation) problem with batch type techniques but is lacking the investigation of a batch approach for the translation free motion model. Of course the threading type methods yield reasonable results and the more general batch type methods could probably solve the problem of initialization for the constrained case too, but the methods proposed in the following drastically reduce the complexity compared to both threading and batch type techniques and some of them provide results at least as good as those resulting from threading type methods – as will be seen later in chapter 5 and 6 – while turning obsolete the question about interpolating missing data on the one hand side and provides a means on how to use all the known data to full extent.

The idea behind the proposed closed-form solutions is that the global homographies $\mathtt{U}_i, i \in [n]$ are not only already contained in the local homographies

$$\mathtt{H}_{i,j} \sim \mathtt{U}_j \mathtt{U}_i^{-1}, \forall i, j \in [n], \tag{4.1}$$

but that they are even contained redundantly.

Expressed in a very large and unspecific sense, it should be possible to extract these redundant information somehow out of the measured $\hat{\mathtt{H}}_{i,j}$ and "average out" the extracted information to get – from a probabilistic point of view – a better initialization for the Bundle Adjustment.

For the following sections, the inter-frame homographies $\mathtt{H}_{i,j}$ are considered as measured input data and will *not* be noted as $\hat{\mathtt{H}}_{i,j}$ even though they are (usually) the MLE of the real homographies.

4.1 Non-Homogeneous Group for Full-Rank-Homographies

The first issue to handle is that the homographies are in general only defined up to scale. Hence Equation 4.1 cannot be used as is to compare how close approximated and/or mea-

sured homographies might be one to each other. The homographies therefore have to be normalized so that they can be compared linearly.

The solution to this problem is to rescale each homography $\mathtt{H}$ so that

$$\det \mathtt{H} = 1 \tag{4.2}$$

by dividing them by the 3rd root of their determinant:

$$\mathtt{H} \leftarrow \frac{\mathtt{H}}{\sqrt[3]{\det \mathtt{H}}} \tag{4.3}$$

Implicating this normalization for any of the used homographies makes a linear group of the homographies in regard to the standard matrix multiplication. Precisely, they will be part of the special linear group $\mathbb{SL}_3$. $\mathbb{SL}_n$ is the special linear group of all invertible $n \times n$ matrices with determinant 1:
Let $\mathtt{A}, \mathtt{B}, \mathtt{C}$ be any invertible $n \times n$ matrices of determinant 1, then ...

Closure ... $\mathtt{AB}$ is an invertible $n \times n$ matrix too ($(\mathtt{AB})^{-1} = \mathtt{B}^{-1}\mathtt{A}^{-1}$) and also of determinant 1 ($\det(\mathtt{AB}) = \det(\mathtt{A}) \det(\mathtt{B}) = 1$),

Associativity ... $(\mathtt{AB})\mathtt{C} = \mathtt{A}(\mathtt{BC})$, which directly results from the properties of matrix multiplication and the closure,

Identity Element ... $\mathtt{AI} = \mathtt{IA} = \mathtt{A}$ and $\det \mathtt{I} = 1$,

Inverse Element ... $\mathtt{A}\mathtt{A}^{-1} = \mathtt{I}$ and $\det(\mathtt{A}^{-1}) = \det(\mathtt{A})^{-1} = 1$.

□

This normalization of all homographies in use enables us now to rewrite Equation 4.1 with a strict equality:

$$\forall (i,j) \in [n]^2 : \mathtt{H}_{i,j} = \mathtt{U}_j \mathtt{U}_i^{-1}, \qquad \mathtt{H}_{i,j}, \mathtt{U}_i, \mathtt{U}_j \in \mathbb{SL}_3. \tag{4.4}$$

The establishment of this equality is absolutely mandatory for the further deduction of the closed-form solutions as will be seen during the following sections!

4.2 Full Data Solutions

When creating mosaics from multiple images, there is usually a big chance that many images do not overlap or do not have enough corresponding points in order to determine their local alignment directly. In a first step however this situation is ignored and it is assumed that for each inter-frame homography the MLE is known, which is herein denoted as "full data". The "missing data" scenario is being treated later on in Section 4.3.

4.2.1 Full Data SVD (FDS)

As the measured inter-frame homographies $\mathtt{H}_{i,j}$ suffer from noise too, it won't be possible to find a set of homographies $\mathtt{U}_1, \ldots, \mathtt{U}_n$ which hold exactly for Equation 4.4. Therefore we reformulate latter to

$$(4.4) \Longrightarrow \forall (i,j) \in [n]^2 : \mathtt{H}_{i,j} - \mathtt{U}_j \mathtt{U}_i^{-1} = \mathtt{0}. \tag{4.5}$$

Under exclusive consideration of the measured $\mathtt{H}_{i,j}$ and not knowing anything more about the complex type of noise which the homographies really suffer from, it stands to reason to assume that these homographies are normally distributed data too [SS01a]. With that reasoning and with Equation 4.5 the MLE for the global homographies $\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n$ minimize the distance between the measured and estimated homographies in terms of the Frobenius distance. (The Frobenius norm is denoted $\|\cdot\|$ herein.) This leads to the cost

$$\mathcal{C}_{\mathsf{FDS}}(\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathtt{H}_{i,j} - \hat{\mathtt{U}}_j \hat{\mathtt{U}}_i^{-1}\|^2 \tag{4.6}$$

which is to be minimized. This cost can be rewritten as

$$\mathcal{C}_{\mathsf{FDS}}(\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \left\| \begin{bmatrix} (\mathtt{H}_{1,1} - \hat{\mathtt{U}}_1 \hat{\mathtt{U}}_1^{-1}) & \ldots & (\mathtt{H}_{n,1} - \hat{\mathtt{U}}_1 \hat{\mathtt{U}}_n^{-1}) \\ \vdots & \ddots & \vdots \\ (\mathtt{H}_{1,n} - \hat{\mathtt{U}}_n \hat{\mathtt{U}}_1^{-1}) & \ldots & (\mathtt{H}_{n,n} - \hat{\mathtt{U}}_n \hat{\mathtt{U}}_n^{-1}) \end{bmatrix} \right\|^2 \tag{4.7}$$

$$= \left\| \begin{bmatrix} \mathtt{H}_{1,1} & \ldots & \mathtt{H}_{n,1} \\ \vdots & \ddots & \vdots \\ \mathtt{H}_{1,n} & \ldots & \mathtt{H}_{n,n} \end{bmatrix} - \begin{bmatrix} \hat{\mathtt{U}}_1 \hat{\mathtt{U}}_1^{-1} & \ldots & \hat{\mathtt{U}}_1 \hat{\mathtt{U}}_n \\ \vdots & \ddots & \vdots \\ \hat{\mathtt{U}}_n \hat{\mathtt{U}}_1 & \ldots & \hat{\mathtt{U}}_n \hat{\mathtt{U}}_n \end{bmatrix} \right\|^2. \tag{4.8}$$

Introducing $\hat{\mathtt{U}}'_i = \hat{\mathtt{U}}_i^{-1}$, leads to a relaxation of constraints and with

$$\mathcal{H} = \begin{bmatrix} \mathtt{H}_{1,1} & \ldots & \mathtt{H}_{n,1} \\ \vdots & \ddots & \vdots \\ \mathtt{H}_{1,n} & \ldots & \mathtt{H}_{n,n} \end{bmatrix}, \qquad \hat{\mathcal{U}} = \begin{bmatrix} \hat{\mathtt{U}}_1 \\ \vdots \\ \hat{\mathtt{U}}_n \end{bmatrix}, \quad \text{and} \quad \hat{\mathcal{U}}' = \begin{bmatrix} \hat{\mathtt{U}}'_1 & \ldots & \hat{\mathtt{U}}'_n \end{bmatrix}$$

the cost can be altered to

$$\mathcal{C}_{\mathsf{FDS}}(\hat{\mathcal{U}}, \hat{\mathcal{U}}') = \|\mathcal{H} - \hat{\mathcal{U}}\hat{\mathcal{U}}'\|^2. \tag{4.9}$$

In the noise free case, $\mathcal{H}$ is of $\mathrm{rank}(\mathcal{H}) = 3$ [Stu00, MC02] and has three non-zero singular values. Under presence of noise, the problem is therefore the same as the one of approximating $\mathcal{H}$ with its three most significant singular vectors. This is, in a first step, achieved by performing a Singular Value Decomposition (SVD) of $\mathcal{H}$ which yields

$$\mathtt{U}\Sigma\mathtt{V}^{\mathsf{T}} \overset{\mathsf{SVD}}{\longleftarrow} \mathcal{H}. \tag{4.10}$$

Then $\hat{\mathcal{U}}$ can easily be constructed of the three left-most column vectors of $\mathtt{U}\sqrt{\Sigma}$ (resp. $\hat{\mathcal{U}}'$ of the three top-most row vectors of $\sqrt{\Sigma}\mathtt{V}^\mathsf{T}$) with the implication that

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{3n} \end{bmatrix} \qquad \text{s.t. } \sigma_1 \geqslant \dots \geqslant \sigma_{3n} \geqslant 0. \tag{4.11}$$

Appendix A shows that in this case $\hat{\mathcal{U}}\hat{\mathcal{U}}'$ best approximates $\mathcal{H}$ in terms of the Frobenius norm.

In [Stu00] Sturm uses a very similar approach in order to factorize rotations, and so he faces the problem that the factorized matrices might probably not be pure rotations under presence of noise and has to rectify this in a following step. As the model herein only describes homographies which are far less constrained, the factorized matrices can be used without further treatment even though their determinant might not exactly be 1.

The 3×3 ambiguity which resides here and which might be modeled by any invertible 3×3 matrix $\mathtt{C}$ in $\mathtt{AB} = \mathtt{ACC}^{-1}\mathtt{B}$ can be ignored.

4.2.2 Full Data Eigenvector (FDE)

Although the preceding approach already provides a closed-form solution it is not very satisfactory because, due to the required relaxation of constraints, it provides two solutions of lower quality ($\hat{\mathcal{U}}$ and $\hat{\mathcal{U}}'$) instead of one single solution of better quality. Hence, the next step is to formulate a new cost with $\hat{\mathcal{U}}$ as the only parameter.

Instead of subtracting $\mathtt{U}_j\mathtt{U}_i^{-1}$ from Equation 4.4 (which was the step in order to get Equation 4.5), it will be multiplied by $\mathtt{U}_i$ which leads to the fact that

$$\forall k \in [n] : \sum_{i=1}^{n} \mathtt{H}_{i,k}\mathtt{U}_i = \sum_{i=1}^{n} \mathtt{U}_k\mathtt{U}_i^{-1}\mathtt{U}_i = n\mathtt{U}_k. \tag{4.12}$$

With the same reasoning about the noise distribution of the measured homographies as that one made in Section 4.2.1, the relation of 4.12 is best approximated by minimizing

$$\mathcal{C}_{\mathsf{FDE}}(\hat{\mathtt{U}}_1, \dots, \hat{\mathtt{U}}_n) = \sum_{k=1}^{n} \left\| n\hat{\mathtt{U}}_k - \sum_{i=1}^{n} \mathtt{H}_{i,k}\hat{\mathtt{U}}_i \right\|^2 \tag{4.13}$$

In order to better illustrate how to get to the final cost in Equation 4.19, let's consider the noise free case for a short moment again. At first the $\sum_{i=1}^{n}$ in Equation 4.12 is replaced by an appropriate matrix product and afterwards the n equations (one for each k) are stacked up

to a single matrix equation

$$(4.12) \Leftrightarrow \qquad \forall k \in [n]: \quad \underbrace{\begin{bmatrix} \mathtt{H}_{1,k} & \dots & \mathtt{H}_{n,k} \end{bmatrix} \begin{bmatrix} \mathtt{U}_1 \\ \vdots \\ \mathtt{U}_n \end{bmatrix}}_{\mathcal{U}} = n\mathtt{U}_k \tag{4.14}$$

$$\Leftrightarrow \qquad \begin{bmatrix} \left(\begin{bmatrix} \mathtt{H}_{1,1} & \dots & \mathtt{H}_{n,1} \end{bmatrix} \mathcal{U} \right) \\ \vdots \\ \left(\begin{bmatrix} \mathtt{H}_{n,1} & \dots & \mathtt{H}_{n,n} \end{bmatrix} \mathcal{U} \right) \end{bmatrix} = \begin{bmatrix} n\mathtt{U}_1 \\ \vdots \\ n\mathtt{U}_n \end{bmatrix} \tag{4.15}$$

$$\Leftrightarrow \qquad \underbrace{\begin{bmatrix} \begin{bmatrix} \mathtt{H}_{1,1} & \dots & \mathtt{H}_{n,1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathtt{H}_{1,n} & \dots & \mathtt{H}_{n,n} \end{bmatrix} \end{bmatrix}}_{\mathcal{H}} \mathcal{U} = n \underbrace{\begin{bmatrix} \mathtt{U}_1 \\ \vdots \\ \mathtt{U}_n \end{bmatrix}}_{\mathcal{U}} \tag{4.16}$$

$$\Leftrightarrow \qquad \mathcal{H}\mathcal{U} = n\mathcal{U}. \tag{4.17}$$

It can easily be seen that Equation (4.17) represents some sort of an eigenvector problem. Any set of three non-equal eigenvectors of $\mathcal{H}$ corresponding to an eigenvalue equal to n provides the column vectors for a solution to $\mathcal{U}$.

