Revisiting Rotation Averaging: Uncertainties and Robust Losses

Ganlin Zhang
ETH Zürich
zhangga@student.ethz.ch

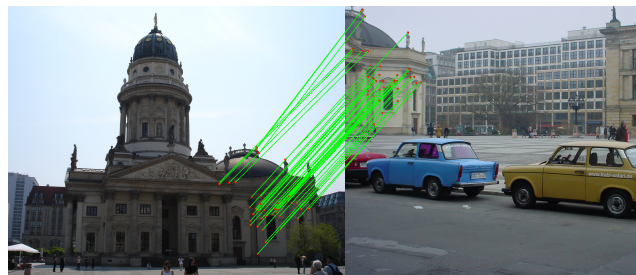Viktor Larsson
Lund University
viktor.larsson@math.lth.se

Daniel Barath
ETH Zürich
dbarath@inf.ethz.ch

(a) Uncertainty = **0.63**
Rotation error = **0.05°**, Translation error = **6.30°**

(b) Uncertainty = **7.61**
Rotation error = **4.81°**, Translation error = **21.12°**

Figure 1. Two example pairs where, even though the relative pose estimation found a large number of correct inliers (shown by green line segments), the quality of the estimated poses differs greatly. The uncertainties, used as weighting terms, help in rotation averaging to account for this quality difference. Here, the shown numbers as the uncertainties are the traces of the covariance matrices.

Abstract

In this paper, we revisit the rotation averaging problem applied in global Structure-from-Motion pipelines. We argue that the main problem of current methods is the minimized cost function that is only weakly connected with the input data via the estimated epipolar geometries. We propose to better model the underlying noise distributions by directly propagating the uncertainty from the point correspondences into the rotation averaging. Such uncertainties are obtained for free by considering the Jacobians of two-view refinements. Moreover, we explore integrating a variant of the MAGSAC loss into the rotation averaging problem, instead of using classical robust losses employed in current frameworks. The proposed method leads to results superior to baselines, in terms of accuracy, on large-scale public benchmarks. The code is public.
https://github.com/zhangganlin/GlobalSfMpy

1. Introduction

Building large 3D reconstructions from unordered image collections is an essential component in any system that relies on crowd-sourced mapping. The current paradigm is to perform this reconstruction via *Structure-from-Motion* [29] which jointly estimates the camera parameters and the scene geometry represented with a 3D point cloud. Meth-

ods for Structure-from-Motion can generally be categorized into two classes; *Incremental methods* [29, 32, 33, 38] that sequentially grows a seed reconstruction by alternating triangulation and registering new images, and *Global methods* [8, 22, 24, 27] which first estimate pairwise geometries and then aggregate them in a bottom-up approach. Historically, incremental methods are more robust and accurate, but the need for frequent bundle adjustment [35] comes with significant computational cost which limits their scalability. In contrast, global (or non-sequential) methods require much lower computational effort and can in principle scale to larger image collections. However, in practice, current methods are held back by the lack of accuracy and have not enjoyed the same level of success as incremental methods.

Global methods work by first estimating a set of pairwise epipolar geometries between co-visible images. Next, via *rotation averaging*, a set of globally consistent rotations are estimated by ensuring they agree with the pairwise relative rotations. Once the rotations are known, the camera positions and 3D structure are estimated, and refined jointly in a single final bundle adjustment.

Rotation averaging has a long history in computer vision (see *e.g.* [16, 22] for early works) and is a well-studied problem. Most methods formulate it as an optimization problem, finding the rotation assignment that minimizes some energy. A common choice is the *chordal distance*, measuring the

discrepancy in the rotation matrices in the L_2-sense

$$\min_{\{R_i\}_{i=1}^N} \sum_{i=1}^N \sum_{j\in\mathcal{N}(i)} \|\hat{R}_{ij}R_i - R_j\|_F^2 \qquad (1)$$

where $\hat{R}_{ij}$ is the relative rotation estimated between image i and j. There are also other choices such as using angle-axis [34] or quaternion [16] as rotation representation, or optimising over a Lie algebra [17], however the overall idea (measuring some consistency with the relative estimates) remains the same. Many works have focused on the optimization problem itself, both theoretically [11, 36] and by providing new algorithms [9], but did not consider whether the cost itself is suitable for the task. In (1), each relative rotation measurement is given the same weight. However in practice, the quality of the epipolar geometries varies significantly. Figure 1 shows two images with wildly different uncertainties (and errors) in the rotation estimate. To address this problem, there is a line of work [6, 18, 31] which augment the cost in (1) with robust loss functions that give a lower weight to large residuals. However, the same loss function is generally applied to each residual, independent of the measurement uncertainty.

In this paper we revisit the rotation averaging problem. We argue that the main problem in current methods is that the cost functions that are minimized are only weakly connected with the input data via the estimated epipolar geometries. We propose to better model the underlying noise distributions (coming from the keypoint detection noise and spatial distribution) by directly propagating the uncertainty from the point correspondences into the rotation averaging problem, as shown in Figure 2. While the idea itself is simple, we show that this allows us to get significantly more accurate estimates of the absolute rotations; reducing the gap between incremental and global methods. Note that the uncertainties we leverage are essentially obtained for free by considering the Jacobians of the two-view refinement.

