Closed-Form Solutions to Multiple-View Homography Estimation

Pierre Schroeder[1] schroedp@in.tum.de
Adrien Bartoli[2] adrien.bartoli@gmail.com
Pierre Georgel[3] pierre.georgel@gmail.com
Nassir Navab[1] navab@in.tum.de

[1]Chair for Computer Aided Medical Procedures & Augmented Reality, Technische Universität München, Germany
[2] Image Science for Interventional Techniques, Université d'Auvergne, Clermont-Ferrand, France
[3]UNC Chapel Hill, Department of Computer Science, USA

Abstract

The quality of a mosaic depends on the projective alignment of the images involved. After point-correspondences between the images have been established, bundle adjustment finds an alignment considered optimal under certain hypotheses. This procedure minimizes a nonlinear cost and has to be initialized with care. It is very common to compose inter-frame homographies which have been computed with standard methods in order to get an initial global alignment. This technique is suboptimal if there is noise or missing homographies as it typically uses a small part of the available data.

We propose four new closed-form solutions. They all provide non-heuristic initial alignments using all the known inter-frame homographies. Our methods are tested with synthetic and real data and are compared to the standard method. These experiments reveal that our methods are more accurate, taking advantage of the redundant information available in the set of inter-frame homographies.

1. Introduction

In computer vision, many applications require panoramic stitching [18] from a collection of images or frames from a video. This technique allows one to create for instance wide-angle recordings without using special hardware such as fish-eye lenses. Panoramic stitching is also used in more advanced applications such as super-resolution imaging [4, 15], video compression [10], and camera auto-calibration [8, 14].

Creating a mosaic from only two overlapping images is a relatively easy task and standard techniques provide very good results [3]. Unfortunately aligning multiple images is more complicated, particularly if some input images do not overlap.

We propose a means to linearly extract and entirely use the redundantly contained information from inter-frame homographies in order to better initialize the final bundle adjustment.

Paper organization. A brief introduction to the mathematical background of stitching is given in Section 2. We discuss prior work (threading and batch methods) in Section 3. We then introduce in Section 4 our proposed methods. They fall into two categories: those which require all the inter-frame homographies and those which handle missing information. We experimentally show the improvements brought by our methods in Section 5 and provide a conclusion in Section 6.

Notation. $\mathbb{P}^2$ represents the 2D-projective space and $\sim$ equality up to scale. Vectors are denoted using bold fonts. In general but not exclusively, small letters (e.g. $\mathbf{q}$) refer to homogeneous point coordinates in images and capitals ($\mathbf{Q}$) represent point coordinates in 3D; matrices are denoted with type-writer capitals (e.g. $\mathtt{M}$). $\|\cdot\|$ refers to the standard two-norm when used for vectors and to the Frobenius norm if used for matrices.

2. Theoretical Background

In order to create a mosaic, it is necessary to warp an image from its own to the mosaic's coordinate frame. A projective camera projects a point $\mathbf{Q} \in \mathbb{R}^3$ in space to a point $\mathbf{q} \in \mathbb{P}^2$ in the image plane as:

$$\mathbf{q} \sim \mathtt{K}\mathtt{R}\mathbf{Q}, \tag{1}$$

with $\mathtt{K}$ the matrix of intrinsic parameters and $\mathtt{R}$ the rotation matrix specifying the camera orientation. (We assume that the origin coincides with the centre of projection; hence there is no translational component.) A point $\mathbf{q}^i$ from image $\mathcal{I}_i$ is related to its corresponding point $\mathbf{q}^j$ in image $\mathcal{I}_j$ by:

$$\mathbf{q}^j \sim \mathtt{K}\mathtt{R}_j\mathtt{R}_i^\mathsf{T}\mathtt{K}^{-1}\mathbf{q}^i. \tag{2}$$

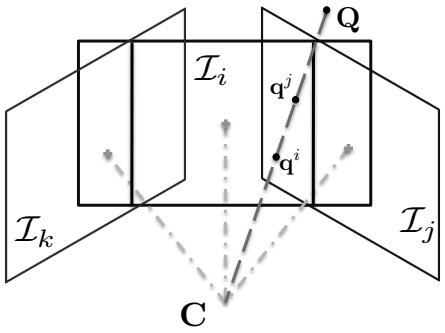


Figure 1. Sweep of images with a static center of projection **C** (no parallax): the real 3D point **Q** projects to $\mathbf{q}^i$ in image $\mathcal{I}_i$ and its corresponding point $\mathbf{q}^j$ in $\mathcal{I}_j$, but is not visible in $\mathcal{I}_k$.

This is illustrated in Figure 1. Thus, an inter-frame homography is given by

$$\mathtt{H}_{i,j} \sim \mathtt{P}_j \mathtt{P}_i^{-1}, \ \forall k : \mathtt{P}_k \sim \mathtt{K}\mathtt{R}_k \tag{3}$$

from which we derive the *consistency relationship*

$$\mathtt{H}_{i,j} \sim \mathtt{P}_j \mathtt{P}_k^{-1} \mathtt{P}_k \mathtt{P}_i^{-1} \sim \mathtt{H}_{k,j} \mathtt{H}_{i,k}. \tag{4}$$

A homography between two images can easily be estimated from point correspondences using standard procedures [9]. Assuming that the image points are perturbed by Gaussian noise, these methods provide Maximum Likelihood Estimates (MLE) $\hat{\mathtt{H}}_{i,j}$ for each overlapping pair of images $(\mathcal{I}_i, \mathcal{I}_j)$. These methods generally ignore the additional information from the remaining images in the set and under presence of noise the consistency relationship (4) is violated, thus in general $\hat{\mathtt{H}}_{i,j} \not\sim \hat{\mathtt{H}}_{k,j} \hat{\mathtt{H}}_{i,k}$. That is to say, two images may align well to a reference image but may at the same time be badly aligned with each other.