Whereas in the forward reasoning (supposedly knowing $\mathtt{U}_1, \dots, \mathtt{U}_n$) the fact that every $\mathtt{U}_i$ is invertible implicates on the one hand that $\mathcal{U}$ cannot be the trivial solution $\mathtt{0}$ to Equation 4.17, and on the other hand that the three column vectors of $\mathcal{U}$ differ one from each other, these two requirements must be fulfilled explicitly for the backward reasoning where only $\mathcal{H}$ is known and $\mathcal{U}$ is to be determined.

A further issue is represented by the fact that an eigenvector decomposition might yield complex valued vectors which are absolutely valid eigenvectors, but do not deliver an appropriate solution to the problem stated herein. Hence, Equation 4.17 is rewritten as

$$\mathcal{H}\mathcal{U} - n\mathcal{U} = (\mathcal{H} - n\mathtt{I})\mathcal{U} = \mathtt{0} \tag{4.18}$$

which yields the final cost to minimize

$$\mathcal{C}_{\mathsf{FDE}}(\hat{\mathcal{U}}) = \|(\mathcal{H} - n\mathtt{I})\hat{\mathcal{U}}\|^2, \qquad \text{s.t. } \hat{\mathcal{U}}^{\mathsf{T}}\hat{\mathcal{U}} = \mathtt{I} \tag{4.19}$$

and which can be solved by performing a SVD

$$\mathtt{U}\Sigma\mathtt{V}^{\mathsf{T}} \overset{\mathsf{SVD}}{\longleftarrow} \mathcal{H} - n\mathtt{I}. \tag{4.20}$$

Then the three right singular vectors (contained as column vectors in $\mathtt{V}$) corresponding to the three smallest (ideally 0) singular values $\sigma_{n-2}, \sigma_{n-1}, \sigma_n$ represent the column vectors of the solution to $\hat{\mathcal{U}}$ which minimizes $\mathcal{C}_{\mathsf{FDE}}(\hat{\mathcal{U}})$. Appendix B gives the proof that the SVD solves this minimization problem. (The constraint $\hat{\mathcal{U}}^{\mathsf{T}}\hat{\mathcal{U}} = \mathtt{I}$ ensures that the earlier mentioned requirements, that all of the three column vectors of $\hat{\mathcal{U}}$ are different ones and that $\hat{\mathcal{U}} \neq \mathtt{0}$, are fulfilled.)

4.2.3 Explicit Missing Data Interpolation

In order to use FDS or FDE (as described earlier) even with lacking data, the missing entries in $\mathcal{H}$ have to be filled with some artificially created data. Sturm also faces the missing data problem when he factorizes rotations [Stu00], and similar to his solution to the problem, the missing entries of $\mathcal{H}$ can be composed of known ones. In the case that there are multiple possibilities to compose the missing entries, all possible combinations are used and averaged. Although, the averaged solution can be constrained stronger for rotations, the idea will still be used herein.

In order to limit the propagation of errors due to combinations of noisy homographies, only the "shortest" combinations possible are being used to interpolate missing entries. If known homographies are noted as $\mathtt{H}_{i,j}^{(0)}$ and $\mathtt{H}_{i,j}^{(l)}, l \in \mathbb{N}^*$ denote missing ones composed of other homographies, then missing homographies are interpolated with

$$\mathtt{H}_{i,j}^{(l)} = \frac{1}{|\Gamma_{i,j}^{(l)}|} \sum_{k \in \Gamma_{i,j}^{(l)}} \mathtt{H}_{k,j}^{(l-1)} \mathtt{H}_{i,k}^{(l-1)} \tag{4.21}$$

with

$$\Gamma_{i,j}^{(l)} = \left\{ x | \exists \mathtt{H}_{i,x}^{(l-1)} \wedge \exists \mathtt{H}_{x,j}^{(l-1)} \right\}. \tag{4.22}$$

Alhtough this interpolation will probably work not too bad for a small ratio of missing to known data, it is not recommended if that ratio increases. As the FDE method will be extended to handle missing data implicitly without the need for interpolation, this method is mainly intended for application with FDS, because factorization of sparse data [Jac97, TK92, SS01b] becomes much more complicated and it is tried to avoid the need for it herein.

4.3 Missing Data Solutions

In the preceding Section 4.2 two solutions were introduced which are both based on the assumption that for n images the $n(n-1)$ local homographies, inter-aligning each image to each other, would be known (referred to as "full data" case). Unfortunately this is hardly anytime the case (referred to as "missing data" case) and very often many of the local homographies – in many scenarios even most of those homographies – have to be interpolated by composing those (few) local homographies which can be estimated with standard methods. As mentioned in Chapter 1 this will propagate errors throughout the system and most probably worsen any result. Hence it is desirable to find a means not to have to close these gaps of data explicitly.

Based on the reasoning in Section 4.2.2 two more closed-form solutions will be introduced in the following. Both of them handle the missing data problem implicitly, though. The only requirement is that the graph, the nodes of which represent the images and, the edges of which represent the local homographies, has obviously to be a connected one.

Actually, Equation 4.12 could be considered the answer to the question "For a destination image $\mathcal{I}_k$, and n source images $\mathcal{I}_1, \ldots, \mathcal{I}_n$. How many times is $\mathtt{U}_k$ contained in the local

homographies $\mathtt{H}_{1,k}, \ldots, \mathtt{H}_{n,k}$?". Answer: "$n$ times!" And in fact, that is – in some way — already the answer to how to handle missing data implicitly, too. The only limitation that equation suffers from, is that the same number of source images is being supposed for every destination image. Hence, it suffices to extend the equation to consider each destination image seperately:
In order to formulate whether a local homography $\mathtt{H}_{i,j}$ is known or not,

$$\gamma_{i,j} = \begin{cases} 1, & \text{if } \mathtt{H}_{i,j} \text{ is known,} \\ 0, & \text{otherwise} \end{cases} \tag{4.23}$$

is introduced. Then the number of source images which provide a known local homography to a destination image $\mathcal{I}_k$ can be formulated as

$$\zeta_k = \sum_{i=1}^{n} \gamma_{i,k} \tag{4.24}$$

which allows to extend Equation 4.12 to

$$\forall k \in [n]: \quad \sum_{i=1}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \mathtt{U}_i = \zeta_k \mathtt{U}_k. \tag{4.25}$$

Whereas, in the noise free case, it doesn't matter whether Equation 4.25 is divided by ζ_k (scaling of local homographies) or not and left as is (scaling of global homographies), it turns out, as discussed hereafter, that in the real case, it *does* matter in regard to the cost function. Both solutions are therefore considered separately in the following Sections 4.3.1 and 4.3.2.

4.3.1 Locally Scaled Homographies (LSH)

For the first of the proposed missing data solutions

$$(4.25) \Leftrightarrow \qquad \forall k \in [n]: \quad \frac{1}{\zeta_k} \sum_{i=1}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \mathtt{U}_i = \mathtt{U}_k \tag{4.26}$$

serves as the theoretical base leading to the cost function

$$\mathcal{C}_{\mathsf{LSH}}(\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \sum_{k=1}^{n} \left\| \hat{\mathtt{U}}_k - \frac{1}{\zeta_k} \sum_{i=1}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \hat{\mathtt{U}}_i \right\|^2. \tag{4.27}$$

Again $\mathcal{C}_{\mathsf{LSH}}$ may be expressed in matrix form with an approach very similar to the one in

Section 4.2.2:

$$(4.26) \Leftrightarrow \qquad \forall k \in [n]: \quad \frac{1}{\zeta_k} \begin{bmatrix} \gamma_{1,k}\mathtt{H}_{1,k} & \dots & \gamma_{n,k}\mathtt{H}_{n,k} \end{bmatrix} \begin{bmatrix} \mathtt{U}_1 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} = \mathtt{U}_k \tag{4.28}$$

$$\Leftrightarrow \qquad \underbrace{\begin{bmatrix} \frac{1}{\zeta_1} \begin{bmatrix} (\gamma\mathtt{H})_{1,1} & \dots & (\gamma\mathtt{H})_{n,1} \end{bmatrix} \\ \vdots \\ \frac{1}{\zeta_n} \begin{bmatrix} (\gamma\mathtt{H})_{1,n} & \dots & (\gamma\mathtt{H})_{n,n} \end{bmatrix} \end{bmatrix}}_{\mathcal{S}} \begin{bmatrix} \mathtt{U}_1 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} = \begin{bmatrix} \mathtt{U}_1 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} \tag{4.29}$$

$$\Leftrightarrow \qquad \mathcal{S}\mathcal{U} = \mathcal{U} \tag{4.30}$$

$$\Leftrightarrow \qquad (\mathcal{S} - \mathtt{I})\mathcal{U} = \mathtt{0}. \tag{4.31}$$

Equation 4.31 shows that $\mathcal{C}_{\mathsf{LSH}}$ can be written in matrix form as

$$\mathcal{C}_{\mathsf{LSH}}(\hat{\mathcal{U}}) = \|(\mathcal{S} - \mathtt{I})\hat{\mathcal{U}}\|^2, \quad \text{s.t. } \hat{\mathcal{U}}^\mathsf{T}\hat{\mathcal{U}} = \mathtt{I}. \tag{4.32}$$

A solution $\hat{\mathcal{U}}$ to $\underset{\hat{\mathcal{U}}|\hat{\mathcal{U}}^\mathsf{T}\hat{\mathcal{U}}=\mathtt{I}}{\arg\min}\ \mathcal{C}_{\mathsf{LSH}}(\hat{\mathcal{U}})$ can be found in the same way as it is done for $\mathcal{C}_{\mathsf{FDE}}$, by performing a SVD of $\mathcal{S} - \mathtt{I}$.

This solution is referred to as "Locally Scaled Homographies", because – expressed naively – the rescaling, made in order to make the equations suit according to the number of available data, is performed on the known local homographies. Although it represents a solution which implicitly handles missing data, it contains two drawbacks of which one gets aware of while taking a second look at the cost used to solve data suffering from noise and which will be rectified in the following section.

4.3.2 Globally Scaled Homographies (GSH)

On the one hand, in the preceding solutions, the homographies $\mathtt{H}_{i,i}$ between any image and itself have been considered as being part of the measured data. Thus, those cost functions misleadingly take into account that $\mathtt{H}_{i,i}$ might be off it's real value although it is absolutely sure that $\mathtt{H}_{i,i} = \mathtt{I}$. As a consequence this falsifies the weighting on the really measured data, containing valuable information about the system. Correcting this weighting should yield more likely solutions.

On the other hand, $\mathcal{C}_{\mathsf{LSH}}$ contains another crucial weighting issue. One might not get aware of it in the noise-free case but taking a closer look on $\mathcal{C}_{\mathsf{LSH}}(\hat{\mathtt{U}}_1, \dots, \hat{\mathtt{U}}_n)$ one realizes that this cost function weights errors arising from the sum of transformations leading to a destination image with only few source images available much more as the errors arising from the sum of transformations leading to a destination image with a lot of source images available.

Taking it over again from Equation 4.25 and applying both mentioned changes by taking out the term $(\gamma \mathtt{H})_{k,k}\mathtt{U}_k$ and leaving the scaling of the local homographies unchanged, leads to

$$(4.25) \Leftrightarrow \quad \forall k \in [n]: \quad \mathtt{U}_k + \sum_{i=1, i\neq k}^{n} (\gamma \mathtt{H})_{i,k}\mathtt{U}_i = \zeta_k \mathtt{U}_k \tag{4.33}$$

$$\Leftrightarrow \quad \forall k \in [n]: \quad \sum_{i=1, i\neq k}^{n} (\gamma \mathtt{H})_{i,k}\mathtt{U}_i = (\zeta_k - 1)\mathtt{U}_k \tag{4.34}$$

$$\Leftrightarrow \begin{bmatrix} \mathtt{0} & (\gamma\mathtt{H})_{2,1} & \dots & (\gamma\mathtt{H})_{n,1} \\ (\gamma\mathtt{H})_{1,2} & \mathtt{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & (\gamma\mathtt{H})_{n,n-1} \\ (\gamma\mathtt{H})_{1,n} & \dots & (\gamma\mathtt{H})_{n-1,n} & \mathtt{0} \end{bmatrix} \begin{bmatrix} \mathtt{U}_1 \\ \mathtt{U}_2 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} = \begin{bmatrix} (\zeta_1 - 1)\mathtt{U}_1 \\ (\zeta_2 - 1)\mathtt{U}_2 \\ \vdots \\ (\zeta_n - 1)\mathtt{U}_n \end{bmatrix} \tag{4.35}$$

$$\Leftrightarrow \begin{bmatrix} \mathtt{0} & (\gamma\mathtt{H})_{2,1} & \dots & (\gamma\mathtt{H})_{n,1} \\ (\gamma\mathtt{H})_{1,2} & \mathtt{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & (\gamma\mathtt{H})_{n,n-1} \\ (\gamma\mathtt{H})_{1,n} & \dots & (\gamma\mathtt{H})_{n-1,n} & \mathtt{0} \end{bmatrix} \begin{bmatrix} \mathtt{U}_1 \\ \mathtt{U}_2 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} - \begin{bmatrix} (\zeta_1 - 1)\mathtt{I}_{3\times 3} & \mathtt{0} & \dots & \mathtt{0} \\ \mathtt{0} & (\zeta_2 - 1)\mathtt{I}_{3\times 3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathtt{0} \\ \mathtt{0} & \dots & \mathtt{0} & (\zeta_n - 1)\mathtt{I}_{3\times 3} \end{bmatrix} \begin{bmatrix} \mathtt{U}_1 \\ \mathtt{U}_2 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} = \mathtt{0} \tag{4.36}$$

$$\Leftrightarrow \underbrace{\begin{bmatrix} (1-\zeta_1)\mathtt{I} & (\gamma\mathtt{H})_{2,1} & \dots & (\gamma\mathtt{H})_{n,1} \\ (\gamma\mathtt{H})_{1,2} & (1-\zeta_2)\mathtt{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & (\gamma\mathtt{H})_{n,n-1} \\ (\gamma\mathtt{H})_{1,n} & \dots & (\gamma\mathtt{H})_{n-1,n} & (1-\zeta_n)\mathtt{I} \end{bmatrix}}_{\mathcal{G}} \begin{bmatrix} \mathtt{U}_1 \\ \mathtt{U}_2 \\ \vdots \\ \mathtt{U}_n \end{bmatrix} = \mathtt{0} \tag{4.37}$$

$$\Leftrightarrow \quad \mathcal{G}\mathcal{U} = \mathtt{0} \tag{4.38}$$

and for the noise suffering case is solved by minimizing the cost

$$\mathcal{C}_{\mathsf{GSH}}(\hat{\mathcal{U}}) = \|\mathcal{G}\hat{\mathcal{U}}\|^2, \quad \text{s.t. } \hat{\mathcal{U}}^{\mathsf{T}}\hat{\mathcal{U}} = \mathtt{I}. \tag{4.39}$$

Once again, $\hat{\mathcal{U}}$ which minimizes $\mathcal{C}_{\mathsf{GSH}}(\hat{\mathcal{U}})$ can be retrieved by performing a SVD of $\mathcal{G}$ and proceeding analogue to Sections 4.2.2 and 4.3.1.