As a second contribution, we explore integrating a variant of the MAGSAC [3] loss into the rotation averaging problem, instead of using the classical robust losses employed in current frameworks. MAGSAC [3] was originally proposed as a threshold-free estimator for two-view epipolar geometry, where the idea is to marginalize over an interval of acceptable thresholds, *i.e.*, noise range. We show that this fits well into the context of rotation-averaging, as it is not obvious how to set the threshold for deciding on inlier/outlier relative rotation measurements, especially in the uncertainty-reweighted cost that we propose.

2. Related Work

Rotation averaging is a long standing problem in computer vision with some of the early works dating back more
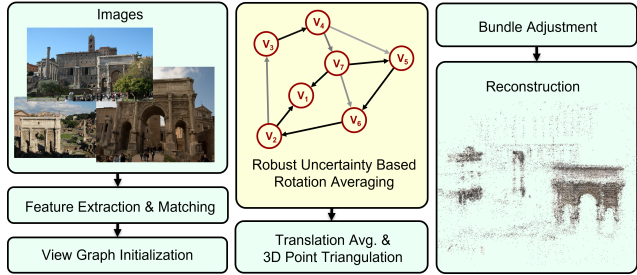

Figure 2. **Global Structure from Motion pipeline** with the proposed uncertainty-based rotation averaging. In the yellow box, the opacity of the view graph edges indicates uncertainty difference.

than two decades. Govindu [16] proposes an approximation in which the problem becomes linear in terms of the quaternions, after heuristically resolving the sign ambiguity. Similarly, Martinec and Pajdla [22] parameterize the problem in terms of the full 3×3 matrix, but drop the nonlinear constraints to obtain a tractable problem. Wilson et al. [36] investigate, more generally, under what conditions the rotation averaging problem is easy.

When the chordal distance is used, *i.e.* (1), the optimization problem can be solved via an SDP-based relaxation, see Arie et al. [2] and Fredriksson et al. [13] for some of the earlier papers in this line of work. For this particular relaxation, Eriksson et al. [11] show that under some noise assumptions, the SDP relaxation obtains the globally optimal solution. In [11], the authors propose a block-coordinate descent method specialized for the dual formulation of the rotation averaging problem. Dellaert et al. [9] propose an optimization scheme based on sequentially lifting the problem into higher-dimensional rotations $SO(n)$. The method, named *Shonan* rotation averaging, is shown to avoid some local minima in which standard optimization techniques, such as Levenberg-Marquardt [23], might be stuck in. In our work, our contributions are related to changing the cost function minimized, and it is possible that the methods from these works could be applied in our setting as well.

To deal with outliers in the relative rotation measurements, Hartley et al. [18] propose a generalization of the Weiszfeld-algorithm to minimize the L_1-loss over the rotation residuals. This method was later extended by Chatterjee and Govindu in [6]. To obtain more robust estimations other robust losses have been explored, *e.g.* $L_{0.5}$ [7] and Geman-McClure [31]. In our work, we experimentally evaluate these loss-functions (in addition to many others [19, 20]) and compare against the loss function based on MAGSAC [3] that we propose. In [14], Gao et al. propose an iterative scheme for solving the rotation averaging problem where they weight the view graph edges based on the two-view inliers. In our experiments, we compare against this re-weighting scheme as well.
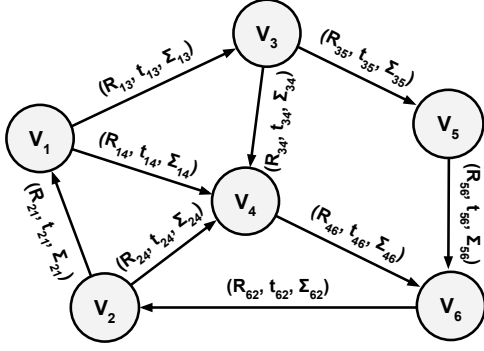
Figure 3. **View graph** with its vertices representing images and its edges 3D rotations and translations between the views. Edge weight Σ_{ij} comes from the uncertainty of the pair-wise estimation.

Global Structure-from-Motion builds the reconstruction by aggregating pair-wise estimates of epipolar geometries. In most cases, this is done via rotation averaging [24, 27], but there are also works that perform the averaging in SE(3) instead [8]. Once rotations are known, there are different approaches for recovering the translations and structure. Wilson and Snavely [37] solve the translation averaging problem using an outlier filter based on projecting the translations onto 1D subspaces. Moulon et al. [24] formulates the problem as an L_∞-optimization. Olsson and Enqvist [27] also rely on L_∞-optimization but solve jointly for both 3D points and camera positions.

There are several open-source frameworks for global Structure-from-Motion, such as Theia [34] and Open-MVG [25]. In our experiments, we integrate our updated optimization objectives into the framework from Theia [34], and use the remaining pipeline unchanged. Our contributions are, however, not specific to this pipeline.

3. Rotation Averaging with Uncertainties

In this paper, we propose a way to leverage uncertainties, coming from pair-wise relative pose estimations, in rotation averaging. This additional signal indicating the quality of the input relative poses allows to further improve 3D reconstruction by global Structure from Motion (SfM) methods.