In other words, aligning n images $\mathcal{I}_1, \ldots, \mathcal{I}_n$ to a (real or virtual) reference image in order to create a mosaic, requires a set of homographies $\mathtt{U}_1, \ldots, \mathtt{U}_n$ which by construction *consistently* aligns *all* images to a common reference frame. These global homographies respect the relationship:

$$\mathtt{H}_{i,j} \sim \mathtt{U}_j \mathtt{U}_i^{-1}. \tag{5}$$

As we assume an independent and identically distributed noise model for the image points, the optimal global homographies can be found by minimizing

$$\mathcal{C}_{\mathtt{BA}}(\hat{\mathbf{q}}_1, \ldots, \hat{\mathbf{q}}_m, \hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \sum_{z=1}^{m} \sum_{i=1}^{n} \delta_{z,i} d^2(\mathbf{q}_z^i, \hat{\mathtt{U}}_i \hat{\mathbf{q}}_z) \tag{6}$$

which represents the reprojection error and

$$\delta_{z,i} = \begin{cases} 1 & \text{if } \mathbf{q}_z^i \text{ exists} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

In order to solve this problem, one has to non-linearly optimize the cost [9] which is then called Bundle Adjustment (BA). The results of non-linear optimization though depend on the initialization parameters provided to the optimization algorithm [16]. The latter can only yield reasonable results if they are being run in a specific trust region in order to converge at all and a good initialization increases the probability that the algorithm converges to a better local minimum. Additionally, BA becomes quite resource intensive for larger problems and an initialization closer to the minimum which the algorithm will converge to may reduce the number of necessary iterations. We hereafter introduce four analytical batch methods which provide initial parameters for the bundle adjustment and make use of all known inter-frame homographies.

3. State of the Art

For this paper we suppose the image set topology graph known. The nodes of this graph consist of the images and its edges represent the known inter-frame homographies. The known homographies are the ones computed either with feature-based methods [3] or direct methods [2] between the input images.

3.1. Threading Type Methods

Threading methods treat the problem of finding an initialization based on local homographies as a searching algorithm and generally determine a reference image $\mathcal{I}_r$ from the set of images. The coordinate frame of $\mathcal{I}_r$ acts as the mosaic coordinate frame. Thus we see from Equation (5) that the global homographies $\mathtt{U}_i$ which align each image $\mathcal{I}_i$ to the mosaic are $\forall i : \mathtt{U}_i = \mathtt{H}_{r,i}$ since $\mathtt{U}_r = \mathtt{I}$. $\mathcal{I}_r$ is not chosen arbitrarily: it is either chosen as the image for which each homography to the other images in the set is known or, if such an image does not exist, the reference is chosen as an image with a maximum number of connections to other images in the image set topology. In the latter case, missing homographies must be hallucinated from known ones, following Equation (4). More complex threading methods are based on finding paths in the topology graph. The corner stone of all threading methods is the cost they impose on edges and paths: Kang et al. accumulate the residual error [11], Marzotto et al. impose constraints on the ratio of overlapping area and residual error [15], and Verges-Llahi et al. and Bajramovic and Denzler use notion of uncertainty propagation for Structure from Motion (SfM) [20, 1].

3.2. Batch Methods

Batch approaches find a global solution by linearly approximating the motion model. For instance, Govindu uses such an approach for SfM. In order to build a linear system, he uses quaternions [6] or Lie algebra [7].

Trying to solve a similar SfM problem, Sturm [17] uses homographies induced by planes as input. In order to com-

pute the initial rotation and translation for his nonlinear optimization, he uses a factorization based approach which suffers from the missing data problem, which he solves by averaging known rotations. Zelnik-Manor and Irani [21] combine inter-frame homography estimates and multi-view subspace constraints derived from the underlying 3D-planar structures to increase the accuracy of the computed global homographies.

Malis and Cipolla [14] impose the constraints in collineation matrices of a planar structure in multiple views for efficient camera self-calibration, mosaicing, and reconstruction in order to reduce geometric errors.

Contrarily to the problem we tackle in this paper, previous works use other constraints, as they additionally assume that the camera translation does not vanish between the different images.

4. Proposed Closed-Form Solutions

In order to build the linear system which is to be solved in Section 4.2 and 4.3, we require the homographies to represent a group with respect to standard matrix multiplication which is created in Section 4.1.