Chapter 5

Synthetic Data Experiments

In order to know whether the proposed solution(s) would eventually increase the accuracy of the bundle adjustment, and if it would so, to what extent, the solutions have been compared to a naive threading type method with synthetic data. Synthetically generated data allows, on the one hand, to compare many differing scenes in order to get statistically reasonable results, and on the other hand, it allows to exactly control which effects do influence the experiments. In contrast to real data experiments, the latter argument allows for example to silently eliminate outliers and erroneous matches. Those would not necessarily influence the statistical representativity of the generated data, but complicate the evaluation of the computed results.

5.1 Implementation Details

The experimental framework was implemented and realized in Matlab. Although it is not a very fast language it represents a very convenient, powerful, and comfortable tool for developing and performing prototype-like applications and maths. The most powerful feature in that context is probably that new implementation steps can be executed directly from the command line and results can immediately be visualized and evaluated in detail, which drastically simplifies location of programming mistakes and/or quick detection of erroneous assumptions in the underlying theoretical approach.

5.1.1 Data Generation

Scene The first step in regard to running synthetic experiments, is to generate artificial scenes, which sufficiently reflect a similar behaviour expected or desired from the modeled real world examples. For the experiments herein, it is desired to avoid erroneous point correspondences between the generated images. Hence, it suffices absolutely to immediately generate points in space which will represent the ”real-world”, noise-free instances of the feature points which would be detected in the images in reality. Additionally, as the model forbids parallax, the (virtual) user won't have any depth perception of the points and ran-

domly generating normally distributed points centered to the origin will give the user the impression of being surrounded by uniformly distributed points in space. (Really generating uniformly distributed points would require to distribute them uniformly inside a sphere, otherwise the user would get the impression that they were not distributed in a uniform manner; this would result in supplementary but completely needless processing.) For the experiments herein those points have been generated using $\mathcal{N}(0, 10)$ for each of the three components but any other value for the standard deviation would generate similar results as long as the values are consistently equal for all the components. The generated points form the synthetic scene.

Viewpoints The next step is to simulate multiple cameras pointing into several directions. As the centres of projection are constrained to being the same for every camera, the camera centres are fixed to the origin. Hence, the extrinsic parameters of a camera i are fully modeled by the rotation $\mathtt{R}_i^\mathsf{T}$. Imitating a camera mounted on a tripod, the rotation is simply composed by a rotation $\mathtt{R}_i^{(y)}$ about the y-axis, followed by a rotation $\mathtt{R}_i^{(x)}$ about the x-axis. Hence $\mathtt{R}_i = \mathtt{R}_i^{(y)}\mathtt{R}_i^{(x)}$ as the scene will be rotated instead of the camera. Both rotational components are limited to a maximum rotation angle α which can be adjusted in order to make sure that with any possible camera configuration there is always a reference frame which does not produce any points at infinity if all the points visible in any of the images are transformed to that reference frame. The rotation angles described by $\mathtt{R}_i^{(x|y)}$ are randomly generated with uniform distribution on $[-\alpha; \alpha]$.

Furthermore a camera is characterized by its camera matrix $\mathtt{K}$ which models its intrinsic parameters and which projects points from the (rotated) scene to the camera's homogeneous coordinate frame. But for computing convenience, $\mathtt{K}$ is not used till the following step.

Images Having generated the scene (points) and the viewpoints, the synthetic images can be generated from those. The images are certainly not images in the common sense, but they are represented by a list (matrix) of projected point coordinates and a list containing identifiers putting into correspondence the 3d points in space with their two dimensional projections in the respective image.

For each camera, the scene is rotated now according to $\mathtt{R}_i$ and the points are projected to the resulting image's coordinate frame, resulting in points of the form $\mathbf{q} = \mathtt{K}\mathtt{R}_i\mathbf{Q} = [uw\ vw\ w]^\mathsf{T}$. Gaussian noise with standard deviation σ is added to the u- and v-component of the projected points and points $\mathbf{q} \notin [0; 640] \times [0; 480] \times \mathbb{R}_+^*$ are clipped because they lie outside the image boundaries. (Obviously images of size 640×480 pixels have been simulated.)

Local Homographies Finally, the local homographies themselves have still to be computed. As point correspondences between images can be perfectly matched in this case, robust matching techniques can obviously be omitted. In reality however, one would have to perform a robust matching with RANSAC for example. But still, even in the perfectly matched case, this implementation will require a minimum of twenty point correspondences which can be established in order to guarantee more or less that local homographies are not

exaggeratedly off, because it has been computed with a theoretical minimum of four point correspondences with very unluckily distributed noise.

The DLT (Section 2.2.2) as presented in [HZ04] yields an initial guess $\tilde{\mathtt{H}}_{i,j}$ for the homography $\mathtt{H}_{i,j}$ between $\mathcal{I}_i$ and $\mathcal{I}_j$. The MLE $\hat{\mathtt{H}}_{i,j}$ of that homography is then non-linearly optimized according to $\mathcal{C}_{\mathsf{RP2}}$ (Section 2.2.1), using $\tilde{\mathtt{H}}_{i,j}$ as initialization.

A DLT implementation for Matlab, written by Peter Kovesi [Kov09], has been used during this work.

5.1.2 Data Evaluation

Bundle Adjustment Not being part of the data generation part anymore, but more of the data evaluation part, a BA had to be used in order to finalize the registration. Bartoli kindly provided his implementation of a BA he contributed to the AirPhoto Software [Bar04]. The code (actually the cost function) has been slightly modified in order to fit for this application. The BA is based on an implementation of the Levenberg-Marquardt-Algorithm for the optimization procedure.

Threading Solution The closed form solutions have been compared to a threading type solution. As it was not sure at the beginning of this work, whether the proposed solutions would yield any usable results or not, they were only compared to a very naive threading solution, in order to avoid implementing complex threading algorithms and realizing afterwards that the closed form solutions couldn't even compete with the simplest threading solutions.

The method simply looks for a reference image which provides a maximum number of inter-frame connections to other images. If there are more images fulfilling this constraint, the one for which the sum of mean squared errors of the homographies ($\sum \mathcal{C}_{\mathsf{RP2}}$) is minimal is chosen. The images which do not provide a direct homography to the reference frame, are being connected with interpolated homographies, according to Section 4.2.3. This might probably not yield the smallest tree of the graph, but it should yield one which is not much bigger than the smallest one. (Anyways, the "smallest" tree, depends on the criteria with which edges and (even worse) paths are weighted, and trying to develop a most meaningful weighting was not the target of this work.)

5.2 Measurands

Root Mean Squared Residual One value representing a measure of quality to evaluate throughout the experiments made, is the Root Mean Squared Residual (RMSR)

$$\epsilon = \min_{\hat{\mathbf{q}}_1,\ldots,\hat{\mathbf{q}}_m} \sqrt{\frac{\sum_{z=1}^{m}\sum_{i=1}^{n} \delta_{z,i} d^2(\mathbf{q}_z^i, \mathtt{U}_i \hat{\mathbf{q}}_z)}{\sum_{z=1}^{m}\sum_{i=1}^{n} \delta_{z,i}}} \tag{5.1}$$

which is closely related to $\mathcal{C}_{\mathsf{BA}}$, the bundle adjustment's cost function. Actually it is the reprojection error averaged out over all measured data (that is to say, image points) and rooted to represent a value expressed in pixels. The value should be close to the noise level the measured points suffer from. The RMSR of a single sample s of a test with noise level σ is referred to as $\epsilon(\sigma, s)$; $\tilde{\epsilon}$ and $\hat{\epsilon}$ denote that the RMSR had been evaluated before and after the BA.

Standard Deviation of RMSR A second value to observe, is the standard deviation of the RMSR $\mathrm{std}\{\epsilon(\sigma,1),\ldots,\epsilon(\sigma,s_{\mathsf{max}})\}$ which will be written shortly as $\mathrm{std}(\epsilon(\sigma))$. Although it does not provide any quantitative conclusions, it reflects the accuracy of the optimizations (or initial guesses) and hence shows – if compared directly to that one of another method applied to the same sample(s) – whether one method provides more reliable results as may do another solution.

Iterations of Bundle Adjustment Because there is a chance that the proposed solutions may reduce the number of iterations needed during the BA, this number will also be logged for each sample and each initialization method. It is denoted as Φ in the following.

Mean Reprojected Image Corner Distance Although the RMSR is in fact a statistical valid measure, it might, depending on the situation, have almost the same value for, on the one hand, a local minimum which is very close to the MLE, and on the other hand, a local minimum which is much closer to the ground truth. This could be the case if multiple images on the outer side of the mosaic have some kind of drift to the same direction. Therefore, an additional measurand has been introduced which gives information about the "distance" between the estimated alignment of the set of images compared to the ground truth. One attempt would have been to directly compute the squared distance between the homographies used to generate the ground truth and the estimated ones. However, as the homographies are all defined in terms of an arbitrary homography which can be altered at any time which isn't an equidistant transform, different reference frames would yield different errors. Therefore – after multiple debates – a quite uncommon error had been constructed:
Every image in the set will once serve as the reference frame (in order to handle the varying distances for different reference frames). Then the corners $\mathbf{p}_1,\ldots,\mathbf{p}_4$ of the image serving as the reference frame will be projected to every image in the set using the estimated

(global) homographies, and backprojected to its reference using the true transforms; then the distance between those backprojected corners to their original roots is averaged. This measurand will be referred to as

$$\hat{\eta} = \frac{\sum_{r=1}^{n}\sum_{i=1}^{n}\sum_{z=1}^{4} d(\mathbf{p}_z, \mathtt{U}_r\mathtt{U}_i^{-1}\hat{\mathtt{U}}_i\hat{\mathtt{U}}_r^{-1}\mathbf{p}_z)}{4n^2}. \tag{5.2}$$

To avoid any misunderstanding: $\hat{\eta}$ is *neither* a statistically quantifying value, *nor* is it a theoretically founded means of measure. It is only introduced to get a slight impression *whether or not* the proposed initializations approach the estimated solutions closer to the ground truth in terms of the image frames; how much or to what extent they do so cannot be determined upon this measure.

5.3 Results

Any experiment consists of a certain number of takes, which, at its own, consists of a certain number of samples. A sample refers to one single observation, that is to say, it refers to generating a scene with a certain level of noise σ and evaluating the different initializations. A take is composed of a fixed number of samples with increasing level of noise.

E.g. h samples per take, means that h samples with different levels of noise have been evaluated and t takes made during an experiment, means, that each level of noise has been evaluated t times.

5.3.1 Low Projective and Full Data Experiment (LPFDE)

In the first experiment, the scenes were generated with a big number of scene points (10^5). 100 images per sample were taken. The camera rotations for these images were limited to a very small $\alpha = \pi/48$ and a small field of view. As a consequence, local inter-frame homographies could be estimated between any two images, thus resulting in a "full data" experiment. Additionally, the small field of view causes only a decent projective transformation.

There were 100 takes and 10 samples per take (with noise levels $\sigma = 0.1, 0.2, \ldots, 1.0$) made for this experiment, yielding the results discussed in the following. Figure 5.1(a) shows the RMSR of the initial guesses resulting from the different methods.

In the full data case, which it is for this experiment, FDE, LSH solve exactly the same problem and GSH does *almost* do so too. Therefore $\tilde{\epsilon}_{\mathsf{FDE}}(\sigma)$ and $\tilde{\epsilon}_{\mathsf{LSH}}(\sigma)$ are plotted with a single line, which also sufficiently reflects $\tilde{\epsilon}_{\mathsf{GSH}}(\sigma)$ because the relative error

$$\left|\frac{\tilde{\epsilon}_{\mathsf{GSH}}(\sigma, s) - \tilde{\epsilon}_{\mathsf{LSH}}(\sigma, s)}{\tilde{\epsilon}_{\mathsf{LSH}}(\sigma, s)}\right| < 10^{-7} \tag{5.3}$$

is insignificantly small.