Overview of Rotation Averaging. Rotation averaging is one of the main components of global SfM algorithms allowing to estimate the global rotation in such a way that it is decoupled from the translations. It is usually formalized as a graph optimization problem, where the vertices are the image poses and the edges are relative poses estimated as a preliminary step. See Fig. 3. Each edge imposes a constraint on the global poses of its neighboring vertices (*i.e.*, images). The primary goal is to use these constraints to find the absolute poses of the cameras.

Standard rotation averaging proceeds as follows: first, a vertex is chosen as the origin. Second, the orientations of all vertices that fall into the same connected component as the origin-defining one are initialized by a maximum spanning tree. Finally, the rotations are optimized jointly, leveraging the information coming from the relative rotations. Rotation R_i^* of the ith camera is calculated as

$$\{R_i^*\}_{i=1}^n = \operatorname*{arg\,min}_{\{R_i\}_{i=1}^n} \sum_{(i,j)\in\mathcal{E}} \rho\left(\|L_e(R_i, R_j, R_{ij})\|^2\right), \quad (2)$$

where $\mathcal{E}$ is the set of edges in the view graph $\mathcal{G} = (\mathcal{E}, \mathcal{V})$, function L_e measures the error of the relative rotation $\bar{R}_{ij} = R_i R_j^\mathsf{T}$ coming from the optimized global orientations R_i and R_j of the ith and jth views w.r.t. the estimated rotation R_{ij}. Function $\rho(\cdot) : \mathbb{R} \to \mathbb{R}$ is the robust loss used to deal with potential outliers in the data. The optimization procedure outputs $\{R_i^*\}_{i=1}^n$, the estimated absolute orientations of all the n views in the same connected component. In case multiple disjoint graphs are generated from the images, the procedure runs on all of them independently.

Uncertainty-Aware Rotation Averaging. In Eq. (2), every edge $(i,j) \in \mathcal{E}$ of the view graph $\mathcal{G}$ is treated equally. However, the estimated relative poses depend on many factors in practice (*e.g.*, inlier number, baseline) and thus, they are of different quality. This quality measure can be captured by the uncertainty of the two-view estimation. In rotation averaging, the uncertainty in each measurement can be considered in a weighting scheme as follows:

$$\{R_i^*\}_{i=1}^n = \operatorname*{arg\,min}_{\{R_i\}_{i=1}^n} \sum_{(i,j)\in E} \rho\left(\left\|D_{ij}^\mathsf{T} L_e(R_i, R_j, R_{ij})\right\|^2\right), \quad (3)$$

where term D_{ij} depicts the uncertainty of edge (i,j).

Relative Rotation Uncertainty. To propagate the uncertainty from the features to the estimated relative rotation, we first express the rotation via the 2D-2D point correspondences. The relative rotation R_{ij} and translation t_{ij} between the ith and jth views is written as follows:

$$\begin{aligned} R_{ij}, t_{ij} &= \operatorname*{arg\,min}_{R,t} L_{ij} \\ &= \operatorname*{arg\,min}_{R,t} \sum_{(p,p')\in\mathcal{M}_{ij}} L_s{}^2(R, t, p, p', K_i, K_j), \end{aligned} \quad (4)$$

where $\mathcal{M}_{ij} = \{(p, p') \mid p \in \text{view}_i \wedge p' \in \text{view}_j\}$ is the set of point correspondences in views i and j. Matrices K_i and K_j contain the intrinsic parameters, *e.g.* focal length and principal point, of the cameras. L_s is the point-to-model residual, *e.g.*, Sampson distance or symmetric epipolar error. Due to its robustness, we use Sampson distance as L_s

(a) Example image from scene Gendarmenmarkt.

(b) Ground truth.
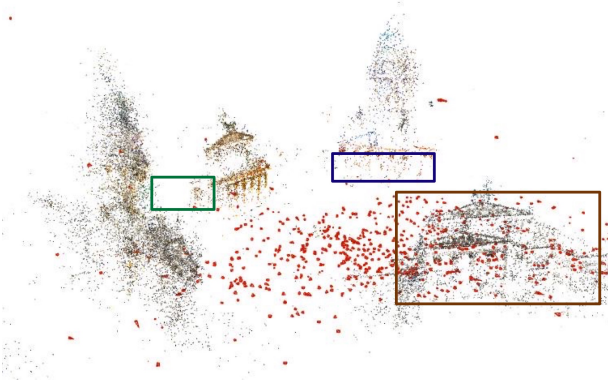
(c) Baseline.

(d) Ours.

Figure 4. Scene Gendarmenmarkt from the 1DSfM dataset, containing a concert hall in the middle and two cathedrals on both sides. Fig. 4a is an image from the scene. Fig. 4b is the ground truth reconstruction. Fig. 4c and Fig. 4d are results of the baseline [34] and the proposed method, respectively. The camera poses are drawn as red cones. Compared to the baseline, our method has achieve better reconstruction. Example differences are shown by colored boxes in Fig. 4c and Fig. 4d. *Green boxes*: walls of the concert hall are missing in the baseline, while being reconstructed in ours. *Blue boxes*: the bottom of the right cathedral is almost missing in the baseline, but is well-preserved in ours. *Brown box*: baseline reconstructs an additional concert hall which does not exist. Fig. 4d shows that the proposed method is correct.

written as follows:

$$
L_s(R_{ij}, t_{ij}, p, p', K_i, K_j) =
\frac{p^{\mathrm{T}\prime} F_{ij} p}{((Fp)_1^2 + (Fp)_2^2 + (F^{\mathrm{T}}p')_1^2 + (F^{\mathrm{T}}p')_2^2)^{1/2}}, \quad (5)
$$

where F_{ij} is the fundamental matrix as

$$
F_{ij} = K_j^{-\mathrm{T}} R_{ij} [t_{ij}]_\times K_i^{-1}, \quad (6)
$$

and the lower-indices refer to coordinates of Fp and $F^{\mathrm{T}}p'$.

