Global Structure-from-Motion Revisited

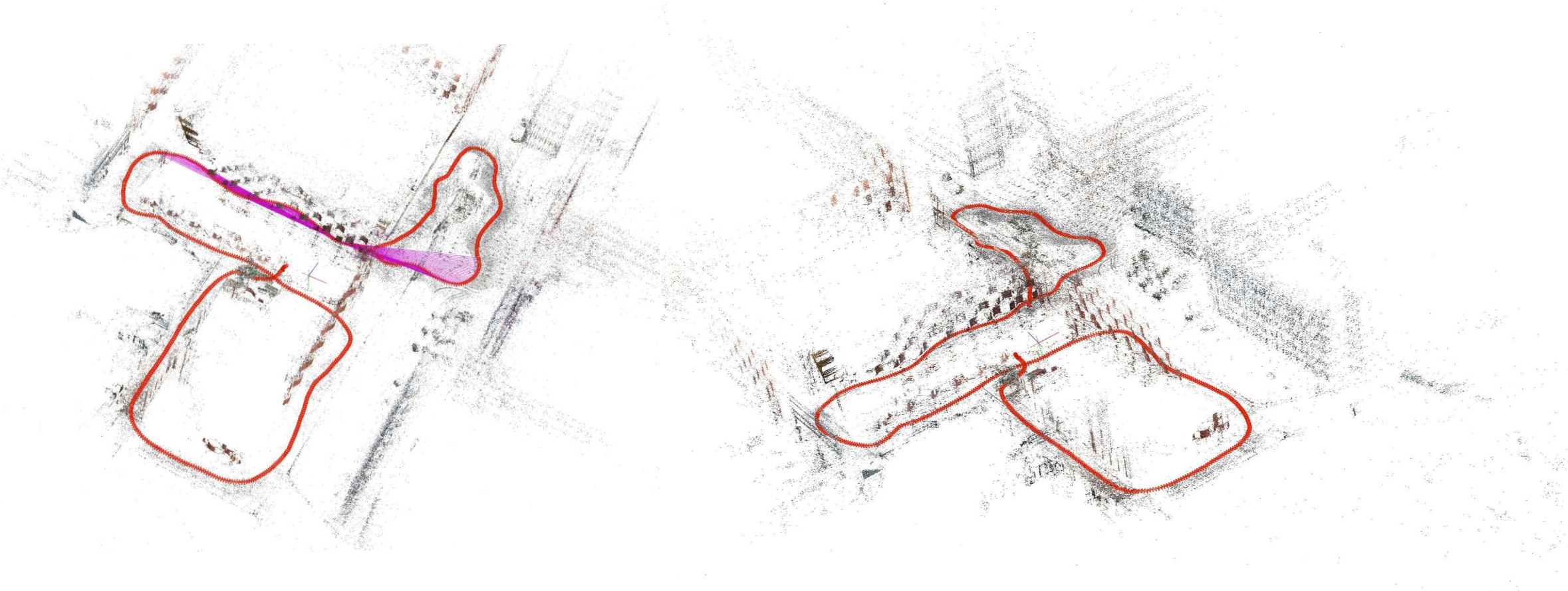
GLOMAP

Linfei Pan[1], Dániel Baráth[1], Marc Pollefeys[1,2], and Johannes L. Schönberger[2]
[1] ETH Zurich [2] Microsoft

Agenda

- Short Introduction
- Literature Review of Global Structure-from-Motion
- GLOMAP – Technical contributions
- Results
- DEMO
- Questions

Structure from Motion



Two main categories

INCREMENTAL APPROACHES

- Superior accuracy
- Superior Robustness
- Computationally expensive

e.g. COLMAP

GLOBAL APPROACHES

- More scalable
- More efficient
- Less accurate/robust

e.g Theia

Structure from Motion

Image-based feature extraction and matching

Two-view geometry estimation

INCREMENTAL APPROACHES

- Starts from two views and sequentially expands by adding images and their 3D structure
- At each/few steps we need to run absolute camera pose estimation, triangulation and BUNDLE ADJUSTMENT (~O^3)

GLOBAL APPROACHES

- Camera geometry for all images are computed at once by considering ALL two-view geometries in the view graph (R+t)
- The globally estimated camera geometry is used as initialization for the triangulation of 3D structure
- Final global BUNDLE ADJUSTMENT step

Where is the accuracy gap?