4.1. A Non-Homogeneous Group for Full-Rank Homographies

Our proposed methods require one to resolve the non-uniform scaling of the homographies. As homographies are represented by invertible square matrices, we normalize them so that $\forall \mathtt{H} : \det(\mathtt{H}) = 1$ by dividing $\mathtt{H}$ by $\sqrt[3]{\det(\mathtt{H})}$. As a consequence it makes them be part of the special linear group $\mathbb{SL}_3$ in regard to standard matrix multiplication:

- **Closure** $\forall \mathtt{A}, \mathtt{B} \in \mathbb{SL}_3$:

$$\det(\mathtt{AB}) = \det(\mathtt{A})\det(\mathtt{B}) = 1 \Rightarrow \mathtt{AB} \in \mathbb{SL}_3. \tag{8}$$

- **Associativity** follows immediately from associativity of matrix multiplication and from closure.

- **Identity Element** is $\mathtt{I}$, since $\det(\mathtt{I}) = 1 \Rightarrow \mathtt{I} \in \mathbb{SL}_3$.

- **Inverse Element**

$$\forall \mathtt{A} \in \mathbb{SL}_3 : \det(\mathtt{A}^{-1}) = \det(\mathtt{A})^{-1} = 1. \tag{9}$$

This solves the scaling issue. A similar idea was used in [14] to normalize the supercollineation matrix. Hence, imposing this normalization for each and every inter-frame homography allows us to rewrite equation (5) as a strict equality

$$\mathtt{H}_{i,j} = \mathtt{U}_j \mathtt{U}_i^{-1}. \tag{10}$$

This normalization of the homographies is absolutely mandatory in order to derive the following closed-form solutions.

4.2. Full Data Solutions

For the next two methods we suppose that all inter-frame homographies are known, hence the "full data" denomination.

4.2.1 Full Data SVD (SVD)

Since all $\mathtt{H}_{i,j}$ are known, we can rewrite (10) to

$$\forall i, j : \mathtt{H}_{i,j} - \mathtt{U}_j \mathtt{U}_i^{-1} = \mathtt{0} \tag{11}$$

for the noise free case. In presence of noise the estimated $\hat{\mathtt{U}}_i$ should minimize the cost

$$\mathcal{C}_{\mathsf{FDS}}(\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left\| \mathtt{H}_{i,j} - \hat{\mathtt{U}}_j \hat{\mathtt{U}}_i^{-1} \right\|^2. \tag{12}$$

We additionally introduce $\hat{\mathtt{U}}_i'^{-1} = \hat{\mathtt{U}}_i^{-1}$, leading to the matrix form for the cost

$$\mathcal{C}_{\mathsf{FDS}}(\hat{\mathcal{U}}, \hat{\mathcal{U}}') = \|\mathcal{H} - \hat{\mathcal{U}} \hat{\mathcal{U}}'^{\top}\|^2, \tag{13}$$

with

$$\hat{\mathcal{U}} = [\hat{\mathtt{U}}_1^{\top} \ldots \hat{\mathtt{U}}_n^{\top}]^{\top}, \quad \hat{\mathcal{U}}' = [\hat{\mathtt{U}}_1'^{-1} \ldots \hat{\mathtt{U}}_n'^{-1}], \tag{14}$$

$$\mathcal{H} = \begin{bmatrix} \mathtt{H}_{1,1} & \ldots & \mathtt{H}_{n,1} \\ \vdots & \ddots & \vdots \\ \mathtt{H}_{1,n} & \ldots & \mathtt{H}_{n,n} \end{bmatrix}. \tag{15}$$

We know from equation (12) that $\mathcal{H}$ is of rank 3 theoretically since $\mathcal{H}$ is a product of $\hat{\mathcal{U}}\hat{\mathcal{U}}'^{\mathsf{T}}$. The best least squares approximation rank 3 matrix of the noisy $\mathcal{H}$ is obtained via its Singular Value Decomposition (SVD). This observation is similar to the one achieved by Tomasi and Kanade in the seminal paper on factorization [19]. Using the SVD ($\mathtt{U}\Sigma\mathtt{V}^{\mathsf{T}} \overset{\mathsf{SVD}}{\leftarrow} \mathcal{H}$) we obtain a solution for $\hat{\mathcal{U}}$ which is contained in the three columns of $\mathtt{U}\sqrt{\Sigma}$ corresponding to the three smallest singular values. $\hat{\mathcal{U}}'$ would then be the three columns of $\sqrt{\Sigma}\mathtt{V}$ corresponding to the three smallest singular values, but we only keep $\hat{\mathcal{U}}$.

This is a major drawback of FDS since we cannot guarantee that $\hat{\mathcal{U}}'$ is composed of the block by block inverse of the matrices in $\hat{\mathcal{U}}'$ as we originally wanted to in (10).

4.2.2 Full Data Eigenvector (FDE)

In the noise free case, each $\mathtt{U}_k$ is contained in any $\mathtt{H}_{i,k}$ and hence

$$\forall k \in \{1, \ldots, n\} : \sum_{i=1}^{n} \mathtt{H}_{i,k} \mathtt{U}_i = \sum_{i=1}^{n} \mathtt{U}_k \mathtt{U}_i^{-1} \mathtt{U}_i = n \mathtt{U}_k. \tag{16}$$