The angle-axis representation of rotation R_{ij}^A is used to avoid singularity. The uncertainty propagation to get covariance $C(R_{ij}^A)$ of relative rotation R_{ij}^A is done as

$$
C(R_{ij}^A) = (J^{\mathrm{T}}(R_{ij}^A) J(R_{ij}^A))^{-1} \quad (7)
$$

where $J(R_{ij}^A)$ is the Jacobian of L_{ij}

Covariances are plugged into Eq. 3 by calculating D_{ij} as the LLT decomposition of $C(R_{ij}^A)^{-1}$ [26] that equals to

$$
C(R_{ij}^A)^{-1} = D_{ij} D_{ij}^{\mathrm{T}}. \quad (8)
$$

Using such weighting scheme in the optimization procedure allows us to consider the pair-wise uncertainties of the rotations to reason about their qualities in a theoretically justifiable manner. In practice, L_s, and thus Jacobian $J(R_{ij})$, is calculated only on the inlier correspondences after the robust estimation, *e.g.* RANSAC [12], finishes.

3.1. Marginalizing over the Noise Scale

In MAGSAC++ [3], a robust loss function is designed by marginalizing over the noise standard deviation to reduce the dependence on a manually set inlier-outlier threshold parameter. The MAGSAC++ loss assumes that the in-

lier residuals follow a χ^2-distribution and does not make any assumptions about the outliers. While the MAGSAC++ loss has only been applied to residual functions measuring the consistence of point correspondences with a projective transformation (*e.g.*, homography and essential matrix), it can be used as a general robust loss like the Huber loss [20].

Let us define the MAGSAC++ loss for rotation averaging. The data points, in our case, are a set of relative rotations, and the models to estimate are the global orientations of the cameras. This means that for a pair of global rotations R_i and R_j, we are given relative rotation R_{ij} constraining both. The inlier weight of R_{ij} is

$$w(r(R_i R_j^{\mathrm{T}}, R_{ij})) = \int_0^{+\infty} \mathbf{P}(R_{ij} \mid R_i, R_j, \sigma) f(\sigma) \mathrm{d}\sigma, \quad (9)$$

where $r(R_i R_j^{\mathrm{T}}, R_{ij})$ is the residual, σ is the noise standard deviation, $f(\sigma)$ is the prior distribution of σ assumed to be uniform on range $[0, \sigma_{\max}]$.

By rewriting Eq. (9) as the marginal density of the inlier residual, we can get

$$w(r) = \int_0^{+\infty} g(r \mid \sigma) f(\sigma) \mathrm{d}\sigma, \quad (10)$$

where $g(r \mid \sigma)$ is the density of the residual r given σ. Assume that σ is uniformly distributed $\sigma \sim U(0, \sigma_{\max})$, then Eq. (10) becomes

$$w(r) = \frac{1}{\sigma_{\max}} \int_0^{\sigma_{\max}} g(r \mid \sigma) \mathrm{d}\sigma. \quad (11)$$

Assume that the residual r is in some ν-dimensional space and the error along each axis of this ν-dimensional space is independent and normally distributed with the same variance σ^2. Then r^2/σ^2 has χ^2-distribution with ν degrees of freedom. For a given σ, r has the trimmed χ-distribution with ν degrees of freedom. Let $\tau(\sigma) = k\sigma$ be the chosen quantile of the χ-distribution.

For $r \geq k\sigma_{\max}$, $w(r) = 0$. For $0 \leq r \leq k\sigma_{\max}$

$$w(r) = \frac{1}{\sigma_{\max}} \int_{r/k}^{\sigma_{\max}} g(r \mid \sigma) \mathrm{d}\sigma =$$
$$\frac{1}{\sigma_{\max}} M(\nu) 2^{\frac{\nu-1}{2}} \left(\Gamma\left(\frac{\nu-1}{2}, \frac{r^2}{2\sigma_{\max}^2} \right) - \Gamma\left(\frac{\nu-1}{2}, \frac{k^2}{2} \right) \right). \quad (12)$$

Here, M is the normalizing constant as follows:

$$M(\nu) = \left(2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2} \alpha \right) \right)^{-1}, \quad (13)$$

where α means that $\tau(\sigma)$ is set to the α-quantile, *i.e.* CDF$(k) = \alpha$. The CDF here is the cumulative distribution function of χ-distribution with ν degrees of freedom. In
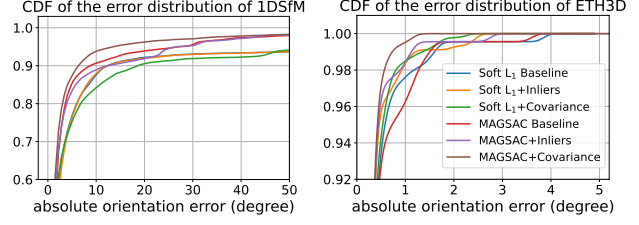