One can see in Figure 5.1(a) that the initialization parameters delivered by the eigenvector solving closed form solutions seem to be "better" as the ones provided by the threading

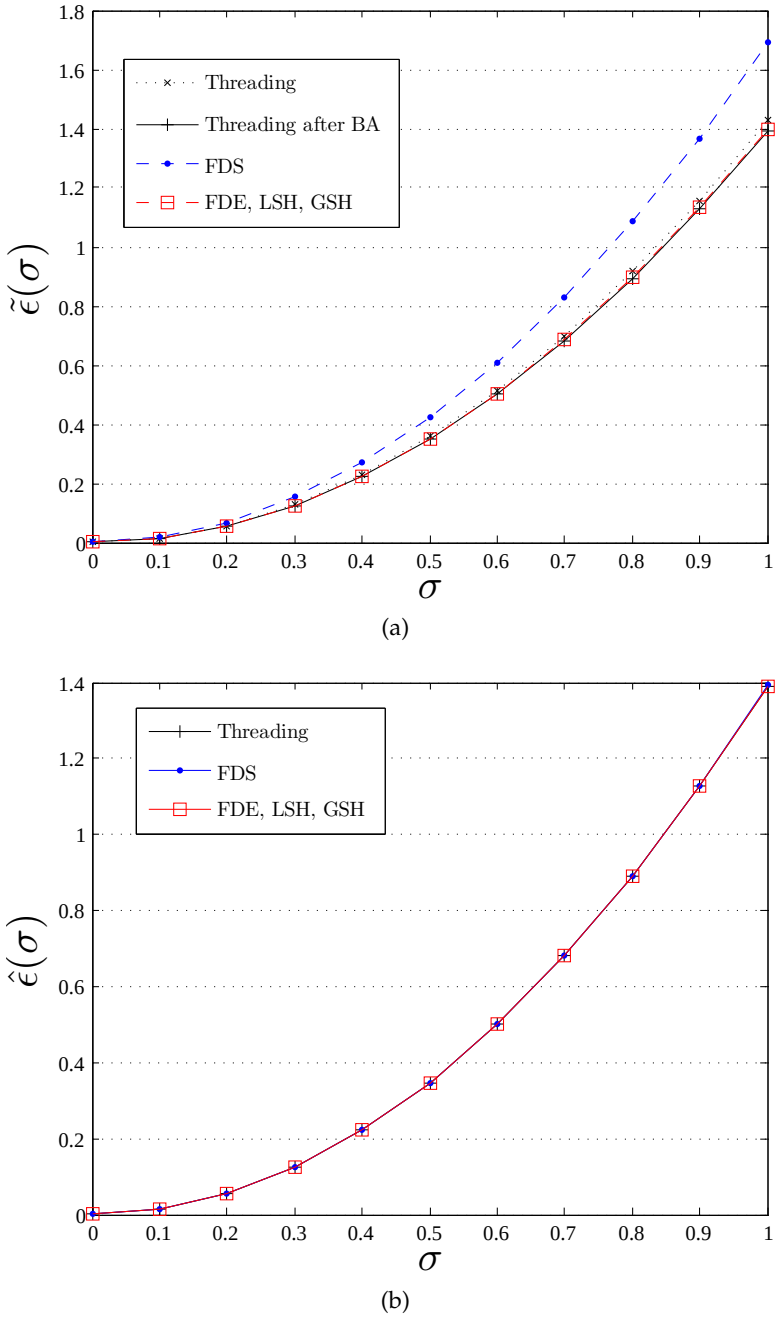


Figure 5.1: *(a) Plot of the Root Mean Squared Residual before the Bundle Adjustment. The initialization parameters found determined with FDE, LSH, and GSH (red □) is very close to the final adjustment found with the threading (bottom most black solid +); whereas the initialization found with FDS (upper most blue •) yields a RMSR which is much higher than the one achieved by the threading initialization (black ×).*
(b) Plot of the Root Mean Squared Residual after the Bundle Adjustment. The RMSR achieved with the different initializations after the bundle adjustment seems to be the same or extremely close one to each other.

solution in terms of the RMSR evaluated before the BA, whereas the factorization solution, seems not to work very well. A look at Figure 5.1(b) though reveals that after the BA all the initializations, even the one acquired by FDS, result in a similarly good minimum in regard to $\mathcal{C}_{\mathsf{BA}}$.

If the closed form solutions did not allow to achieve better results as the threading method, the next question is whether they allowed to reduce the number of iterations required by the optimization algorithm or not. The histogram in Figure 5.2 displays the dis-

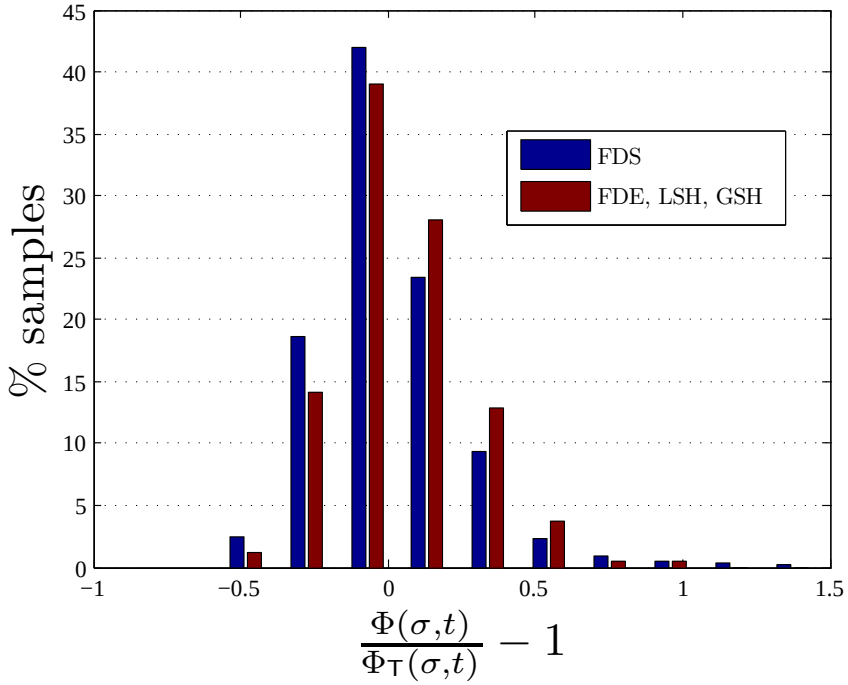


Figure 5.2: *The histogram shows for the different solutions the distribution of the ratio of iterations of the optimization algorithm in regard to those needed with initialization with the threading method.*

tribution of the ratio of iterations needed, compared to the threading type initialization and reveals, together with the key values

	$\text{mean}\left(\frac{\Phi(\sigma,t)}{\Phi_{\mathsf{T}}(\sigma,t)} - 1\right)$	$\text{std}\left(\frac{\Phi(\sigma,t)}{\Phi_{\mathsf{T}}(\sigma,t)} - 1\right)$
FDS	0.0015	0.24
FDE, LSH, GSH	−0.041	0.22

that although for many samples the number of iterations has not been the same, no significant improvement could be achieved with any of the solutions.

A really close look on the RMSR of each sample for different methods though, raises the suspicion, that in almost any sample, the bundle adjustment reached slightly different minima. That would make the comparison of Φ obsolete, as a better initialization might only reduce the number of iterations if the optimization converges to the same minimum. If not,

the optimization will most probably need a similar amount of iterations in order to reach the same precision.

Obviously, in this experiment, the closed-form solutions didn't bring any advantage to the optimization process, neither in terms of performance, nor in terms of precision. Thus, the following experiments aimed at complicating the experiments until any solution would provide any significantly deviating results.

5.3.2 Low Projective and Sparse Data Experiment (LPSDE)

For this experiment, the preceding one has been slightly modified in order to generate less known local homographies by slightly reducing the field of view even more but leaving α untouched. This resulted in samples with $72.5\% \pm 6.84$ of unknown local homographies overall, and $54.2\% \pm 13.3$ of missing local homographies leading to the reference image.

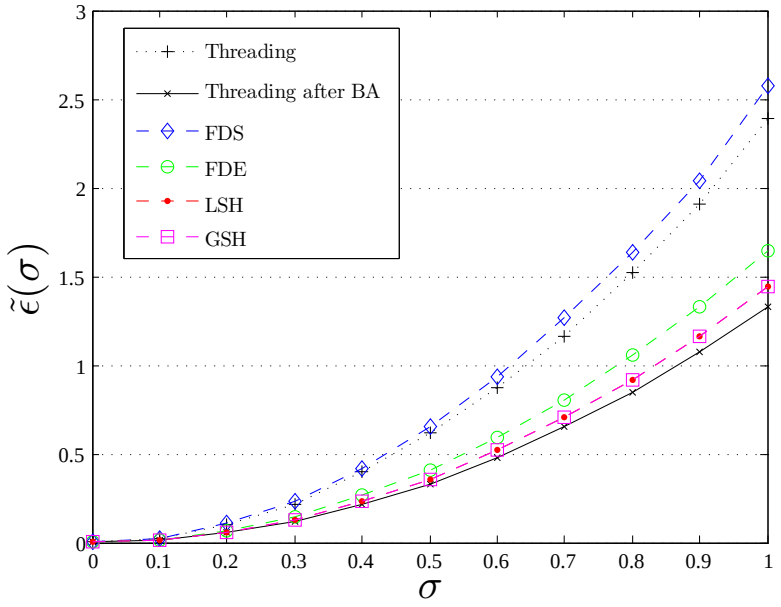
In order to be able to include FDS and FDE in this experiment, the missing data is approximated according to the interpolation technique mentioned in Section 4.2.3.

The results of the evaluation of the RMSR before the bundle adjustment are visualized in Figure 5.4(a). The diagrams show the threading and FDS are much more sensible to noise if considerably many homographies are unknown, as in this experiment the RMSR is for both much higher as it is for the other solutions and the standard deviation of the RMSR drastically increases for higher noise levels, compared to the standard deviation of the RMSR evaluated for FDE, LSH, and GSH. LSH and GSH yield again results which can be considered to be equal, because the relative difference between both is

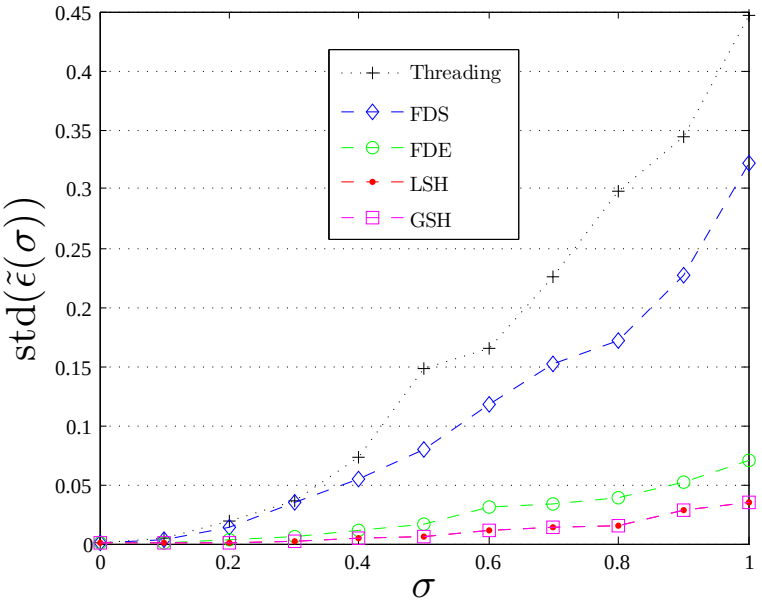
$$\left| \frac{\tilde{\epsilon}_{\mathsf{LSH}}}{\tilde{\epsilon}_{\mathsf{GSH}}} - 1 \right| < 10^{-8}. \tag{5.4}$$

Figure 5.4(a) visualizes that, regardless the significantly differing RMSR before the BA, the RMSR after the BA still results in solutions which are quite close to each other, but looking at Figure 5.4(b) indicates that for a certain number of alignments the RMSR ratio achieved with the closed-form solutions in regard to that one achieved with the threading initialization, differ from 1 by a few percent.

Furthermore the figure shows that, even though FDS yields solutions for which $\tilde{\epsilon}_{\mathsf{FDS}} > \tilde{\epsilon}_{\mathsf{T}}$, they allow the optimization algorithm to reach lower minima. A likely – though not further investigated – reason could be that in a number of samples the factorization in FDS "badly" approaches the initialization to a better solution, while the threading method better approaches the initialization to a solution which is finally worse. This situation is illustrated in Figure 5.5 by means of an arbitrary one-dimensional cost $f(x)$. An optimization algorithm initialized with $\tilde{x}_1$ will converge to $\hat{x}_1$ and initialized with $\tilde{x}_2$ it converges to $\hat{x}_2$, but $f(\hat{x}_1) < f(\hat{x}_2)$ although $f(\tilde{x}_1) > f(\tilde{x}_2)$.