In order to best approximate this relationship in the noisy case we propose to minimize the cost

$$\mathcal{C}_{\mathsf{FDE}}(\hat{\mathtt{U}}_1, \ldots, \hat{\mathtt{U}}_n) = \sum_{k=1}^{n} \left\| n\hat{\mathtt{U}}_k - \sum_{i=1}^{n} \mathtt{H}_{i,k} \hat{\mathtt{U}}_i \right\|^2, \tag{17}$$

which can be rewritten as

$$\mathcal{C}_{\mathsf{FDE}}(\hat{\mathcal{U}}) = \left\|\mathcal{H}\hat{\mathcal{U}} - n\hat{\mathcal{U}}\right\|^2 \text{ s.t. } \hat{\mathcal{U}}^{\mathsf{T}}\hat{\mathcal{U}} = \mathtt{I}. \tag{18}$$

Thus, the theoretical problem is to find the three eigenvectors $\lambda_a, \lambda_b, \lambda_c$ of $\mathcal{H}$ with eigenvalues equal (in practice, close) to n which provides a solution $\hat{\mathcal{U}} = \begin{pmatrix} \lambda_a & \lambda_b & \lambda_c \end{pmatrix}$. In practice however, due to numerical instability, the eigenvectors might not be real. We therefore use another means to solve the problem.

The solution $\hat{\mathcal{U}}$ is in the null space of $(\mathcal{H} - n\mathtt{I})$. We can again use the SVD, since it gives the best approximation in a least squares sense of the null space under the constraint that $\hat{\mathcal{U}}^{\top}\hat{\mathcal{U}} = \mathtt{I}$. Let $\mathbf{v}_{a'}, \mathbf{v}_{b'}, \mathbf{v}_{c'}$ be the singular vectors associated to the smallest singular values of $(\mathcal{H} - n\mathtt{I})$. They provide the global minimum of $\mathcal{C}_{\mathsf{FDE}}$ with $\mathbf{v}_{a'}, \mathbf{v}_{b'}, \mathbf{v}_{c'}$ being orthonormal. Our global alignment will be $\hat{\mathcal{U}} = [\mathbf{v}_{a'}\, \mathbf{v}_{b'}\, \mathbf{v}_{c'}]$. (See appendix for proof.)

Unfortunately both methods are unable to handle missing inter-frame homographies. This requires a mechanism which composes missing homographies from known ones in order to cope with such a situation. This is not desirable as it requires the preceding step of iteratively finding compositions for the unknown homographies and additionally one might argue about its theoretical justification.

4.3. Missing Data Solutions

From now on we relax the constraint on the full-data and the need for hallucination and present two methods which handle the missing data issue implicitly.

4.3.1 Locally Scaled Homographies (LSH)

Our third method we propose can be considered as a generalization of the FDE solution as it extends equation (16) to cope with missing homographies. Each inter-frame homography $\mathtt{H}_{i,k}$ is re-scaled depending on the interconnectivity of each image k. This re-weighting is done as follows:

$$\forall k \in \{1 \dots n\} : \mathtt{U}_k = \frac{1}{\zeta_k} \sum_{i=1}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \mathtt{U}_i \tag{19}$$

with $\gamma_{i,j} = \begin{cases} 1 & \text{if } \mathtt{H}_{i,j} \text{ is known} \\ 0 & \text{otherwise} \end{cases}$ and $\zeta_k = \sum_{i=1}^{n} \gamma_{i,k}$.

We propose the cost function

$$\mathcal{C}_{\mathsf{LSH}}(\hat{\mathtt{U}}_1, \dots, \hat{\mathtt{U}}_n) = \sum_{k=1}^{n} \left\| \hat{\mathtt{U}}_k - \frac{1}{\zeta_k} \sum_{i=1}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \hat{\mathtt{U}}_i \right\|^2. \tag{20}$$

The construction of matrix

$$\mathcal{S} = \begin{pmatrix} \frac{\gamma_{1,1}}{\zeta_1}\mathtt{H}_{1,1} & \dots & \frac{\gamma_{n,1}}{\zeta_1}\mathtt{H}_{n,1} \\ \vdots & \ddots & \vdots \\ \frac{\gamma_{1,n}}{\zeta_n}\mathtt{H}_{1,n} & \dots & \frac{\gamma_{n,n}}{\zeta_n}\mathtt{H}_{n,n} \end{pmatrix} \tag{21}$$

allows us to rewrite the cost in matrix form as

$$\mathcal{C}_{\mathsf{LSH}}(\hat{\mathcal{U}}) = \left\|\mathcal{S}\hat{\mathcal{U}} - \hat{\mathcal{U}}\right\|^2 = \left\|(\mathcal{S} - \mathtt{I})\hat{\mathcal{U}}\right\|^2. \tag{22}$$

The problem is to minimize $\mathcal{C}_{\mathsf{LSH}}(\hat{\mathcal{U}})$ under the constraint that $\hat{\mathcal{U}}^{\mathsf{T}}\hat{\mathcal{U}} = \mathtt{I}$. This is solved using the singular value decomposition $\mathtt{U}\Sigma\mathtt{V}^{\mathsf{T}} \overset{\text{SVD}}{\leftarrow} (\mathcal{S} - \mathtt{I})$ which provides the three columns of $\hat{\mathcal{U}}$ as the three right singular vectors of $(\mathcal{S} - \mathtt{I})$ associated to the smallest singular values since $\hat{\mathcal{U}}$ is in the null space of $(\mathcal{S} - \mathtt{I})$ and as stated before the SVD gives the best approximation.