Figure 5. Cumulative distribution functions of the rotation errors (in degrees) after the rotation averaging on the 1DSfM [37] (left; 8824 poses) and ETH3D datasets [30] (right; 451 poses). Being close to the top-left corner is preferred.

our case, $\nu = 3$. Function $\Gamma(a, x)$ is the upper-incomplete gamma function

$$\Gamma(a, x) = \int_x^{+\infty} t^{a-1} \exp(-t) \mathrm{d}t, \quad (14)$$

and

$$\Gamma(a) = \Gamma(a, 0) = \int_0^{+\infty} t^{a-1} \exp(-t) \mathrm{d}t. \quad (15)$$

From Eq. (12), we can calculate the weight of each edge by passing its residual to the formula. In order to convert it to a robust loss that can be used in the optimization procedure while keeping its beneficial properties, we use loss

$$\rho(r) = w(0) - w(r), \quad (16)$$

where $w(0)$ is the weight if $R_{ij} = R_i R_j^{\mathrm{T}}$. It is the maximal weight. Scalar $w(r)$ is the robust weight that residual r implies. Loss $\rho(r)$ is positive and increasing on interval $(0, k\sigma_{\max})$. Therefore, it can be minimized by IRLS and each iteration guarantees a non-increase in the loss ([21], chapter 9). Consequently, it converges to a local minimum.

4. Experiments

In this section, we test the global SfM implemented in the Theia [34] library with and without considering the uncertainties coming from the estimated relative poses in the proposed way. Moreover, we evaluate popular loss functions, including Soft L_1 loss which is used in Theia. For rotation averaging, Theia uses angle-axis rotation parameterization to minimize the relative rotation error via a numerical optimization implemented in the Ceres [1] library. It then performs the 1DSfM translation averaging [37].

Benchmarks. We test the proposed uncertainty-based rotation averaging on the 1DSfM [37] and ETH3D [30] datasets. The 1DSfM dataset consists of 14 different landmarks with images, collected from the Internet, of varying

	Soft L$_1$ [5]			MAGSAC [3]		
	Baseline	+ Inliers	+ Covariance	Baseline	+ Inliers	+ Covariance
Ellis Island	68.2	70.0	67.1	73.8	**76.4**	76.3
Gendarmenmarkt	8.9	7.0	6.1	49.6	54.0	**54.2**
Montreal Notre Dame	77.2	76.4	74.0	79.3	78.6	**79.3**
Notre Dame	77.5	**80.2**	73.9	80.1	78.7	79.5
NYC Library	59.0	61.7	60.6	63.7	**68.6**	65.4
Piazza del Popolo	60.2	59.3	62.1	60.3	**62.2**	60.7
Roman Forum	57.7	50.8	60.3	62.5	65.5	**70.2**
Tower of London	48.8	49.4	66.8	51.9	57.8	**67.2**
Union Square	27.2	24.6	31.3	28.0	**38.0**	35.0
Yorkminster	63.3	64.0	64.5	62.3	64.1	**67.1**
Vienna Cathedral	**67.1**	66.8	60.6	62.9	46.8	66.7
Piccadilly	33.2	33.6	30.6	46.7	50.0	**51.6**
Alamo	63.3	65.4	62.4	65.4	65.4	**66.8**
Average	54.7	54.6	55.4	60.5	62.0	**64.6**

Table 1. The Area Under the recall Curve (AUC) at 5° of estimated rotations after rotation averaging by [6] (Baseline) with different losses (Soft L$_1$ [5] and MAGSAC [3]) and weighting strategies: by the number of inliers (+ Inliers), by the proposed covariance (+ Covariance).

(a) Image of Scene *Roman Forum*.　　(b) Ground truth.　　(c) Inliers [14].　　(d) Covariance.
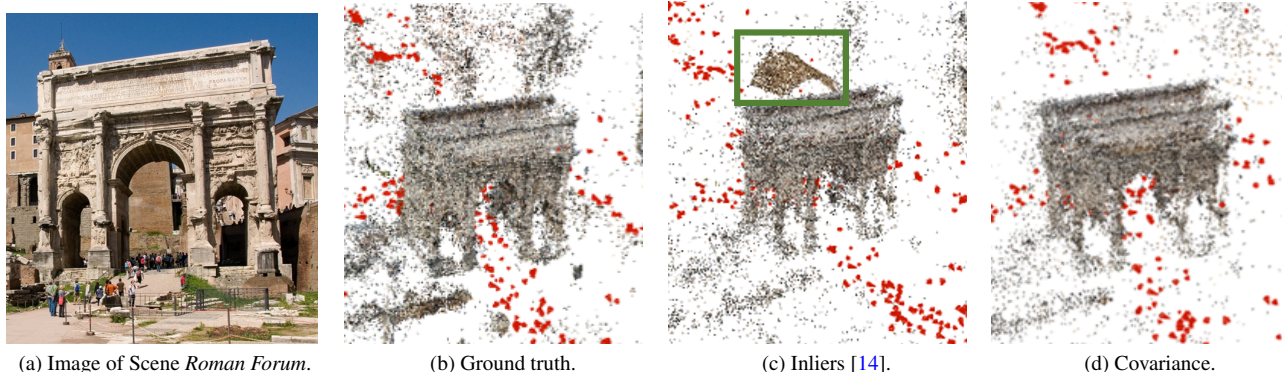
Figure 6. Scene *Roman Forum* from the 1DSfM dataset [30]. Fig. 6a is an example image. Fig. 6b is the ground truth reconstruction. Fig. 6c and Fig. 6d are the reconstructions when the pose graph edges are weighted by the inlier number [14] or by the proposed covariance, respectively. The cameras are represented by red cones. *Green box*: In Fig. 4c, [14] reconstructs an additional roof which does not exist.