The main reason for the accuracy and robustness gap between incremental and global SfM lies in the global translation averaging step. Translation averag-

3 Major challenges in this step:

1. Relative translation from two-view geometry can only be estimated up to a scale. IF the cameras form a skewed triangle --> scales are especially prune to noise in the observations
2. Accurately decomposing two-view geometry into R+t components requires accurate prior knowledge of camera intrinsics --> Otherwise estimated translation direction is often subject to large errors
3. Co-linear motion patterns, especially in sequential datasets (e.g handled videos or self-driving cars).

GLOMAP's approach

global SfM system, termed GLOMAP. The core difference to previous global SfM systems lies in the step of global positioning. Instead of first performing ill-posed translation averaging followed by global triangulation, our proposed method performs joint camera and point position estimation. GLOMAP achieves a similar level of robustness and accuracy as state-of-the-art incremental SfM

Review of Global Structure-from-Motion

Correspondence Search (like in COLMAP)

1. Feature extraction
2. Search for correspondances between image pairs
3. This yields either a homography H_ij or F_ij(uncalibrated) or E_ij
4. The computed two-view geometries with associated inlier correspondences define the view graph G that will be the input to the global reconstruction step

(GLOMAP uses the same code for this part)

Global Camera Pose Estimation

gulation and bundle adjustment, global SfM seeks to estimate all the camera poses $\boldsymbol{P}_i = (\boldsymbol{R}_i, \boldsymbol{c}_i) \in \mathrm{SE}(3)$ at once using the view graph $\mathcal{G}$ as input. To make

geometries [35, 36]. The main challenge lies in dealing with noise and outliers in the view graph by careful modeling and solving of the optimization problems.

Rotation Averaging

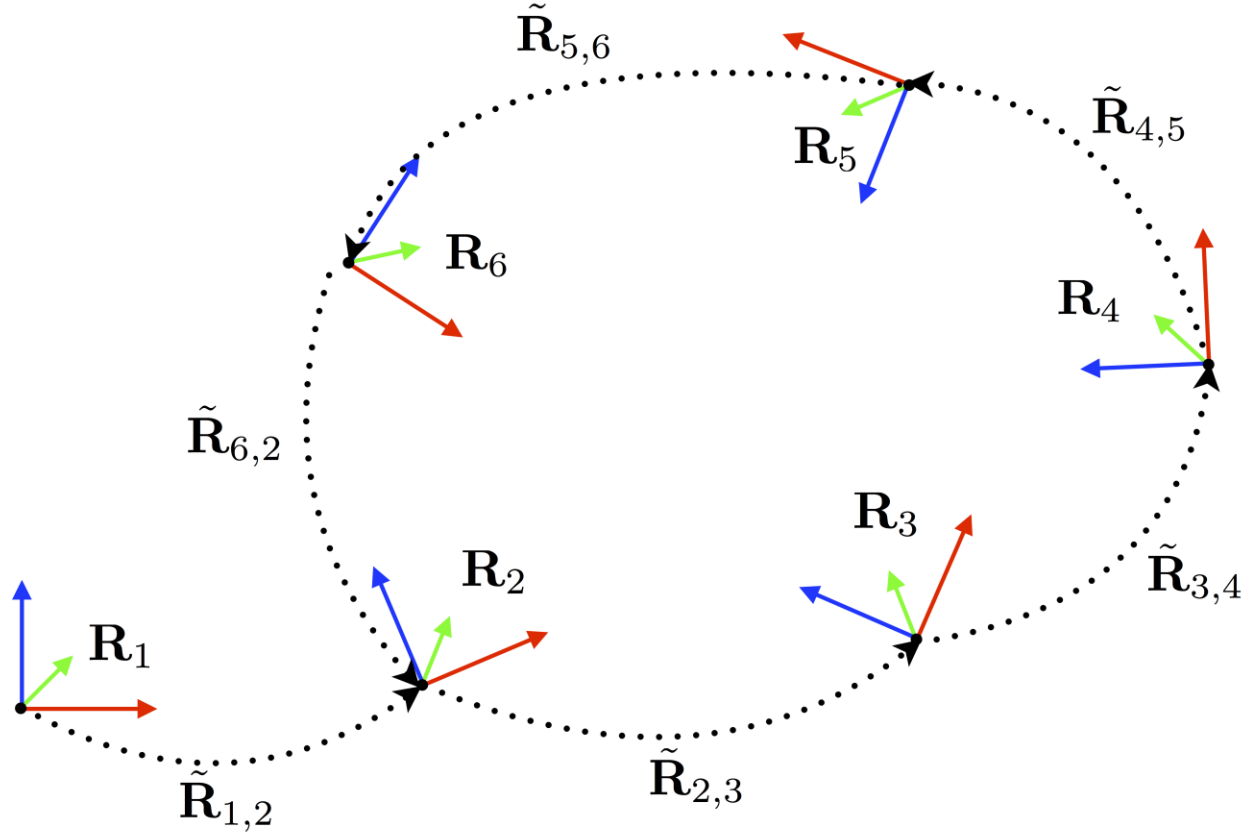
$$\boldsymbol{R}_{ij} = \boldsymbol{R}_j \boldsymbol{R}_i^\top .$$

$$\arg\min_{\boldsymbol{R}} \sum_{i,j} \rho\left(d(\boldsymbol{R}_j^\top \boldsymbol{R}_{ij} \boldsymbol{R}_i, \boldsymbol{I})^p\right) .$$