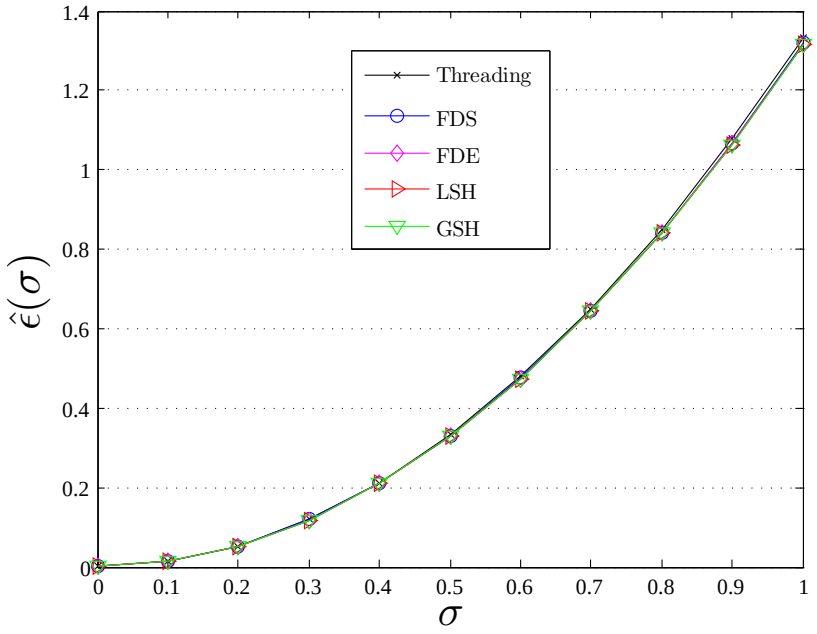


(a)

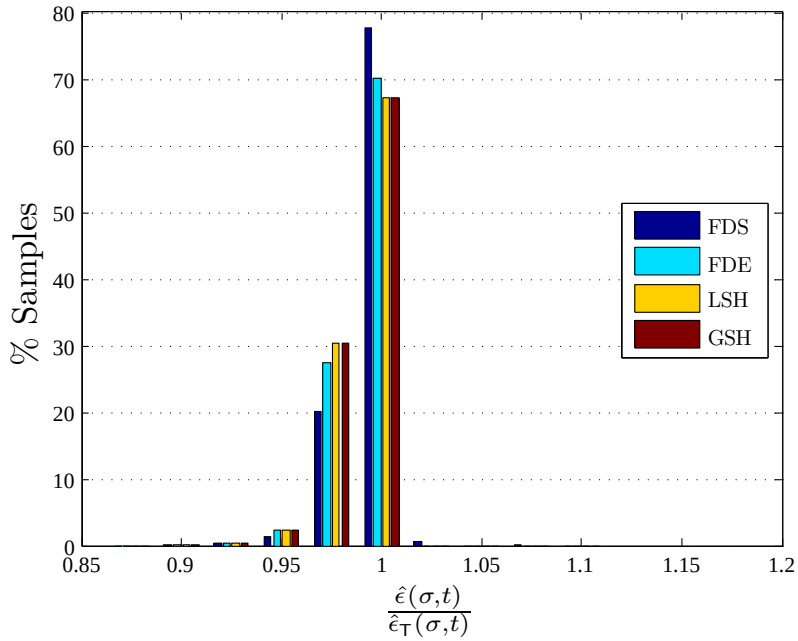


(b)

Figure 5.3: *(a) Plot of the RMSR resulting from the initialization parameters before the BA. The RMSR of the threading solution (black +) is much higher than the RMSR of the FDE (green ○) which is the next one below. Even lower are only LSH (red ●) and GSH (pink □), both almost with identical values. The worst values are achieved by FDS (blue ◇) which is again even worse than the threading type initialization.*
(b) Standard deviation of the RMSR before the bundle adjustment. The mean variation for FDE (green ○), LSH (red ●), and GSH (pink □) stay considerably small in regard to FDS (blue ◇) and – even worse – the threading (black +).



(a)



(b)

Figure 5.4: *(a) RMSR achieved with the different initializations after the bundle adjustment.*
(b) Distribution of RMSR after bundle adjustment relative to the one achieved with threading.

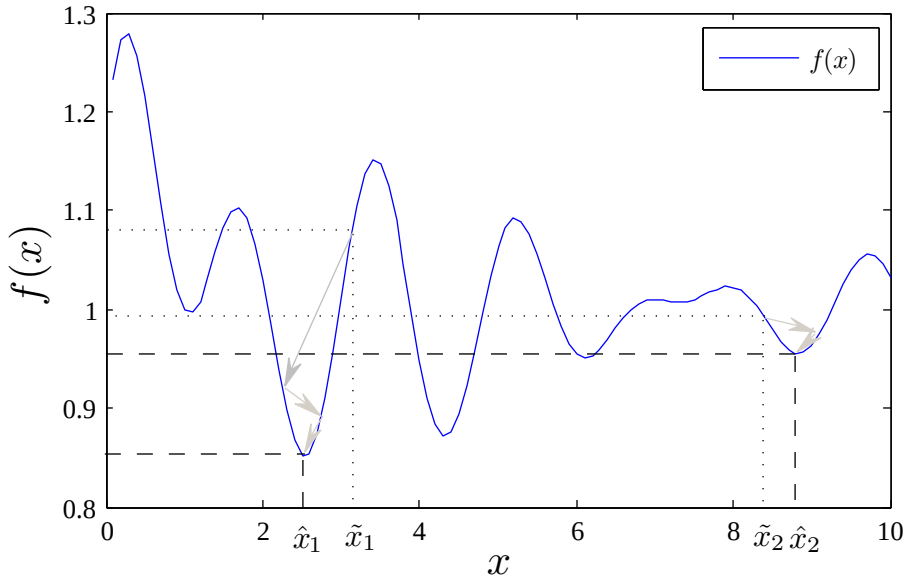


Figure 5.5: *An arbitrary one-dimensional cost function $f(x)$ is given and is to be minized. The initialization with $\tilde{x}_1$ leads the optimization algorithm to the local minimum at $\hat{x}_1$, the initialization $\tilde{x}_2$ lets the optimization converge to the local minimum at $\hat{x}_2$.*

5.3.3 Strong Projective and Sparse Data Experiment (SPSDE)

In this third synthetic experiment, the stress level for the different initialization methods compared herein has been heavily augmented. The maximum camera rotation angle per axis α was chosen as $\alpha = \frac{\pi}{8}$, so that the projective part of the transformation would become more relevant than it was in previous experiments. The field of view was adapted, so that most of the generated synthetic scenes would allow creating a connected frame graph; if the viewpoints generated for a scene wouldn't do so, a new scene would be generated. Additionally the amount of points in the scene was strongly reduced to 10^4 and only $n = 50$ images were made of each scene. The levels of noise were extended up to $\sigma = 1.5$

Under these conditions, all initializations become much worse in terms of the RMSR. The threading type initialization though suffers the most of these hard conditions and yields very poor results as can be seen in Figure 5.6(a). The values evaluated for $\tilde{\epsilon}_{\mathsf{LSH}}$ and $\tilde{\epsilon}_{\mathsf{GSH}}$ stick together just as they did in any other experiment. This is also the first experiment in which the bundle adjustment initialized with the threading solution could clearly not keep up with the results it achieved when it was initialized with parameters determined by the closed form solutions, as reflects the plot in Figure 5.6(b).

An additional look at the histogram displayed in Figure 5.7 reveals that final alignments arising from bundle adjustment initializations with FDS, FDE, and at most LSH and GSH, brought along improvements for most of the samples, and even quite considerable improvements for an important part of them, in regard to the results achieved with the threading type initialization.

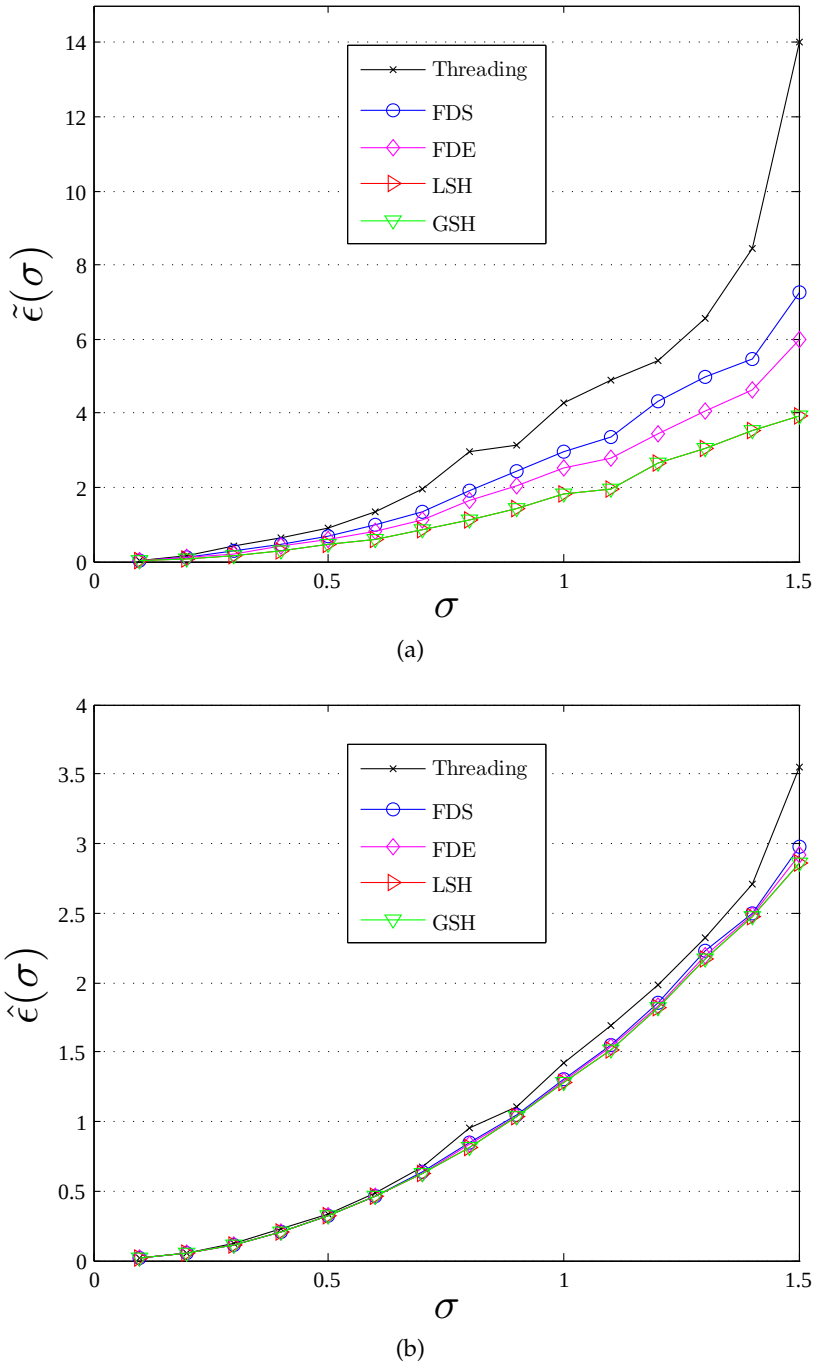


Figure 5.6: *Average RMSR achieved in SPSDE before (a) and after (b) the bundle adjustment. The results achieved with the threading solution (black $\times$) are clearly not as good as the results achieved by any other initialization method.*

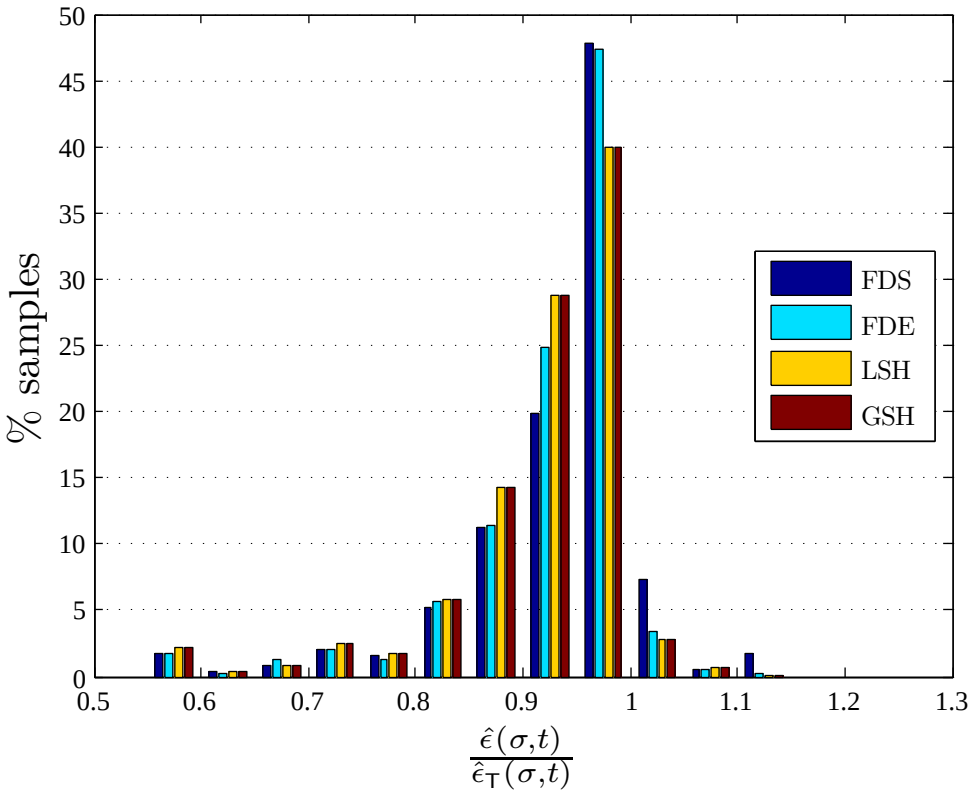


Figure 5.7: *Distribution of the relative RMSR after the BA in regard to the RMSR achieved with an initialization provided by the threading type method. In less than* 10% *of the samples made in this experiment, the closed-form solutions performed better than the threading solution. In at least* 50% *(*40%*) of the samples the RMSR achieved with LSH or GSH (FDS or FDE) was more than* 5% *below the RMSR achieved with the threading type method.*