Setting		AUC α (%) ↑			
		α=2°	α=5°	α=10°	α=20°
Soft L$_1$ [5]	Baseline	37.4	59.4	72.5	82.0
	+ Inliers	35.9	58.9	72.3	81.9
	+ Covariance	39.6	60.3	71.5	79.2
MAGSAC [3]	Baseline	41.1	65.0	77.4	85.6
	+ Inliers	39.2	62.3	74.9	83.3
	+ Covariance	**44.6**	**67.6**	**79.4**	**87.2**

Table 2. AUC scores of rotation errors after the full global SfM pipeline [34] averaged over all scenes of the 1DSfM dataset [37].

sizes and capturing conditions, *e.g.*, day and night. It pro-

vides 2-view matches with epipolar geometries and a reference reconstruction from incremental SfM (computed with Bundler [32, 33]) for measuring error. Since Bundler was published more than ten years ago, we reconstructed the scenes with COLMAP [29] to get a better reconstruction that can be considered as ground truth. We tune the hyperparameters of the tested methods on scene Madrid Metropolis, which are the inlier-outlier threshold of loss function in different settings: baseline 0.02, baseline+inlier 0.06, baseline+covariance 0.02. Therefore, we only report results on the other scenes.

The ETH3D Stereo Muti-view is an indoor-outdoor dataset of 13 scenes with high-resolution images (6000×4000), keypoints, LiDAR depth, and ground truth poses. We use the 13 scenes of the training set, and use all image pairs with at least 500 GT keypoints

in common. For this dataset, we run SuperPoint [10] and SuperGlue [28] to obtain point correspondences from which the relative poses are estimated. As baseline, we also run COLMAP on exactly the same features.

Setting		AUC α (%) $\uparrow$			
		$\alpha=2°$	$\alpha=5°$	$\alpha=10°$	$\alpha=20°$
COLMAP [29]		86.7	90.9	92.8	93.8
Soft L_1 [5]	Baseline	88.6	95.2	97.6	**98.8**
	+ Inliers	89.2	95.2	97.3	98.8
	+ Covariance	84.5	91.4	95.6	97.6
MAGSAC [3]	Baseline	86.9	94.5	97.3	98.6
	+ Inliers	89.5	94.9	96.9	98.3
	+ Covariance	**91.2**	**95.9**	**97.7**	98.6

Table 3. AUC scores of rotation errors after the full global SfM pipeline [34] averaged over all scenes of the ETH3D dataset [30].

The rotations in two reconstructions are not directly comparable since the global rotation of the view-graph is unknown. Therefore, we first align the reconstructed and the ground truth graphs with rotation

$$\hat{R}_{\text{align}} = \arg\min_{R_{\text{align}}} \sum_{i=1}^{i=|\mathcal{V}|} \rho\left(\|L_e(R_i, R'_i, R_{\text{align}})\|^2\right). \quad (17)$$

The loss function $\rho(\cdot)$ we used here is Cauchy loss. We do not use MAGSAC here to avoid bias when comparing methods to the ground truth. Rotations R_i and R'_i are the orientations of view i in GT and reconstructed graph, respectively. Finally, we rotate the reconstructed graph by $\hat{R}_{\text{align}}$.

Uncertainties in Rotation Averaging. To explore the influence of considering the uncertainties of two-view geometry estimation in rotation averaging, we compare the following weighting schemes. The *baseline* is using a constant weight for all rotations. We test using the *inlier number* of the estimated relative pose as weight in the optimization [14]. We use the proposed *uncertainty*-based weighting with covariance matrices as described in Sec. 3. To show its impact on multiple configurations, we run rotation averaging with Soft L_1 [5] and MAGSAC losses [3].

In Tab. 1, the Area Under the recall Curve (AUC) at 5° is reported on the scenes of the 1DSfM dataset. The last row shows the average AUC scores. On average, using any of the compared weighting strategies improves the accuracy. Using the proposed uncertainties leads to the highest AUC score with both robust losses. MAGSAC clearly leads to better results than Soft L_1. The absolute best is obtained by MAGSAC loss and the proposed covariance-based weighting. Compared to the original Theia code (*i.e.*, Baseline

Weight	AUC α (%) $\uparrow$			
	$\alpha=2°$	$\alpha=5°$	$\alpha=10°$	$\alpha=20°$
Covariance	**38.3**	**62.5**	**75.8**	**84.9**
Trace	38.1	58.8	70.9	79.7
F-norm	37.5	58.2	70.4	79.4

Table 4. Area Under the recall Curve (AUC) of the errors of the rotation averaging, on the 1DSfM dataset, when using the trace, the F-norm of the inverse covariance and proposed approach (Covariance) for weighting the view-graph edges.