Robustifier e.g. Huber

Distance metric e.g geodesic distance

control the sensitivity of the distance



Rotation Averaging

Efficient and Robust Large-Scale Rotation Averaging

Avishek Chatterjee Venu Madhav Govindu
Department of Electrical Engineering
Indian Institute of Science, Bengaluru, INDIA
{avishek|venu}@ee.iisc.ernet.in

Translation Averaging

- The rotations R_I can be factored out from camera poses .
- Only c_i needs to be determined
- The problem is described as estimating the global camera positions that are maximally consistent with pairwise relative translation t_ij, where: $\boldsymbol{t}_{ij} = \frac{\boldsymbol{c}_j - \boldsymbol{c}_i}{\|\boldsymbol{c}_j - \boldsymbol{c}_i\|}.$
- Due to the noise, however, outliers and scale of relative translations -> the task becomes especially challenging

Translation Averaging

- Parallel rigidity is the property NEEDED to uniquely determine camera poses
- translation averaging generally only works reliably when the view graph is well connected.
- problem is also inherently ill-posed and sensitive to noisy measurements when cameras are subject to or close to co-linear motion.
- is only possible with known camera intrinsics. When such information is inaccurate, the extracted translations are not reliable
- **GLOMAP skips the step of translation averaging**
- **Directly to global positioning STEP**

Structure for Camera Pose Estimation

era positions with a linear global translation constraint imposed by points. The common theme of these works is that incorporating constraints on the 3D scene structure aids the robustness and accuracy of camera position estimation, which we take as an inspiration for our work.

Global Triangulation

- After recovering the cameras, the global 3D structure can be obtained via triangulation

Global Bundle Adjustment

$$\underset{\pi,\boldsymbol{P},\boldsymbol{X}}{\arg\min} \sum_{i,k} \rho\left(\left\|\pi_i(\boldsymbol{P}_i, \boldsymbol{X}_k) - x_{ik}\right\|_2\right).$$

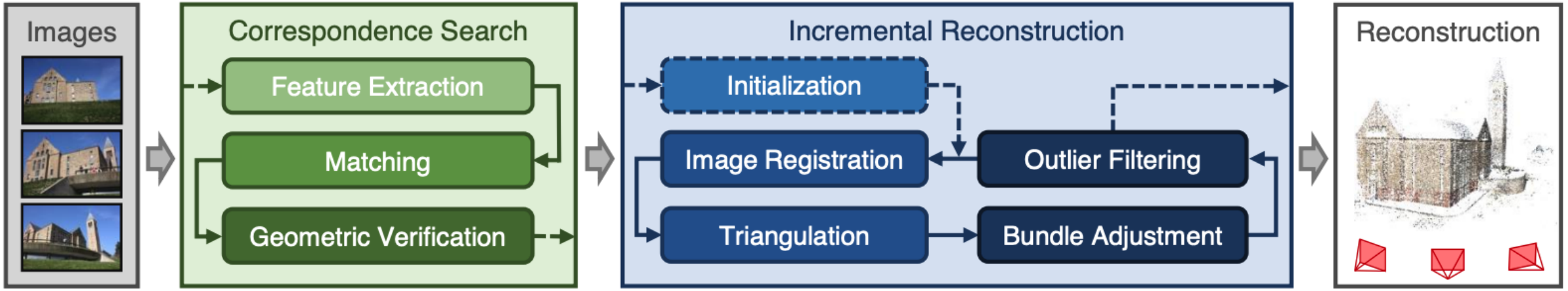
final 3D structure $\boldsymbol{X}_k \in \mathbb{R}^3$, camera extrinsics $\boldsymbol{P}_i$ and camera intrinsics π_i.