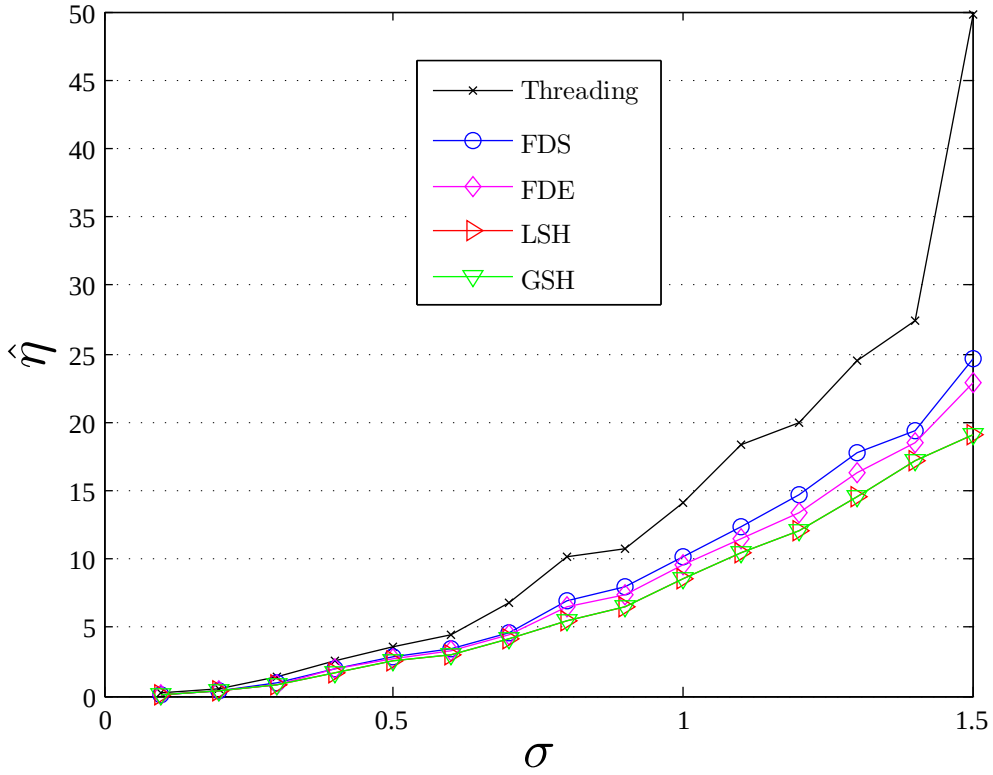


Figure 5.8: *Mean Reprojected Image Corner Distance (Section 5.2) achieved during SPSDE.*

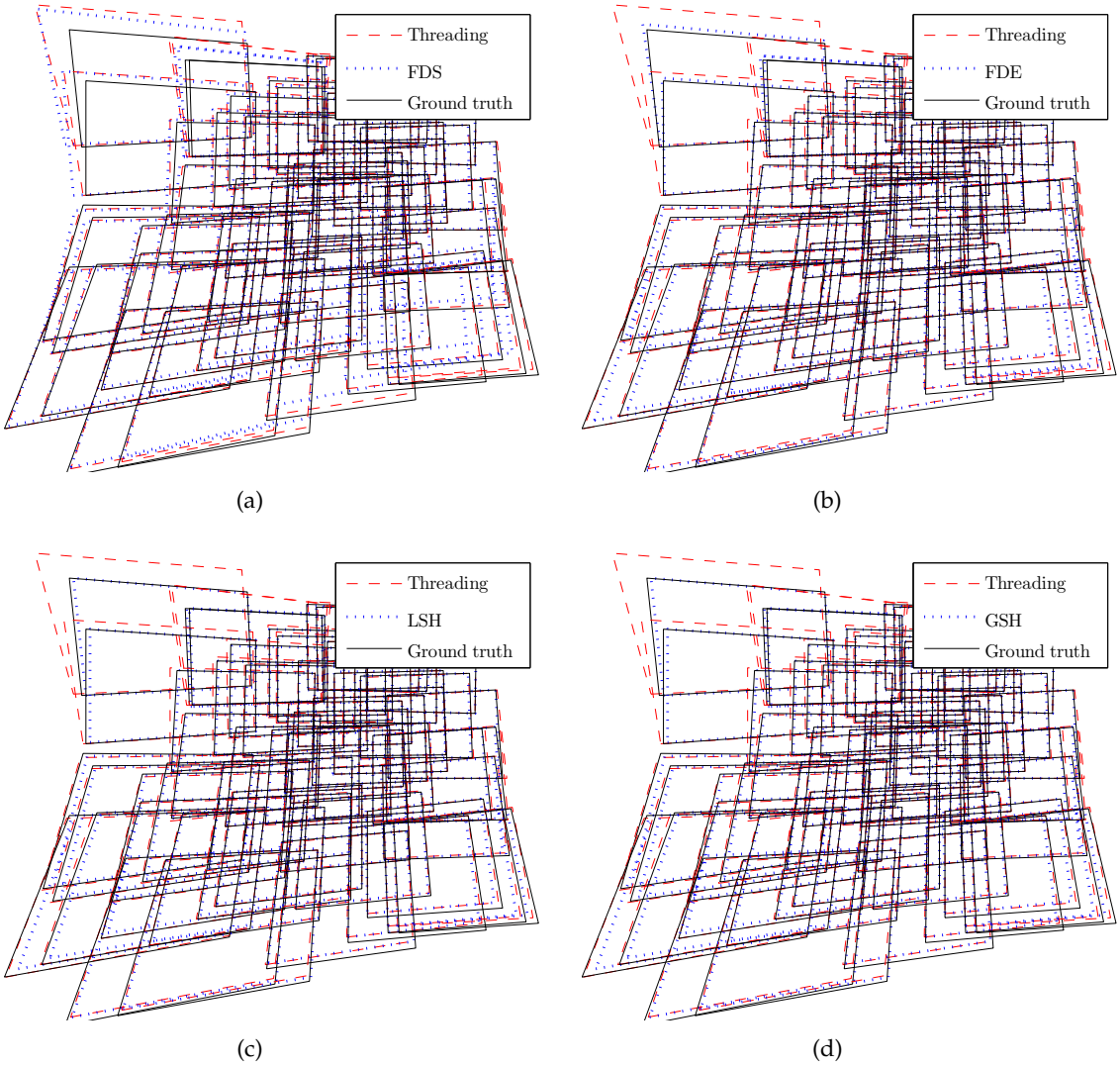


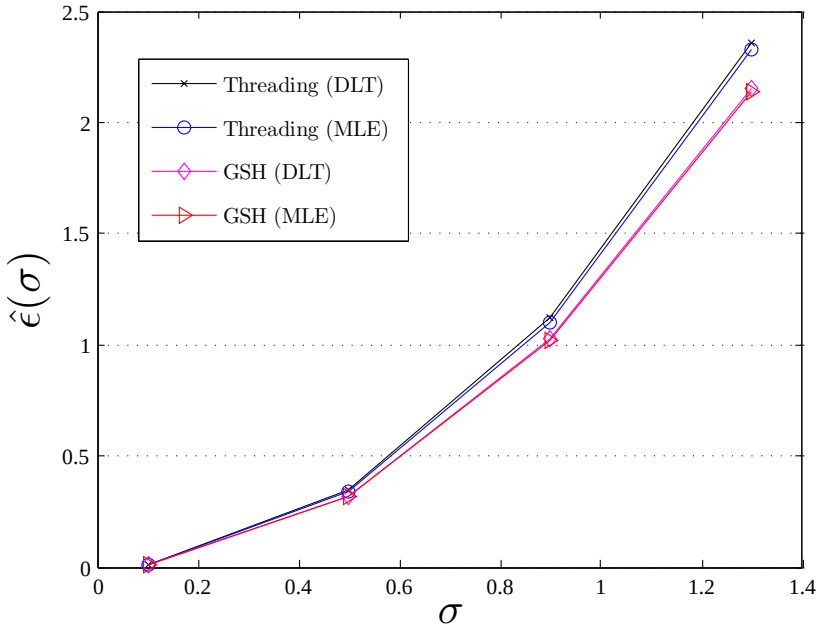
Figure 5.9: *Example of projections of every frame border in respect to the same reference frame, overlaying the ground truth, the threading result, and – in each case – the results of another closed-form solution, all after the bundle adjustment.*

In contrast to LPFDE and LPSDE, $\hat{\eta}$ has also been evaluated. (Again, it should be pointed out, that this measurand does not provide any theoretically founded measure, but is only intended to provide a rough impression, which one of the solutions yields alignments which better reflect the ground truth!) As it can be seen in Figure 5.8, $\hat{\eta}_{\mathsf{T}}$ is by far higher than the values evaluated for FDS, FDE, LSH, and GSH, whereas LSH and GSH provide both similarly the lowest distances to the ground truth. Figure 5.9 tries to better illustrate the matter about the Mean Reprojected Image Corner Distance. In regard to a single reference frame it overlays the reprojection of the frame borders computed with different closed-form solutions with the threading result and the ground truth. At the upper left most image, a significant discrepancy can be noticed between the threading and FDS versus the ground truth borders (Figure 5.9(a)), whereas the alignment resulting from initializations with FDE, LSH, or GSH reflect much better the ground truth of that image (5.9(b), 5.9(c), and 5.9(d) respectively).

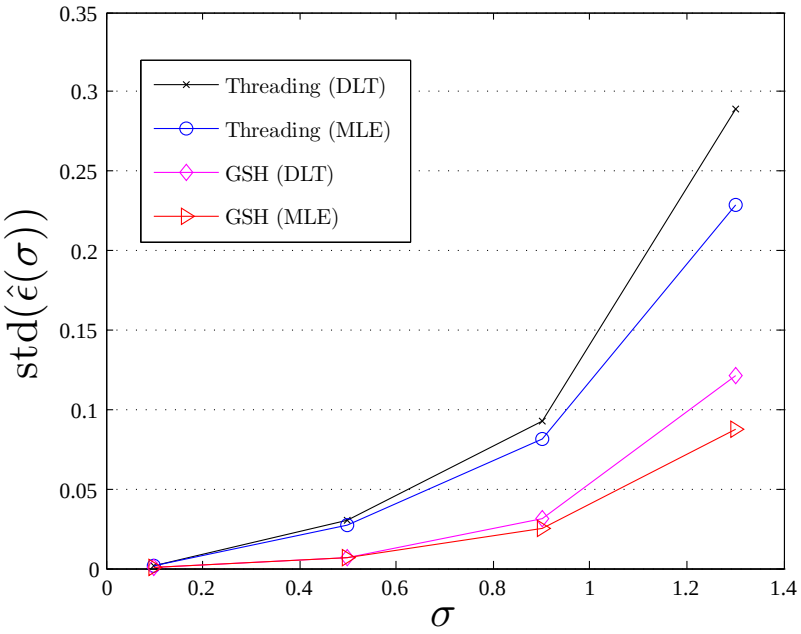
5.3.4 Extended Strong Projective and Sparse Data Experiment (ESPSDE)

At some point, the question came up, whether the closed-form solutions would be able to compensate the inaccuracy of the DLT and allow to find a final global alignment only with local homographies computed with the DLT, but which is not significantly worse than the one which could be found with the MLE of the local homographies. Therefore a small, last experiment was made which used similar synthetic data as it had been used in SPSDE, but with the difference, that only the threading type method based on DLT, the threading type method based on MLE, GSH based on DLT, and GSH based on MLE homographies were evaluated.

Figure 5.10(a) reveals, that both, the threading type and the GSH solution, yield more likely alignments when the MLE are used as input data, but still GSH initializations with DLT homographies converges to a more likely alignment than the threading method based on MLE of the local homographies does. Even though GSH (DLT) performs almost as good as GSH (MLE), the standard deviation of the RMSR is clearly higher for the former one.



(a)



(b)

Figure 5.10: *(a) RMSR achieved after the BA. Threading solution vs. GSH; each of them with DLT vs. MLE initialization. Although the MLE initialized methods perform better than their DLT counterparts, GSH with DLT yields still a better alignment as the threading type solution with MLE.*
(b) The standard deviation of the RMSR obviously is not as low with the DLT initializations as it is with the MLE.

Chapter 6

Real Data Example

Although the previous simulations revealed, that the closed-form solution can compete at all the compared threading type method and which one of them to what extent, it is always possible that – for many reasons – some assumptions which the theoretical reasoning and the data syntheses rely on, might lack important details and thus not suffice to model the real system. Therefore a real experiment had to be made in order to verify that the solutions can handle real data too.

The scene was a city map pinned to a wall which was recorded with a camera mounted on a tripod in front of the wall. The camera was calibrated only in order to undistort the video sequence.

6.1 Implementation Details

The video sequence was then searched for Scale Invariant Feature Transform (SIFT) points [Low04] frame by frame. For each pair of images, these key points were then matched the way Lowe proposed in his paper. For the SIFT point detection and the matching of those, Vedaldi's implementation for Matlab has been used [Ved09].

Outliers in the point correspondences were then removed using the Random Sample Consensus (RANSAC) paradigm [FB81] applied with the DLT algorithm for 2D- homography estimation [HZ04]. When a robustly matched set of correspondences could be found for a pair of images, the local homography was estimated using the DLT in order to retrieve an initial guess, followed by a non-linear optimization yielding the MLE of the homography. For the robust homography estimation, Kovesi's implementation for MATLAB [Kov09] was used and extended by a non-linear optimization.