4.3.2 Globally Scaled Homographies (GSH)

The cost function $\mathcal{C}_{\mathsf{LSH}}$ has two drawbacks. First, homographies connecting an image for which a high number of homographies are known have a stronger influence on the cost than the images for which only few homographies are known. Second, we know that $\mathtt{H}_{k,k} = \mathtt{I}$ and therefore should not be used in the cost function, biasing the weighting.

Thus, we alter equation (16) with a slightly different weighting. Instead of scaling the term containing the local homographies, we scale the term with the global homographies to get:

$$\forall k \in \{1, \dots, n\} : \mathtt{U}_k + \sum_{i=1, i\neq k}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \hat{\mathtt{U}}_i = \zeta_k \mathtt{U}_k, \tag{23}$$

so that each known $\mathtt{H}_{i,j}$ uniformly influences the cost. We then get the cost function:

$$\mathcal{C}_{\mathsf{GSH}}(\hat{\mathtt{U}}_1, \dots, \hat{\mathtt{U}}_n) = \sum_{k=1}^{n} \left\| (\zeta_k - 1)\hat{\mathtt{U}}_k - \sum_{i=1, i\neq k}^{n} \gamma_{i,k} \mathtt{H}_{i,k} \hat{\mathtt{U}}_i \right\|^2 \tag{24}$$

in which this weighting is corrected according to the preceding arguments. We build the matrix

$$\mathcal{G} = \begin{pmatrix} (1-\zeta_1)\mathtt{I} & (\gamma\mathtt{H})_{2,1} & \dots & (\gamma\mathtt{H})_{n,1} \\ (\gamma\mathtt{H})_{1,2} & (1-\zeta_2)\mathtt{I} & \ddots & \vdots \\ \vdots & \ddots & \ddots & (\gamma\mathtt{H})_{n,n-1} \\ (\gamma\mathtt{H})_{1,n} & \dots & (\gamma\mathtt{H})_{n-1,n} & (1-\zeta_n)\mathtt{I} \end{pmatrix} \tag{25}$$

and rewrite

$$\mathcal{C}_{\mathsf{GSH}}(\hat{\mathcal{U}}) = \left\|\mathcal{G}\hat{\mathcal{U}}\right\|^2. \tag{26}$$

The problem is to minimize $\mathcal{C}_{\mathsf{GSH}}(\hat{\mathcal{U}})$ under the constraint $\hat{\mathcal{U}}^{\mathsf{T}}\hat{\mathcal{U}} = \mathtt{I}$. Again the three right singular vectors associated with the 3 smallest singular values of $\mathcal{G}$ provide a solution for the three columns of $\hat{\mathcal{U}}$.

5. Experiments

In order to compare the different methods we are interested in certain statistical values which are described below.

As a measure for quality of the estimation of homographies we used the Root Mean Squared Residual (RMSR):

$$\epsilon = \min_{\hat{\mathbf{q}}_1,\ldots,\hat{\mathbf{q}}_m} \sqrt{\frac{\sum_{z=1}^{m}\sum_{i=1}^{n}\delta_{z,i}d^2(\mathbf{q}_z^i, \mathtt{U}_i\hat{\mathbf{q}}_z)}{\sum_{z=1}^{m}\sum_{i=1}^{n}\delta_{z,i}}}, \tag{27}$$

which we evaluated for each method before the bundle adjustment $\tilde{\epsilon}$ and afterwards $\hat{\epsilon}$. For the synthetic experiments we compared the final alignment based on the noisy projections with the noise-free data. To do so, we introduced and computed

$$\hat{\eta} = \frac{\sum_{r=1}^{n}\sum_{i=1}^{n}\sum_{z=1}^{4} d(\mathbf{p}_z, \mathtt{U}_r\mathtt{U}_i^{-1}\hat{\mathtt{U}}_i\hat{\mathtt{U}}_r^{-1}\mathbf{p}_z)}{4n^2} \tag{28}$$

with $\mathbf{p}_1,\ldots,\mathbf{p}_4$ being the corners of the image borders. We aligned the exact and estimated homographies to the same reference image $\mathcal{I}_r$ and computed the distance between the exact and estimated image corners, then we took the mean in regard of the number of images and corners. The purpose of $\hat{\eta}$ was to compare the performance of each method to the ground truth data without considering the number of points visible in an image. The RMSR itself could miss a drift of multiple estimated images in the same direction, which was the reason to look for an additional measure. At this point, it is once more pointed out explicitly, that $\hat{\eta}$ does not have a direct statistical meaning! It only helps to identify which alignment visually reflects the ground truth better.

5.1. Implementation Details

An implementation of the proposed methods in Matlab served for the experiments. For the synthetic tests each sample was created following the steps hereafter. First we generated about 10^4 normally distributed 3D points around the camera center. For the viewer those points are uniformly distributed around the static camera center. Then we randomly picked n rotations $\mathtt{R}_1,\ldots,\mathtt{R}_n$ to simulate different views for the camera. Simulating a camera mounted on a tripod we used only pitch $\rho \in [-\alpha;\alpha]$ and yaw $\theta \in [-\alpha;\alpha]$ rotations. A small α avoids that the views create images for which the geometric transform of a point from one image to the other will produce points at infinity. Afterwards these points and rotations are used to generate the images while projecting the points to finite projective planes with a surface of 640×480 pixels (the images) and adding some Gaussian noise with standard deviation σ to the projected points.