Hybrid Structure-from-Motion

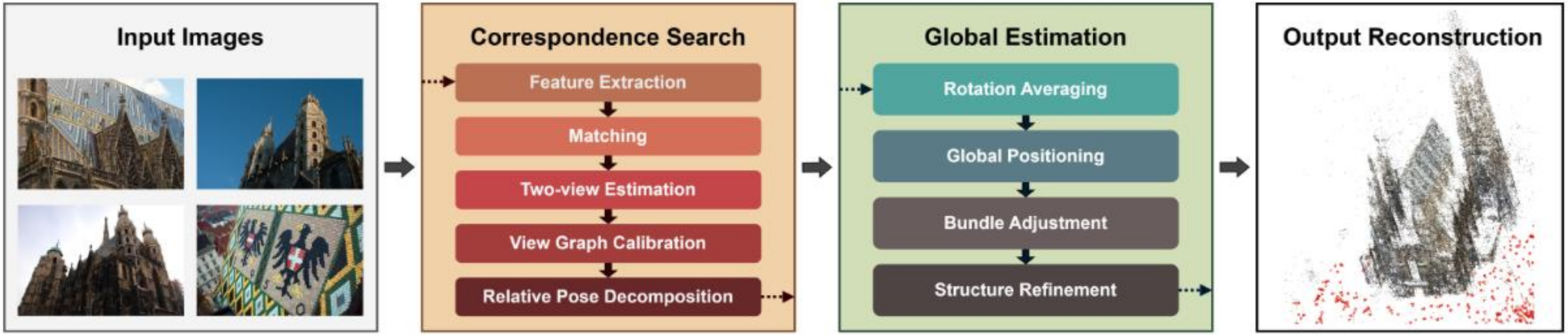
However, such methods are still not applicable when camera intrinsics are inaccurate according to their formulation. Our method overcomes this limitation by different modeling of the objective in the global positioning step.

Technical contributions

COLMAP



GLOMAP



Feature track construction

- Carefully reconstructed
- Only consider inlier feature correspondences produced by two-view geometry verification
- Performing the Cheirality test (in front or not of the camera)
- Matches that are close to the Epipoles or have a small triangulation angles are also removed to avoid singularities
- After filtering of ALL view graph edges -> forming feature tracks by concatenating all remaining matches

Rotation Averaging

Efficient and Robust Large-Scale Rotation Averaging

Avishek Chatterjee Venu Madhav Govindu
Department of Electrical Engineering
Indian Institute of Science, Bengaluru, INDIA
{avishek|venu}@ee.iisc.ernet.in

Rotation Averaging

Algorithm 1 Lie-Algebraic Relative Rotation Averaging

Input: $\{\mathbf{R}_{ij1}, \cdots, \mathbf{R}_{ijk}\}$ ($|\mathcal{E}|$ relative rotations)
Output: $\mathbf{R}_{global} = \{\mathbf{R}_1, \cdots, \mathbf{R}_N\}$ ($|\mathcal{V}|$ absolute rotations)
Initialisation: $\mathbf{R}_{global}$ to an initial guess

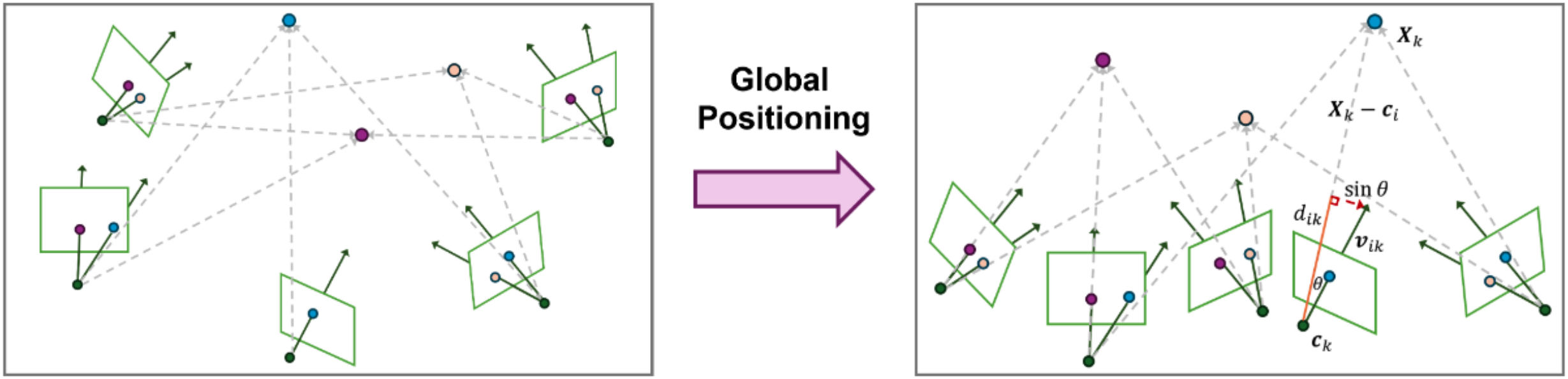
while $||\Delta\boldsymbol{\omega}_{rel}|| < \epsilon$ **do**