Last but not least, the BA, which had already been used for the synthetic experiments, was initialized using the different initialization methods and for each resulting alignment the RMSR (Section 5.2) was evaluated.

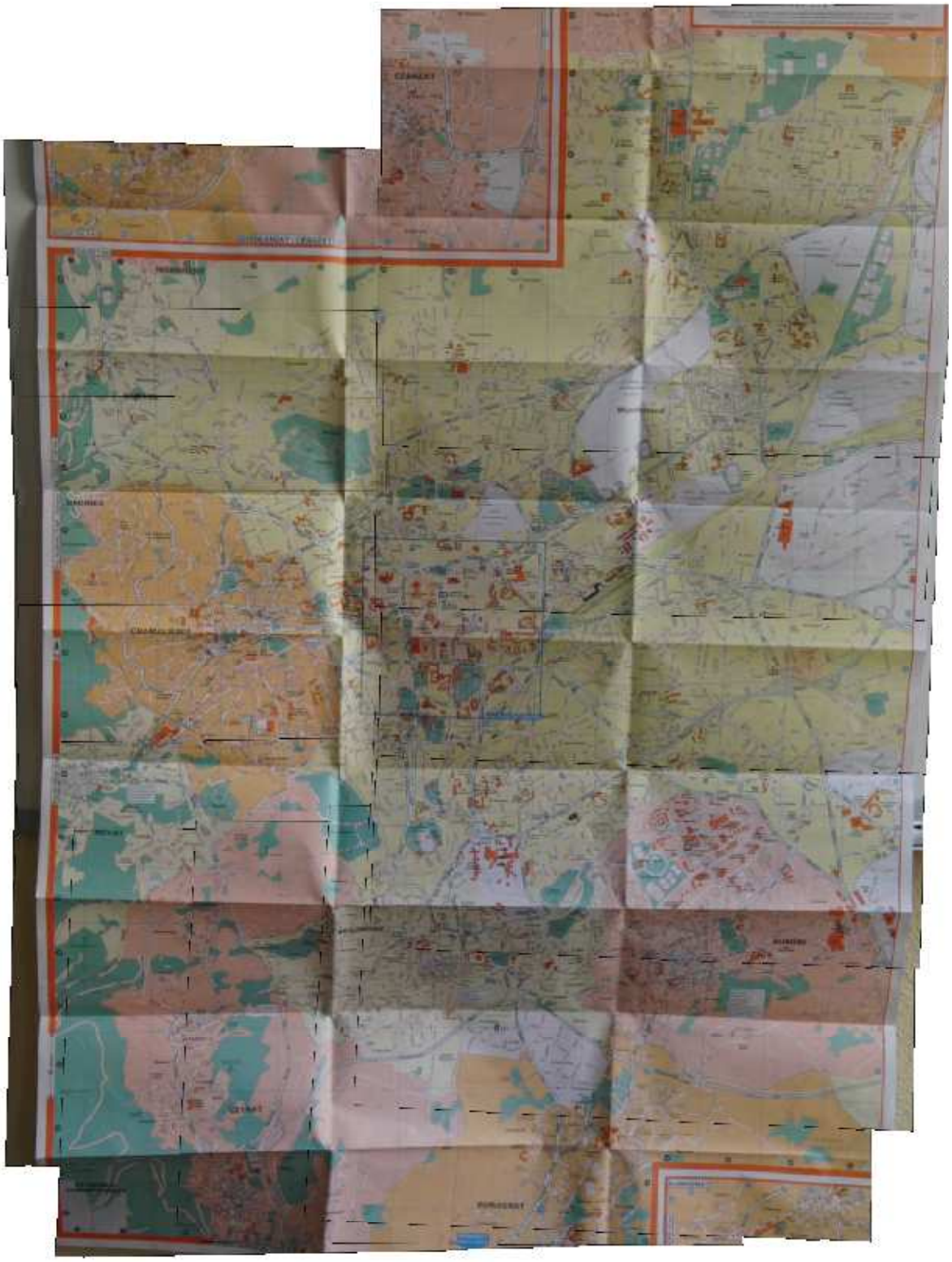


Figure 6.1: *Mosaic of the real video sequence, aligned using an initialization provided by GSH. Only each tenth frame was used for this mosaic.*

6.2 Results

For this video sequence, which consisted of $n = 167$ frames with a resolution of 640×480 pixels, 46.56% of the homographies failed to be estimated and 20.36% of the frames could not be directly aligned to the reference frame (used by the threading method). In average an image was connected to 89 ± 48 other images of the sequence.

The RMSR of the final alignments of the different methods were determined and the relative RMSR compared to the threading type solution was

$$\begin{aligned}
\frac{\hat{\epsilon}_{\mathsf{FDS}}}{\hat{\epsilon}_{\mathsf{T}}} &= 1.0009 \\
\frac{\hat{\epsilon}_{\mathsf{FDE}}}{\hat{\epsilon}_{\mathsf{T}}} &= 0.9979 \\
\frac{\hat{\epsilon}_{\mathsf{LSH}}}{\hat{\epsilon}_{\mathsf{T}}} &= 0.9980 \\
\frac{\hat{\epsilon}_{\mathsf{GSH}}}{\hat{\epsilon}_{\mathsf{T}}} &= 0.9980.
\end{aligned}$$

That is to say, each of the initializations computed with the proposed closed-form solutions but FDS allowed to find a slightly better minimum as the threading initialization did.

Even though it is impossible to determine the ground truth in real examples as for the one denoted here, an overlay of the plots of the frameborders found with different initialization methods reveals, that the impact of the different methods can visually be bigger as one might expect it to be after considering the RMSR only (either relative or absolute). In Figure 6.2 the plots of the frameborders found after the BA initialized once with the threading solution and once with GSH are overlaid and aligned to the same image's reference frame. At the bottom of the plot one may notice the discrepancies between the differently detected alignments which differ from each other by some pixels.

In regard to the observation made during the synthetic experiment described in Section 5.3.3 concerning the Mean Reprojected Image Corner Distance, one can suppose that the results yielded by GSH better reflect the ground truth. To what extent and if this is really the case at all can unfortunately not be verified for real examples.

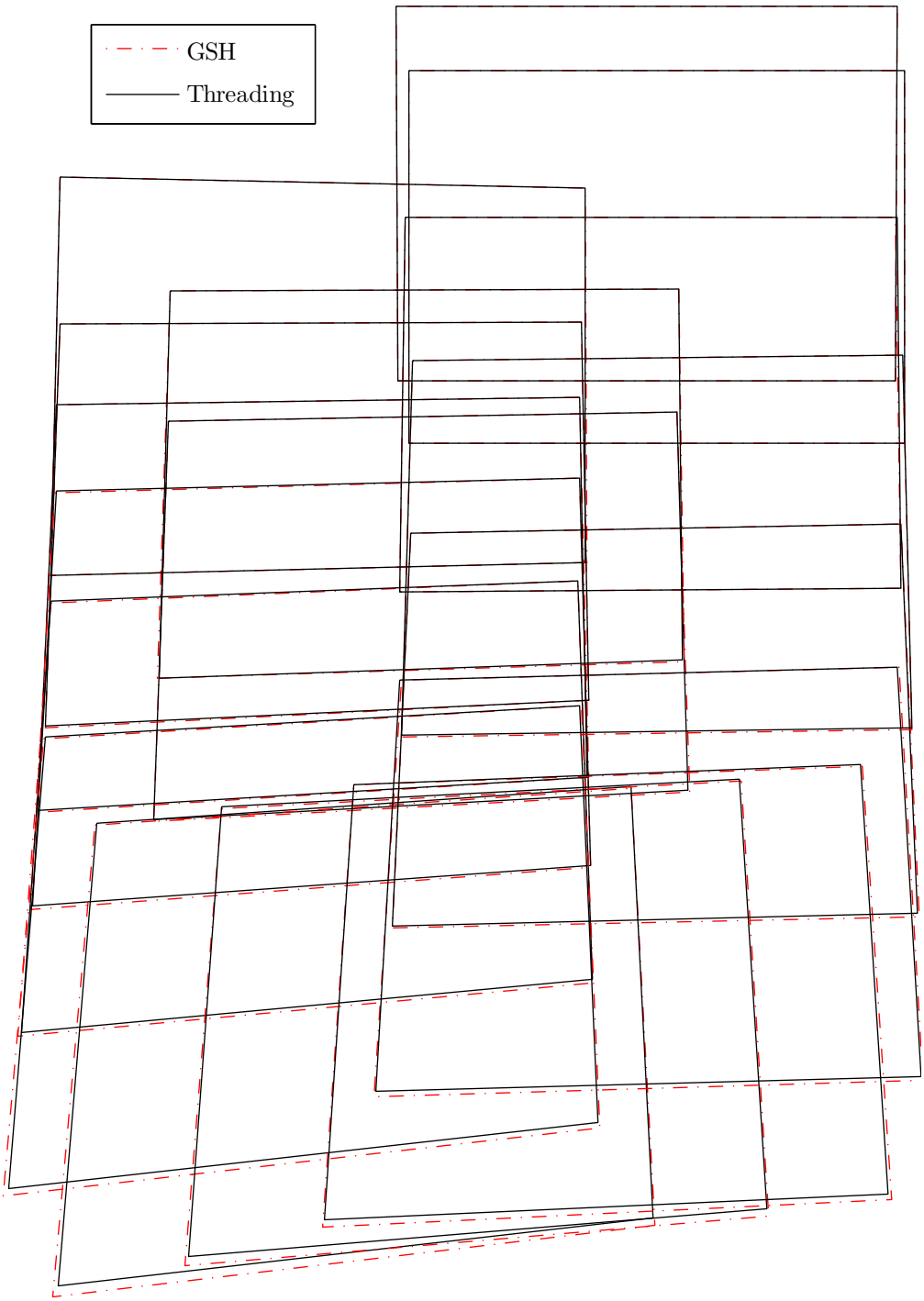


Figure 6.2: *Overlay plot of the frame borders found with Threading versus GSH after the BA. The frame borders of the reference frames (upper right) align well. At the very other extremity of the mosaic (lower left) a discrepancy of several pixels can be detected between the frame borders of the differently aligned mosaics.*

Chapter 7

Conclusion

During this diploma thesis, the potential of the different closed-form solutions for homography estimation, which have been introduced in this thesis, compared to a simple threading solution has been studied. Although the basic idea is not necessarily a new one, but contrary to the previous work this thesis investigates the application of the approach to the constrained scenario of a single projection center.

The introduced solutions provide an analytical approach based on statistical procedures which allow to extract redundantly available information from local inter-frame homographies. This redundant information is – in general – not being considered, and hence lost in threading type approaches. Threading solutions have the disadvantage that they have at some point to interpolate uncomputable homographies by composing known ones. In order to keep error propagation during this interpolation as low as possible, one of the main issues for those solutions, is to guess with, most commonly, heuristical ratings which homographies and combinations of homographies could be the ones suffering the least from noise. In contrast, the herein introduced solutions provide a means to directly estimate global alignments which best fit the measured inter-frame homographies according to proven concepts of statistics theory.

During the synthetic experiments, two major (FDS, GSH) and two minor solutions (FDE, LSH) (latter lead to the second of the major solutions) have been compared to a simple threading solution. The first of the major solutions (FDS) formulates the problem as a factorization task and unfortunately suffers from the missing data problem and requires interpolation of that missing data. The second one (GSH) represents an eigenvector problem and handles missing data implicitly and hence gets rid of one of the main issues of homography estimation without additional processing resources or the need to imagine reasonable ways to interpolate those missing data.

The experiments with synthetic scenarios showed that bundle adjustments initialized with the GSH guesses performed best of the closed-form solutions and also better compared to the threading approach in some specific situations. Namely situations in which the alignments between images contained a strong projective transformation and a certain amount of local homographies could not be determined.

According to the last experiment, it seems that the initialization with GSH could turn obsolete the need for the non-linear estimation of the local homographies and additionally the threading itself. Both aspects could save valuable CPU time in time critical applications, although the processing time impact on the overall process has not been investigated and might represent only a small portion of it, particularly with regard to the final BA.

Although this work represents only an initial study which, in a very first sight, answers mainly the question whether it is worth the effort to further investigate the potential of the proposed methods, one very important conclusion could be made: As it was extremely rare during the experiments that GSH did provide a worse initialization as the threading method, GSH is worth using even in situations where it cannot provide significantly better initializations because it turns the missing data interpolation obsolete without any notable additional implementation effort or supplementary processing resources.

On the one hand, this thesis has proven the feasibility of the proposed closed-form solutions and could identify the clear implementation benefits of using GSH (and to some extent LSH) in advance to simple threading algorithms: The closed-form solutions only require building block matrices straight forward and performing a SVD on them. The SVD algorithm itself should be included in almost any common linear algebra library. Contrarily, the threading solutions require the developer to create searching algorithms and explicit interpolation routines in order to get to a result.

On the other hand though, it raised new questions which couldn't be answered in the scope of this work unfortunately and had to be left for further investigation:
It would, for sure, now be of interest to what extent GSH could compete against more advanced threading techniques as, for example, the approach which Vergés-Llahí et al. [VLMW07] or Bajramovic and Denzler [BD08] are following, and which tries to identify reliable paths in the collineation graph by promoting local inter-frame homographies which approve each other and penalizing edges which might be outliers.

Another question which hasn't been treated at all herein would be, how the methods react in terms of outliers in the point correspondences (that is to say, falsely matched key points).