Knowing the exact point correspondences between the images we then estimated the homography $\mathtt{H}_{i,j}$ for each pair of images $(\mathcal{I}_i, \mathcal{I}_j)$ with sufficiently overlapping regions using the DLT [9] as implemented by Kovesi [12]. This has been followed by a non-linear optimization in order to find the MLE for the homography in question. Additionally, we required a minimum number of 20 point correspondences between both images. This number had been chosen arbitrarily but helped preserving a certain stability of the homographies when we added noise to the points. For the bundle adjustment we used a sparse Levenberg-Marquardt implemented in Matlab.

5.2. Synthetic experiment

We ran experiments on 15 different levels of Gaussian noise $\sigma \in [0.1; 1.5]$ in steps of $\Delta\sigma = 0.1$ pixels. Each experiment consisted of 100 random runs. We used a setup composed of $n = 50$ images and the maximum camera rotation had been fixed to $\alpha = \pi/8$. This setup produced significant projective transformation (Figure 3).

In average an image was connected to 9 other images. $83.5\% \pm 7.4$ of the overal inter-frame homographies were missing, and $70\% \pm 18.0$ of the connections to the reference image needed to be hallucinated.

Taking a look at the RMSR before the bundle adjustment (Figure 2(a)), it turns out that our methods found initializations with a lower RMSR than the standard method did. FDS and FDE brought improvements but could not reach initializations as good as they have been achieved by LSH and GSH. More important though is to see in Figure 2(b) the fact that after the bundle adjustment the RMSR could in our experiments be lowered by initializing with our methods instead of the Threading. Again – even though it might not be well visible in the figure – both methods GSH and LSH achieved the lowest $\hat{\epsilon}$.

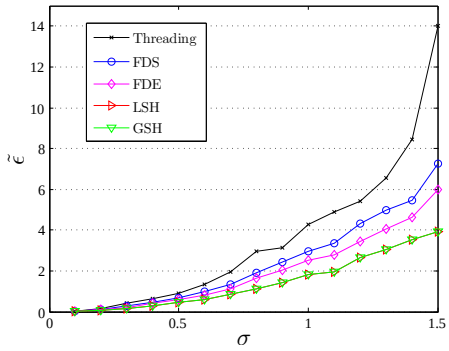
Figure 2(c) shows the distribution of the relative RMSR in terms of the RMSR achieved by the standard method, both after the bundle adjustment. That is to say, the distribution of each sample's $\hat{\epsilon}/\hat{\epsilon}_{\mathsf{Threading}}$. The figure illustrates clearly that all of the proposed solutions initialized the bundle adjustment so that latter converged to a significantly smaller minimum of $\mathcal{C}_{\mathsf{BA}}$ than it did initialized with the standard method in most of the cases.

Furthermore 2(d) visualizes that our methods not only allowed to find a lower local minimum during the optimization, but that the estimated alignment found a solution which better reflected the noise-free data source.

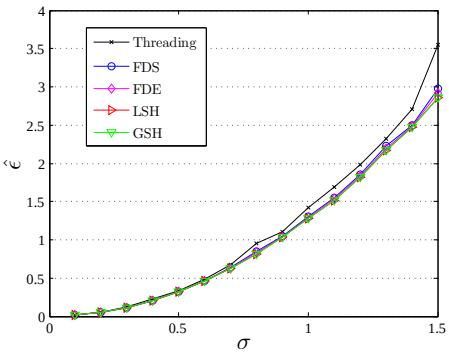
The idea of what $\hat{\eta}$ should penalize is shown in Figure 3. The plots show the projections of the frameborders in the same reference frame for different initializations. In the upper left corner of each sub-figure the difference between each method is visible. It is clear that the bundle adjustment did not recover from the bad initilization provided by the threading, but did so using our batch methods.

5.3. Real Data Experiments

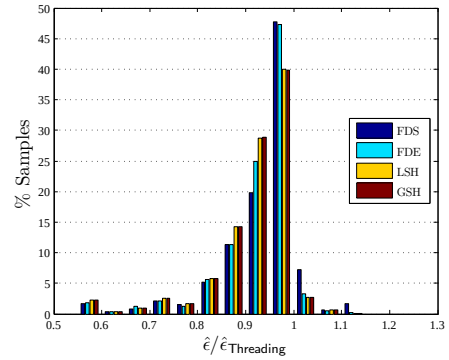
For the real data experiments a video sequence of $n = 167$ undistorted frames with a resolution of 640×480 pixels has been used. The images have been aligned using SIFT features [13] which had been robustly matched using



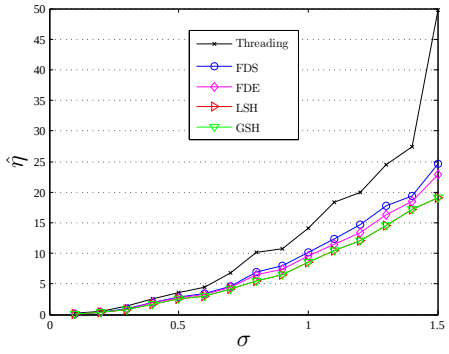
(a) RMSR vs. noise before bundle adjustment



(b) RMSR vs. noise after bundle adjustment



(c) Distribution of $\frac{\hat{\epsilon}}{\hat{\epsilon}_{\text{Threading}}}$ per sample