1. $\Delta\mathbf{R}_{ij} = \mathbf{R}_j{}^{-1}\mathbf{R}_{ij}\mathbf{R}_i$
2. $\Delta\boldsymbol{\omega}_{ij} = \log(\Delta\mathbf{R}_{ij})$
3. Solve $\mathbf{A}\Delta\boldsymbol{\omega}_{global} = \Delta\boldsymbol{\omega}_{rel}$
4. $\forall k \in [1, N], \mathbf{R}_k = \mathbf{R}_k exp(\Delta\boldsymbol{\omega}_k)$

end while

Global positioning of cameras and points

jointly recover point and camera positions



Global positioning of cameras and points

inal formulation was proposed in terms of the relative translations, whereas, in our formulation, we discard the relative translation constraints and only include camera ray constraints. Concretely, our problem is modeled and optimized as:

$$\arg\min_{\boldsymbol{X},\boldsymbol{c},d} \sum_{i,k} \rho\left(\|\boldsymbol{v}_{ik} - d_{ik}(\boldsymbol{X}_k - \boldsymbol{c}_i)\|_2\right), \quad \text{subject to} \quad d_{ik} \geq 0, \tag{3}$$

Global positioning of cameras and points

inal formulation was proposed in terms of the relative translations, whereas, in our formulation, we discard the relative translation constraints and only include camera ray constraints. Concretely, our problem is modeled and optimized as:

$$\arg\min_{\boldsymbol{X},\boldsymbol{c},d} \sum_{i,k} \rho\left(\|\boldsymbol{v}_{ik} - d_{ik}(\boldsymbol{X}_k - \boldsymbol{c}_i)\|_2\right), \quad \text{subject to} \quad d_{ik} \geq 0, \tag{3}$$

error is strictly bounded to the range [0,1]

$$\begin{cases} \sin\theta & \text{if} \quad \theta \in [0, \pi/2), \\ 1 & \text{if} \quad \theta \in [\pi/2, \pi], \end{cases}$$

where θ is the angle between v_ik and X_k −c_i for optimal d_ik

Global positioning of cameras and points

2 Key advantages over the classical translation averaging:

1. Applicability of the method on datasets with inaccurate or unknown camera intrinsics
2. Applicability of GLOBEM in co-linear motion scenarios (which is known degenerate case for translation averaging)

Global bundle adjustment

- Global positioning step is robust estimation but LIMITED in accuracy (especially for unknown camera intrinsics)
- THEREFORE several rounds of global bundle adjustment are performed (Levenberg-Marquardt and the Huber loss)
- First camera rotations are fixed, then jointly optimized with intrinsics and points. -> IMPORTANT for sequential data

Iterations are halted when the ratio of filtered tracks falls below 0.1%

Camera clustering (images collected from web)

- Maybe from images from the internet we have images from places very far away but that will be put in a single reconstruction.
- Therefore, they remove all the pairs of images from edges that have fewer that 5 visible points for each image
- Calculate the median of points for the remaining pairs and set it as τ .
- Now they only use strongly connected components/images with more than τ counts.
- Then they try to merge 2 strong components if there are at least two edges with more than $0.75\,\tau$ counts.
- Each connected component is then output as a separate reconstruction

Results

Experiments – results ETH3D SLAM

- challenging dataset containing sequential data with sparse features, dynamic objects, and drastic illumination changes
- "We evaluated our method on the training sequences that come with millimeter-accurate ground truth"
- Each row in the table averages the results across sequences sharing the same prefix with full results in the suppl. material

Experiments – results ETH3D SLAM (link)

Table 1: Result on ETH3D SLAM [65] dataset. Each row represents the average results on scenes with the same prefix. The proposed system outperforms global SfM baselines by a large margin, and also achieves better results than COLMAP [62].