Furthermore, more synthetic and considerably more real experiments have to be driven in order to well compare different approaches to each other. This and the preceding statements require a better implementation (probably in C/C++) of the experimental environment though, as the currently used one was implemented in Matlab and required elevated resources for synthetic and quite exhaustive resources for real scenarios. A well designed framework could be of interest for plug-in driven experiments. That is to say, multiple initialization methods could simply be compared against each other by dynamically loading the according libraries providing the initializations for whole batches of experiments. This would most probably be not only of use for this work, but for others trying to solve the same problem too.

At the very end of this work, further suggestions came up to improve the closed-form solutions (which would obviously require further investigation). On the one hand, it would be nice to find a means to identify and simply drop outliers in the local homographies, but without drifting into a threading-type solution in order to keep implementation efforts as low as possible.

On the other hand, it would be interesting if more advanced, but still analytically and statistically founded weightings could lead to better initializations too. That is to say, altering the influence of the local homographies on the initializers for the global alignments, based on either the quality of features detected in the images [ZGS+09] or the quality of matchings [RDG08] or even including a measure for the "spreadness of point correspondences", number of correspondences, and "degree of overlap".

Bibliography

[Bar04] Adrien Bartoli. AirPhoto – The Bonn Archaeological Software Package. Technical report, The Unkelbach Valley Software Work, 2004.

[BD08] Ferid Bajramovic and Joachim Denzler. Global Uncertainty-based Selection of Relative Poses for Multi Camera Calibration. In *Proc. BMVC*, volume 2, pages 745–754, September 2008.

[Cap04] David Capel. *Image mosaicing and super-resolution*. Springer, 2004.

[CZ98] D. Capel and A. Zisserman. Automatic Mosaicing with Super-Resolution Zoom. *CVPR*, page 885, 1998.

[DD88] P. Dewilde and E. Deprettere. Singular value decomposition: an introduction. pages 3–41, 1988.

[EY36] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, September 1936.

[FB81] Martin A Fischler and Robert C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), Jun 1981. Notes Seminal RANSAC paper.

[Gov01] V. Govindu. Combining Two-View Constraints for Motion Estimation. In *Proc. CVPR*, 2001.

[Gov04] V. Govindu. Lie-Algebraic Averaging for Globally Consistent Motion Estimation. In *Proc. CVPR*, 2004.

[GVL96] Gene H. Golub and Charles F. Van Loan. *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[Har97] Richard I. Hartley. Self-Calibration of Stationary Cameras. *IJCV*, 22, February 1997.

[HZ04] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[IAB+96] Michal Irani, P. Anandan, Jim Bergen, Rakesh Kumar, and Steve Hsu. Efficient Representations of Video Sequences and Their Applications. In *Signal Processing: Image Communication*, pages 327–351, 1996.

[IHA95] Michal Irani, Steve Hsu, and P. Anandan. Video compression using mosaic representations. *SIGNAL PROCESS-IMAGE*, 7:529–552, November 1995.

[Jac97] David W. Jacobs. Linear Fitting with Missing Data: Applications to Structure-from-Motion and to Characterizing Intensity Images. In *CVPR*, pages 206–212, June 1997.

[KCM00] Eun-Young Kang, Isaac Cohen, and Gerard Medioni. A Graph-based Global Registration for 2D Mosaics. *ICPR*, pages 257–260, May 2000.

[Kov09] Peter Kovesi. MATLAB and Octave Functions for Computer Vision and Image Processing. `http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/`, June 2009.

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[MC02] Ezio Malis and Roberto Cipolla. Camera Self-Calibration from Unknown Planar Structures Enforcing the Multiview Constraints between Collineations. *PAMI*, 24(9):1268–1272, 2002.

[MFM04] Roberto Marzotto, Andrea Fusiello, and Vittorio Murino. High Resolution Video Mosaicing with Global Alignment. In *Proc. CVPR*, pages 692–698, 2004.

[NW99] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 1999.

[RDG08] Julien Rabin, Julie Delon, and Yann Gousseau. A contrario matching of SIFT-like descriptors. In *ICPR*, pages 1–4, 2008.

[SHK98] Harpreet S. Sawhney, Steve Hsu, and Rakesh Kumar. Robust Video Mosaicing through Topology Inference and Local to Global Alignment. In *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*, pages 103–119, London, UK, 1998. Springer-Verlag.

[SS01a] T. Schickinger and A. Steger. *Diskrete Strukturen II*. Springer, 2001.

[SS01b] H.-Y Shum and R. Szeliski. Construction of panoramic image mosaics with global and local alignment. pages 227–268, 2001.

[Stu00] Peter Sturm. Algorithms for Plane-Based Pose Estimation. In *Proc. CVPR*, pages 1010–1017, June 2000.

[TK92] Carlo Tomasi and Takeo Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *IJCV*, 9(2):137–154, Nov 1992.

[Ved09] Andrea Vedaldi. SIFT for Matlab. `http://www.vlfeat.org/~vedaldi/code/sift.html`, July 2009.

[VLMW07] Jaume Verges-Llahi, Daniel Moldovan, and Toshikazu Wada. A new Reliability Measure for Essential Matrices Suitable in Multiple View Calibration. *VISAPP*, page 8, Dec 2007.

[VLW08] Jaumes Vergés-Llahí and Toshikazu Wada. *A General Algorithm to Recover External Camera Parameters from Pairwise Camera Calibrations*, pages 294–304. Springer Berlin / Heidelberg, 2008.

[ZGS+09] B. Zeisl, P. Georgel, F. Schweiger, E. Steinbach, and N. Navab. Estimation of Location Uncertainty for Scale Invariant Feature Points. In *BMVC*, 2009.

Appendix A

Matrix Approximation

According to the Eckart-Young theorem[1] [EY36], a $n \times n$ matrix $\mathtt{M}$ is best approximated in terms of the Frobenius norm with a matrix $\tilde{\mathtt{M}}$ of $\mathrm{rank}(\tilde{\mathtt{M}}) = r$ if

$$\tilde{\mathtt{M}} = \mathtt{U}\tilde{\Sigma}\mathtt{V}^{\mathsf{T}} \tag{A.1}$$

where

$$\mathtt{U}\Sigma\mathtt{V}^{\mathsf{T}} \overset{\mathsf{SVD}}{\longleftarrow} \mathtt{M} \tag{A.2}$$

with

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_n \end{bmatrix}, \qquad \text{s.t. } \sigma_1 \geq \dots \geq \sigma_n \geq 0. \tag{A.3}$$

and

$$\tilde{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & & \dots & & 0 \\ 0 & \ddots & & & & \\ & & \sigma_r & \ddots & & \vdots \\ \vdots & & \ddots & 0 & & \\ & & & & \ddots & 0 \\ 0 & & \dots & & 0 & 0 \end{bmatrix}. \tag{A.4}$$

Proof $\|\mathtt{M} - \tilde{\mathtt{M}}\|$ is to be minimized. Because the column vectors of $\mathtt{U}$ and $\mathtt{V}$ are unitary vectors

$$\arg\min_{\tilde{\mathtt{M}}} \|\mathtt{M} - \tilde{\mathtt{M}}\| \equiv \arg\min_{\tilde{\mathtt{M}}} \|\Sigma - \mathtt{U}^{\mathsf{T}}\tilde{\mathtt{M}}\mathtt{V}\| \tag{A.5}$$

Since Σ is diagonal, $\mathtt{U}^{\mathsf{T}}\tilde{\mathtt{M}}\mathtt{V}$ has to be so too in order to minimize the Frobenius norm, and thus $\mathtt{U}$ and $\mathtt{V}$ are also singular matrices of $\tilde{\mathtt{M}}$:

$$\tilde{\mathtt{M}} = \mathtt{U}\mathtt{S}\mathtt{V}^{\mathsf{T}} \tag{A.6}$$

[1] The proof Eckart and Young gave, is only valid for square matrices, but considered sufficient for this work, as only decompositions of square matrices are made herein.

with $\mathtt{S}$ diagonal with entries s_i.

As a consequence

$$\arg\min_{\mathtt{S}} \|\Sigma - \mathtt{S}\| \equiv \arg\min_{s_i} \sqrt{\sum_{i=1}^{r}(\sigma_i - s_i)^2 + \sum_{i=r+1}^{n} \sigma_i^2}. \tag{A.7}$$

That is to say, $\|\mathtt{M} - \tilde{\mathtt{M}}\|$ is minimized if $\forall i \in [r] : s_i = \sigma_i$ and $\forall i \in [n]\backslash[r] : s_i = 0$. [DD88, GVL96]

□

Appendix B

Orthonormal Linear Least Squares Minimization

Subject to this section is to prove the following statement. The proof given herein is a generalization of a common proof that the solution to $\underset{\mathbf{x}|\mathbf{x}^\mathsf{T}\mathbf{x}=1}{\arg\min}\|\mathtt{A}\mathbf{x}\|^2$ is given by the right singular vector of $\mathtt{A}$ corresponding to its smallest singular value. Latter is implicitly included as a special case of this generalized proof.

Lemma Let $\mathtt{A}$ be a $v \times w$ matrix with real entries and let $\mathtt{X}$ be the $w \times n$ matrix with $n \leq \min\{v, w\}$ which minimizes $\|\mathtt{AX}\|_2^2$ under the constraint that $\mathtt{X}^\mathsf{T}\mathtt{X} = \mathtt{I}$, then the n coloumns of $\mathtt{X}$ are made up of the n right singular vectors of $\mathtt{A}$ corresponding to the latter's n smallest singular values.

Proof

$$\underset{\mathtt{X}|\mathtt{X}^\mathsf{T}\mathtt{X}=\mathtt{I}}{\arg\min}\,\|\mathtt{AX}\|_2^2 \equiv \underset{\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{x}_i^\mathsf{T}\mathbf{x}_i=1,\forall i,j,i\neq j:\mathbf{x}_i^\mathsf{T}\mathbf{x}_j=0}{\arg\min} \sum_{k\in[n]} \|\mathtt{A}\mathbf{x}_k\|_2^2 \tag{B.1}$$

The lagrangian to this problem which encodes the norm- and the pairwise orthogonality-constraint on $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is given by

$$\mathcal{L} = \sum_{k\in[n]} \mathcal{L}_k \tag{B.2}$$

with

$$\forall k \in [n]: \; \mathcal{L}_k = \|\mathtt{A}\mathbf{x}_k\|_2^2 + \lambda_k(1 - \mathbf{x}_k^\mathsf{T}\mathbf{x}_k) + \sum_{r\in[n],r\neq k} \mu_{k,r}(\mathbf{x}_k^\mathsf{T}\mathbf{x}_r)^2 \tag{B.3}$$

which encode the problem in regard of the terms $\|\mathtt{A}\mathbf{x}_1\|_2^2, \ldots, \|\mathtt{A}\mathbf{x}_n\|_2^2$.

$\sum_{k\in[n]} \|\mathtt{A}\mathbf{x}_k\|_2^2$ can only be at a minimum if

$$\forall p \in [n]: \; \frac{\partial \mathcal{L}}{\partial \mathbf{x}_p} = 0 \tag{B.4}$$

and as $\forall k \in [n] : \mathcal{L}_k \geq 0$ is equivalent to

$$\forall p, k \in [n] : \frac{\partial \mathcal{L}_k}{\partial \mathbf{x}_p} = 0. \tag{B.5}$$

Deriving the lagrangian, resp. particularly the terms of the lagrangian yields on the one hand side

$$\forall p, k \in [n],\, p \neq k : \frac{1}{2}\frac{\partial \mathcal{L}_k}{\partial \mathbf{x}_p} = \mu_{k,p}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{x}_p = 0 \tag{B.6}$$

and on the other hand side

$$\forall k \in [n] : \frac{1}{2}\frac{\partial \mathcal{L}_k}{\partial \mathbf{x}_k} = \mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k - \lambda_k\mathbf{x}_k + \sum_{r\in[n],r\neq k} \mu_{k,r}\mathbf{x}_r\mathbf{x}_r^\mathsf{T}\mathbf{x}_k = \mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k - \lambda_k\mathbf{x}_k. \tag{B.7}$$

Both, B.6 and B.7 together with B.5 (resp. B.4) yield that $\sum_{k\in[n]} \|\mathtt{A}\mathbf{x}_k\|_2^2$ can only be minimal if

$$\forall k \in [n] : \mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k = \lambda_k\mathbf{x}_k. \tag{B.8}$$

That is to say, that $\sum_{k\in[n]} \|\mathtt{A}\mathbf{x}_k\|_2^2$ can only be minimal if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are eigenvectors of the matrix $\mathtt{A}^\mathsf{T}\mathtt{A}$ and $\lambda_1, \ldots, \lambda_n$ their corresponding eigenvalues. This insight in regard of the problem statement B.1

$$\underset{\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{x}_i^\mathsf{T}\mathbf{x}_i=1,\mathbf{x}_i^\mathsf{T}\mathbf{x}_j=0}{\arg\min} \sum_{k\in[n]} \mathbf{x}_k^\mathsf{T} \underbrace{\mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k}_{\lambda_k\mathbf{x}_k} \equiv \underset{\lambda_1,\ldots,\lambda_n|\mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_i=\lambda_i\mathbf{x}_i}{\arg\min} \sum_{k\in[n]} \lambda_k \tag{B.9}$$

reveals that the problem is equivalent to the problem of finding the n eigenvectors corresponding to the n smallest eigenvalues of $\mathtt{A}^\mathsf{T}\mathtt{A}$, which – in return – are the n right singular vectors corresponding to the n smallest singular values of $\mathtt{A}$.

□