(d) $\hat{\eta}$ after bundle adjustment

Figure 2. Results of synthetic experiment

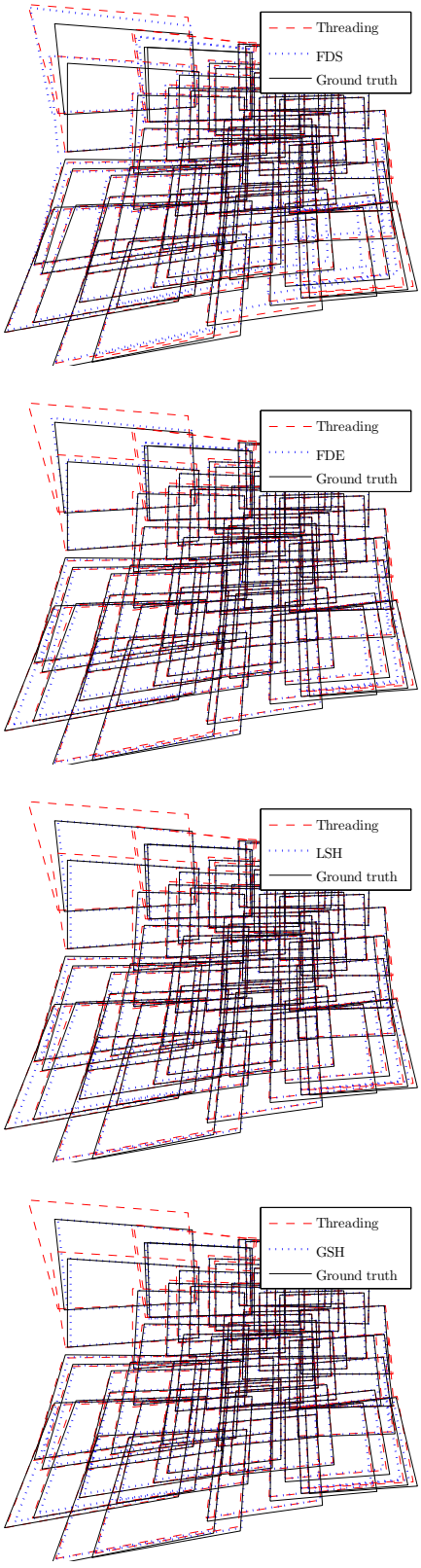


Figure 3. Mosaics of frameborders computed with different initializations.

Figure 4. Mosaic of the real experiment composed of each tenth frame of the video sequence. The bundle adjustment was initialized with homographies computed with GSH.

RANSAC [5]. 46.56% of the homographies failed to be estimated and 20.36% of the homographies to the reference image couldn't be computed. In average an image was connected to 89 ± 48 other images. Figure 4 depicts the results using GSH.

Comparing the RMSR after the bundle adjustment the real data produced

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

That is to say, every of our proposed methods but FDS found an initialization with which the optimization achieved a smaller minimum of $\mathcal{C}_{\mathsf{BA}}$ in this experiment.

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 5 the plots of the frameborders found after

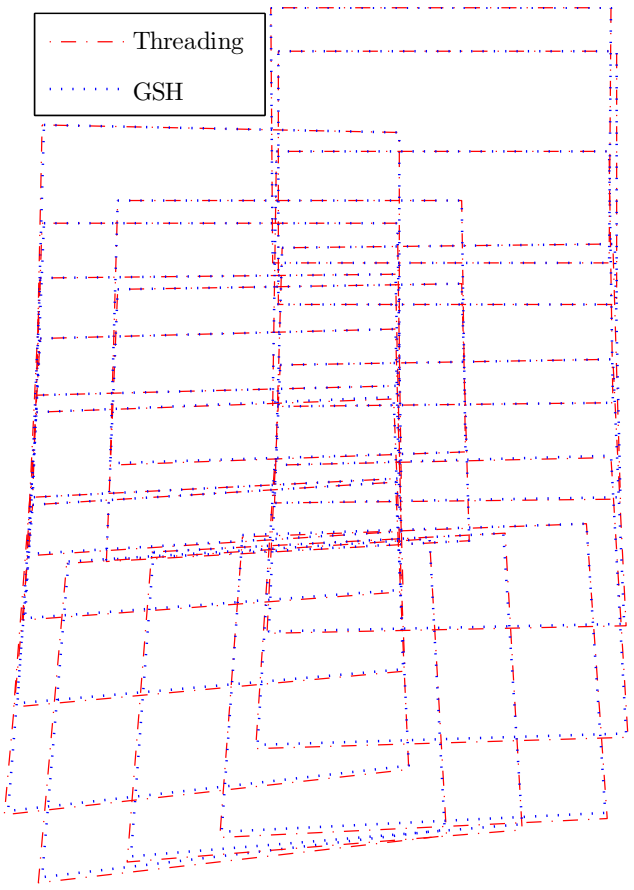


Figure 5. Overlay of frame borders found with Threading versus GSH after BA. The borders of the reference frames (upper right) align well. At the other extremity (lower left) discrepancies of several pixels can be seen.

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 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.