	Recall @ 0.1m				AUC @ 0.1m				AUC @ 0.5m				Time (s)			
	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP
cables	76.8	88.0	88.0	68.0	59.1	60.2	76.2	56.5	77.4	82.4	85.6	70.4	103.1	339.5	195.6	2553.4
camera	2.2	26.6	32.6	34.8	2.0	12.9	21.9	24.7	2.2	26.2	30.8	33.2	1.5	5.3	10.0	196.2
ceiling	6.4	22.8	28.7	28.6	2.9	17.0	22.3	15.3	8.2	21.7	27.4	26.3	78.3	52.6	111.0	1057.7
desk	28.0	32.2	32.3	24.4	16.0	29.2	28.5	21.1	28.6	31.6	31.6	23.7	376.2	195.3	150.0	1115.1
einstein	32.9	47.8	48.5	33.6	22.7	32.1	36.5	25.4	34.2	46.7	47.9	36.7	150.3	70.5	142.1	1230.8
kidnap	73.1	73.3	73.3	73.3	63.4	62.3	70.3	68.5	71.2	71.1	72.7	72.3	114.4	356.7	144.3	731.2
large	35.4	48.6	49.0	44.5	18.1	37.8	45.8	20.7	33.7	46.6	48.4	43.4	91.9	60.2	77.6	983.8
mannequin	49.1	62.2	67.4	59.3	36.7	46.5	61.4	52.8	49.1	59.7	66.4	58.4	33.5	29.7	44.3	301.2
motion	18.8	16.9	39.8	17.7	11.0	11.9	22.5	12.9	23.3	19.2	45.9	19.7	859.7	109.0	788.9	9995.1
planar	30.6	100.0	100.0	100.0	12.5	97.8	98.7	98.3	38.0	99.6	99.7	99.7	313.8	167.5	533.3	2349.7
plant	77.0	89.1	93.3	92.8	62.9	75.3	82.0	82.3	77.1	88.2	93.4	92.5	21.7	35.7	28.7	202.7
reflective	12.6	16.1	22.0	26.2	6.7	9.0	12.1	9.2	16.5	23.0	31.3	33.5	721.3	118.3	434.4	6573.9
repetitive	26.3	28.5	32.7	28.5	23.9	15.2	29.2	27.2	25.8	27.0	32.0	28.3	63.2	136.8	74.5	561.1
sfm	83.2	94.1	97.0	55.9	57.9	53.7	79.6	35.7	80.0	88.7	95.2	55.8	91.7	170.5	239.7	469.6
sofa	11.2	22.0	23.9	32.2	5.5	13.1	22.1	28.6	10.3	21.3	23.5	31.6	9.1	8.9	10.1	157.3
table	79.1	93.7	94.3	99.9	68.2	69.2	84.3	95.6	76.9	89.2	92.3	99.1	182.0	97.8	221.5	2777.7
vicon	64.6	84.2	97.0	81.1	20.4	57.1	80.5	38.9	71.7	87.6	93.7	75.0	50.9	88.2	46.2	474.8
Average	48.2	62.8	66.4	57.9	34.9	46.0	57.0	47.6	48.6	61.1	65.7	57.9	120.8	91.8	133.5	1115.4

Experiments – results ETH3D MVS (rig)

- contains, per scene, about 1000 multi-rig exposures with each 4 images.
- The dataset contains both outdoor and indoor scenes with millimeter-accurate ground truth for 5 training sequences.
- 1 failure from COLMAP

Experiments – results ETH3D MVS (rig)

Table 2: Results on ETH3D MVS (rig) [66]. Proposed GLOMAP largely outperforms other SfM systems by a large margin while maintaining the efficiency of global SfM.

	AUC @ 1°				AUC @ 3°				AUC @ 5°				Time (s)			
	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP
delivery_area	0.0	47.8	75.9	66.6	0.1	81.0	91.2	87.5	0.3	88.2	94.6	92.2	235.7	259.5	519.9	1745.9
electro	0.2	25.9	47.5	38.4	1.8	61.6	72.9	65.2	2.9	73.5	81.7	75.2	99.9	151.8	429.2	3283.1
forest	0.0	65.6	74.1	74.7	0.0	87.1	90.0	90.3	0.1	91.8	93.5	93.7	306.6	563.2	1658.4	7571.6
playground	0.0	23.1	40.0	0.0	0.1	62.1	72.5	0.0	0.2	74.1	81.0	0.0	121.5	598.6	1008.3	350.2
terrains	0.7	39.6	50.2	48.0	2.6	71.2	78.1	76.8	3.1	79.6	85.3	84.3	62.6	179.4	353.0	1333.9
Average	0.2	40.4	57.6	45.5	0.9	72.6	80.9	64.0	1.3	81.4	87.2	69.1	165.3	350.5	793.8	2857.0

https://www.eth3d.net/view_multi_view_result?dataset=9&id=11&tid=1 (failure case for COLMAP visualized)

Experiments – results ETH3D MVS (dslr)