Loss Function	AUC α (%) $\uparrow$			
	$\alpha=2°$	$\alpha=5°$	$\alpha=10°$	$\alpha=20°$
MAGSAC [3]	**35.9**	**60.9**	**75.0**	84.0
Soft L_1 [5]	30.4	54.6	72.0	80.8
$L_{0.5}$ [7]	29.8	55.2	71.3	81.8
Tukey [21]	29.0	56.2	73.7	**84.7**
Cauchy [4]	28.9	53.7	70.3	81.4
Huber [20]	21.6	45.7	63.4	76.7
GM [15]	13.7	42.2	65.3	79.0
Trivial	11.2	30.2	49.2	66.1

Table 5. Area Under the recall Curve (AUC) of the rotation errors of different robust losses inside the rotation averaging of [6] averaged over all scenes in the 1DSfM dataset [37].

Downsample Factor	Baseline	Ours	Improvement
1	88.61	91.19	+2.58
4	81.23	84.74	+3.51
8	80.69	84.62	+3.93

Table 6. The AUC@2° scores of the rot. errors of global SfM [34] on the ETH3D [30] dataset when the images are downsampled.

with Soft L_1), the proposed algorithm leads to a more than 10 points increase in the AUC score.

The left plot of Fig. 5 shows the rotation error distribution over all the scenes of the 1DSfM dataset. The MAGSAC loss with covariance method outperforms all other methods in all error ranges. Soft L_1 loss with covariance performs well when the error is smaller than 5°, but not as good as the Soft L_1 loss baseline when error get larger. One reason is that there are still some outlier edges in the view graph with small uncertainty, in this case, the optimization process would be dragged to the sub-optimal results because of the lack of robustness of the Soft L_1 loss.

We also show the AUCs of rot. errors after the full global SfM pipeline. As shown in Tab. 2, MAGSAC with covariance is still the best. The strategy weighting by the inlier

Setting		Ori. Err Med (degree)	Pos. Err Med (meter)	# Reconst. Views	# Common Views	Time (minute)	Reproj. Err Avg (pixel)
COLMAP [29]		-	-	14097	-	2852	
Soft L$_1$ [5]	Baseline	1.44	1.63	9413	8824	27	1.24
	+ Inliers	1.72	1.65	9398	8806	34	1.21
	+ Covariance	1.45	1.60	9191	8668	29	1.18
MAGSAC [3]	Baseline	1.33	1.72	9404	8814	31	1.04
	+ Inliers	1.18	1.79	9424	8797	55	1.21
	+ Covariance	1.00	1.64	9361	8767	46	0.99

Table 7. Performance of different settings for full SfM pipeline. We report the median orientation (in degrees) and position errors (in meters), the number of reconstructed views and views common with the ground truth, the processing time (in minutes), and the reprojection average error (in pixels). The COLMAP reconstruction (with default parameters) is considered ground truth. Time is the total running time of all scenes in 1DSfM dataset without the time of the feature extraction and matching. It is measured on a laptop with Intel i7-12700H.

numbers is worse than the baseline with both losses. This means that in a large-scale dataset like 1DSfM, the final bundle adjustment step could make up the gap of the information provided by the inlier numbers, which means that a stronger uncertainty extraction method (covariance) is necessary to improve the reconstruction quality.

The results of ETH3D dataset are shown in the right plot of Fig. 5 and in Tab. 3. The combination of MAGSAC and the proposed uncertainties leads to the most accurate results. Another interesting observation is that our global SfM outperforms COLMAP method, confirming the potential in global SfMs over their incremental counterparts.

Tab. 4 shows the results of rotation averaging by different uncertainty-based weighting strategies. We compare using the trace and Frobenius-norm of the covariance matrix to using it directly. The direct strategy, in Eq. (3), clearly outperforms both other tested algorithms.

Robust Loss Functions. We compared the impact of using different loss functions for non-linear optimization in rotation averaging. In Tab. 5, we compare the MAGSAC loss function with other 7 popular loss functions on the scenes from the 1DSfM dataset. We report the AUC scores at $2°, 5°, 10°$, and $20°$. MAGSAC is the best by a large margin on AUC at $2°, 5°$ and $10°$. It is second best, in terms of AUC $20°$, being slightly behind Tukey weighting by 0.7 AUC points. For example, the AUC@$2°$ of MAGSAC is 5.5 points higher that of the second best, *i.e.*, Soft L$_1$.

Image Downsampling. We test the reconstruction on the ETH3D dataset by downsampling the input high resolution images. The AUC $2°$ scores are reported in Tab. 6 at different downsampling factors. We observe that the proposed pipeline is more accurate than the baseline in *all* cases. Actually, the larger the downsampling factor (*i.e.*, image quality reduction), the bigger the improvement compared to the

baseline [34]. This clearly demonstrates the potential in using the uncertainties of the estimated two-view geometries and the importance of choosing the best loss function.