6. Conclusion

We proposed four non-heuristic analytical methods intended to find an initialization, based on inter-frame homographies, for bundle adjustment. We compared these four methods with the standard technique of homography threading and investigated their behavior under the influence of noise and missing data.

Even though we are not able to decrease the number of iterations made by the nonlinear optimization, initializations with our methods FDE, LSH, and GSH lead to better minima than the standard technique. Finally it should be noted that the four methods presented in this paper are very easy to implement.

Appendix

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

$$\arg\min_{\mathtt{X}|\mathtt{X}^\mathsf{T}\mathtt{X}=\mathtt{I}} \|\mathtt{AX}\|_2^2 \equiv \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{x}_i^\mathsf{T}\mathbf{x}_i=1,\mathbf{x}_i^\mathsf{T}\mathbf{x}_j=0} \sum_{k\in[n]} \|\mathtt{A}\mathbf{x}_k\|_2^2 \quad (29)$$

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 $\mathcal{L} = \sum_{k\in[n]} \mathcal{L}_k$ with

$$\mathcal{L}_k = \|\mathtt{A}\mathbf{x}_k\|_2^2 + \lambda_k(1-\mathbf{x}_k^\mathsf{T}\mathbf{x}_k) + \sum_{r\in[n],r\neq k} \mu_{k,r}(\mathbf{x}_k^\mathsf{T}\mathbf{x}_r)^2 \quad (30)$$

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 $\forall p \in [n] : \frac{\partial \mathcal{L}}{\partial \mathbf{x}_p} = 0$ and as $\forall k \in [n] : \mathcal{L}_k \geq 0$ is equivalent to $\forall p,k \in [n] : \frac{\partial \mathcal{L}_k}{\partial \mathbf{x}_p} = 0$.

Differentiating the lagrangian yields

$$\forall p,k \in [n],\, p \neq k : \frac{1}{2}\frac{\partial \mathcal{L}_k}{\partial \mathbf{x}_p} = \mu_{k,p}\mathbf{x}_k\mathbf{x}_k^\mathsf{T}\mathbf{x}_p = 0 \quad (31)$$

$$\begin{aligned} \forall k \in [n] : \frac{1}{2}\frac{\partial \mathcal{L}_k}{\partial \mathbf{x}_k} &= \mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k - \lambda_k\mathbf{x}_k + \sum_{r\in[n],r\neq k} \mu_{k,r}\mathbf{x}_r\mathbf{x}_r^\mathsf{T}\mathbf{x}_k \\ &= \mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k - \lambda_k\mathbf{x}_k. \end{aligned} \quad (32)$$

Both, (31) and (32) together with (6) (resp. (6)) yield that $\sum_{k\in[n]} \|\mathtt{A}\mathbf{x}_k\|_2^2$ can only be minimal if

$$\forall k \in [n] : \mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k = \lambda_k\mathbf{x}_k. \quad (33)$$

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 (29)

$$\arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{x}_i^\mathsf{T}\mathbf{x}_i=1,\mathbf{x}_i^\mathsf{T}\mathbf{x}_j=0} \sum_{k\in[n]} \mathbf{x}_k^\mathsf{T} \underbrace{\mathtt{A}^\mathsf{T}\mathtt{A}\mathbf{x}_k}_{\lambda_k\mathbf{x}_k} \equiv \arg\min_{\lambda_1,\ldots,\lambda_n} \sum_{k\in[n]} \lambda_k \quad (34)$$

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 are the n right singular vectors corresponding to the n smallest singular values of $\mathtt{A}$. □

References

[1] F. Bajramovic and J. Denzler. Global uncertainty-based selection of relative poses for multi camera calibration. In *BMVC*, 2008. 2

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 2004. 2

[3] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *IJCV*, 2006. 1, 2

[4] D. Capel and A. Zisserman. Automatic mosaicing with super-resolution zoom. *CVPR*, 1998. 1

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Com. ACM*, 1981. 7

[6] V. Govindu. Combining two-view constraints for motion estimation. In *CVPR*, 2001. 2

[7] V. Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *CVPR*, 2004. 2

[8] R. I. Hartley. Self-calibration of stationary cameras. *IJCV*, 1997. 1

[9] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. 2, 5

[10] M. Irani, S. Hsu, and P. Anandan. Video compression using mosaic representations. *Signal Process-Image*, 1995. 1

[11] E.-Y. Kang, I. Cohen, and G. Medioni. A graph-based global registration for 2d mosaics. *ICPR*, 2000. 2

[12] P. Kovesi. Matlab and octave functions for computer vision and image processing. 5

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004. 5

[14] E. Malis and R. Cipolla. Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. *PAMI*, 2002. 1, 3

[15] R. Marzotto, A. Fusiello, and V. Murino. High resolution video mosaicing with global alignment. In *CVPR*, 2004. 1, 2

[16] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999. 2

[17] P. Sturm. Algorithms for plane-based pose estimation. In *CVPR*, 2000. 2

[18] R. Szeliski. Image alignment and stitching: A tutorial. Technical report, 2004. 1

[19] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 1992. 3

[20] J. Verges-Llahi, D. Moldovan, and T. Wada. A new reliability measure for essential matrices suitable in multiple view calibration. *VISAPP*, 2007. 2

[21] L. Zelnik-Manor and M. Irani. Multi-view constraints on homographies. In *Trans. PAMI*, 2002. 3