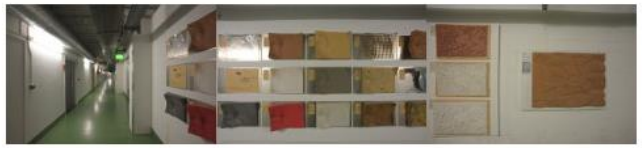










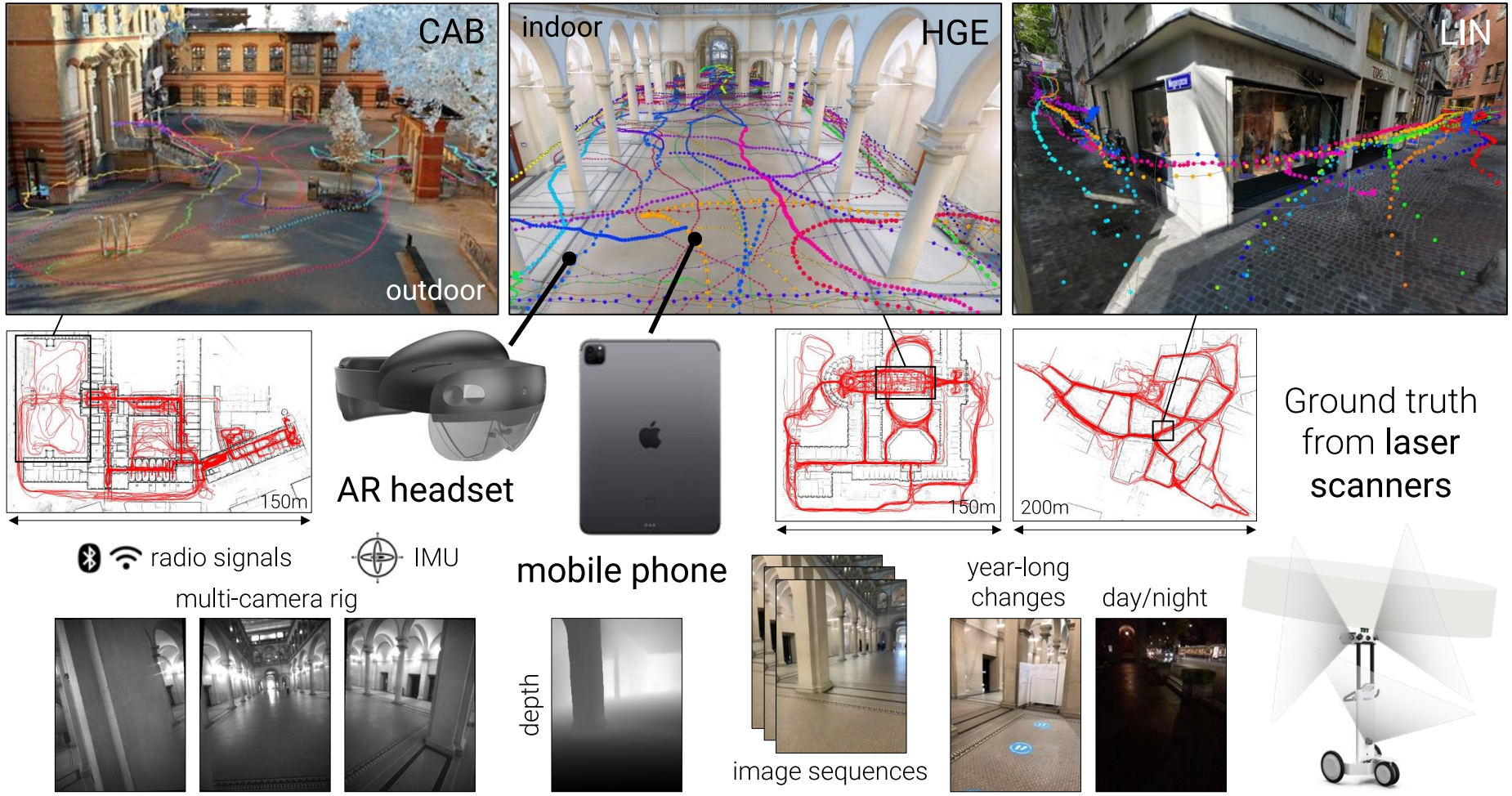



Experiments – results ETH3D MVS (DSLR)

Table 3: Results on ETH3D MVS (DSLR) [66]. On this dataset, the proposed method outperforms other global SfM baselines and is comparable to COLMAP [62].