Full Global SfM Pipeline Results. In Tab. 7, the results after the full global SfM pipeline (finishing with a bundle adjustment) are shown on the 1DSfM dataset. The lowest median orientation error is achieved by the proposed algorithm. The median position errors, the number of reconstructed views, and the number of views that are common with the ground truth, are similar in case of all methods.

Since the complexity of evaluating MAGSAC loss is marginally larger than that of the L$_1$ loss and, also, we need to invert and decompose the covariance matrix, the proposed algorithm is slightly (*i.e.*, by a few minutes) slower than other combinations. The benefit from global approaches is clear in this figure as their run-time is *two order-of-magnitude* lower than that of COLMAP. COLMAP runs for 48 hours. Global approaches run for 30 minutes.

5. Conclusion

In this paper, we revisit the rotation averaging problem applied in global Structure-from-Motion pipelines by leveraging uncertainties in two-view epipolar geometries and investigating robust losses. Our experiments demonstrate that integrating the covariance matrices of the uncertainties directly in the optimization procedure largely improves reconstruction quality. Moreover, carefully choosing the employed robust loss also gives a boost in the accuracy. We believe our work helps in closing the accuracy gap between incremental and global approaches.

References

[1] Sameer Agarwal, Keir Mierle, and The Ceres Solver Team. Ceres Solver, 3 2022.

[2] Mica Arie-Nachimson, Shahar Z Kovalsky, Ira Kemelmacher-Shlizerman, Amit Singer, and Ronen Basri. Global motion estimation from point matches. In *International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 2012.

[3] Daniel Barath, Jana Noskova, and Jiri Matas. Marginalizing sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[4] Michael J Black and Paul Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, 63(1):75–104, 1996.

[5] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on image processing*, 6(2):298–311, 1997.

[6] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *ICCV*, 2013.

[7] Avishek Chatterjee and Venu Madhav Govindu. Robust relative rotation averaging. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):958–972, 2017.

[8] Zhaopeng Cui and Ping Tan. Global structure-from-motion by similarity averaging. In *ICCV*, 2015.

[9] Frank Dellaert, David M Rosen, Jing Wu, Robert Mahony, and Luca Carlone. Shonan rotation averaging: global optimality by surfing $SO(p)^n$. In *European Conference on Computer Vision*, pages 292–308. Springer, 2020.

[10] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018.

[11] Anders Eriksson, Carl Olsson, Fredrik Kahl, and Tat-Jun Chin. Rotation averaging with the chordal distance: Global minimizers and strong duality. *IEEE TPAMI*, 2019.

[12] 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):381–395, 1981.

[13] Johan Fredriksson and Carl Olsson. Simultaneous multiple rotation averaging using lagrangian duality. In *ACCV*, 2012.

[14] Xiang Gao, Lingjie Zhu, Zexiao Xie, Hongmin Liu, and Shuhan Shen. Incremental rotation averaging. *IJCV*, 2021.

[15] Stuart Geman. Statistical methods for tomographic image reconstruction. *Bull. Int. Stat. Inst*, 4:5–21, 1987.

[16] Venu Madhav Govindu. Combining two-view constraints for motion estimation. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 2, pages II–II. IEEE, 2001.

[17] Venu Madhav Govindu. Lie-algebraic averaging for globally consistent motion estimation. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004.

[18] Richard Hartley, Khurrum Aftab, and Jochen Trumpf. L1 rotation averaging using the weiszfeld algorithm. In *CVPR*, 2011.

[19] Paul W Holland and Roy E Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–827, 1977.

[20] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

[21] Ricardo A Maronna, R Douglas Martin, Victor J Yohai, and Matías Salibián-Barrera. *Robust statistics: theory and methods (with R)*. John Wiley & Sons, 2019.

[22] Daniel Martinec and Tomas Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *CVPR*, 2007.

[23] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[24] Pierre Moulon, Pascal Monasse, and Renaud Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *ICCV*, 2013.

[25] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, 2016.

[26] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

[27] Carl Olsson and Olof Enqvist. Stable structure from motion for unordered image collections. In *Scandinavian Conference on Image Analysis*, pages 524–535. Springer, 2011.

[28] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020.

[29] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, pages 4104–4113, 2016.

[30] Thomas Schops, Johannes L Schonberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3260–3269, 2017.

[31] Chitturi Sidhartha and Venu Madhav Govindu. It is all in the weights: Robust rotation averaging revisited. In *2021 International Conference on 3D Vision (3DV)*, pages 1134–1143. IEEE, 2021.

[32] Noah Snavely, Steve Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *IJCV*, 80(2):189–210, 2008.

[33] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM siggraph*. 2006.

[34] Christopher Sweeney, Tobias Hollerer, and Matthew Turk. Theia: A fast and scalable structure-from-motion library. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 693–696, 2015.

[35] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, 1999.

[36] Kyle Wilson, David Bindel, and Noah Snavely. When is rotations averaging hard? In *ECCV*, 2016.

[37] K. Wilson and N. Snavely. Robust Global Translations with 1DSfM. In *ECCV*, pages 61–75, 2014.

[38] Changchang Wu. Towards linear-time incremental structure from motion. In *2013 International Conference on 3D Vision-3DV 2013*, pages 127–134. IEEE, 2013.