	AUC @ 1°				AUC @ 3°				AUC @ 5°				Time (s)			
	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP
courtyard	68.2	91.2	87.8	87.3	81.7	97.0	95.9	95.8	85.9	98.2	97.6	97.5	11.2	11.9	24.5	39.2
delivery_area	91.4	92.0	92.8	92.3	97.1	97.3	97.6	97.4	98.3	98.4	98.6	98.5	235.7	3.8	9.2	26.4
electro	72.6	49.8	79.3	70.0	81.8	54.9	86.7	75.7	83.9	57.9	88.5	77.0	99.9	2.8	7.5	23.9
facade	89.4	88.1	91.1	90.3	96.4	94.3	97.0	96.7	97.8	95.5	98.2	98.0	35.1	47.4	91.4	113.5
kicker	82.0	71.0	86.3	86.6	89.6	77.6	94.3	91.3	91.2	79.2	96.5	92.2	1.7	3.5	6.4	16.2
meadow	10.5	17.3	74.7	61.6	13.8	21.5	90.4	77.8	14.7	23.4	94.2	81.5	0.3	1.8	2.1	5.1
office	19.3	27.6	59.6	45.2	24.1	32.8	81.8	56.5	25.6	35.5	88.6	59.9	0.4	0.7	1.3	19.1
pipes	29.4	41.0	89.8	86.3	36.3	53.5	96.6	95.4	41.1	56.7	98.0	97.3	0.3	0.6	0.8	3.4
playground	49.4	72.3	91.2	90.6	55.6	77.8	97.0	96.8	56.9	78.9	98.2	98.1	121.5	3.0	7.7	29.8
relief	89.8	66.9	93.7	93.3	96.6	73.0	97.9	97.8	97.9	76.4	98.7	98.7	3.7	4.7	12.7	34.1
relief_2	11.9	94.1	95.0	94.9	12.4	98.0	98.3	98.2	12.5	98.8	99.0	98.9	0.7	3.2	5.4	24.1
terrace	91.9	94.0	92.8	92.3	97.3	98.0	97.6	97.4	98.4	98.8	98.6	98.5	1.1	1.6	4.0	10.9
terrains	70.9	86.4	82.2	81.9	89.3	95.3	93.7	93.6	93.4	97.1	96.2	96.1	62.6	2.1	6.8	21.7
botanical_garden	26.0	51.0	87.2	5.1	30.1	74.5	95.6	5.3	30.9	82.4	97.3	5.4	1.0	1.0	4.2	9.0
boulders	89.6	89.8	91.0	90.6	96.4	96.5	97.0	96.9	97.9	97.9	98.2	98.1	4.1	2.6	8.3	18.5
bridge	44.9	89.8	91.8	91.6	47.9	95.3	97.2	97.2	48.5	96.4	98.3	98.3	35.1	31.3	87.2	91.9
door	93.8	93.8	95.1	96.9	97.9	97.9	98.4	99.0	98.8	98.8	99.0	99.4	1.2	1.0	1.8	4.2
exhibition_hall	11.0	84.4	25.5	85.4	15.2	93.5	29.7	94.3	16.3	96.0	30.7	96.5	20.6	49.8	68.3	72.0
lecture_room	80.2	69.4	83.3	82.9	92.7	79.6	94.1	94.1	95.6	82.6	96.5	96.4	3.3	2.4	7.4	10.9
living_room	88.0	84.8	88.0	88.3	95.8	92.6	95.8	95.8	97.4	94.3	97.4	97.5	9.4	10.5	22.5	49.1
lounge	34.1	34.1	34.0	33.9	35.4	35.4	35.3	35.3	35.6	35.6	35.6	35.6	0.4	1.0	1.2	1.8
observatory	58.4	65.8	65.4	64.4	83.7	87.1	87.0	86.5	89.9	92.2	92.1	91.8	5.8	4.8	13.2	24.4
old_computer	23.3	49.8	53.8	78.7	41.5	59.8	61.4	90.2	48.4	62.0	63.0	92.7	5.9	4.1	8.9	19.6
statue	96.4	98.8	98.8	98.7	98.8	99.6	99.6	99.6	99.3	99.8	99.8	99.7	0.9	1.6	5.7	7.4
terrace_2	87.9	90.8	91.0	90.7	95.7	96.9	96.9	96.8	97.4	98.1	98.2	98.1	1.0	1.5	3.3	10.2
Average	60.4	71.8	80.8	79.2	68.1	79.2	88.5	86.5	70.1	81.2	90.3	88.1	26.5	7.9	16.5	27.5

Experiments – results LaMAR



Experiments – results LaMAR

Table 4: Results on LaMAR [60] datasets. The proposed method largely outperforms other baselines as well as COLMAP [62]. For LIN, structure refinement is not performed for GLOMAP due to memory limitation (marked as *).

	Recall @ 1m				AUC @ 1m				AUC @ 5m				Time (s)			
	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP
CAB	-	6.0	11.6	13.0	-	3.2	4.7	5.8	-	9.3	16.9	19.2	-	1345.6	6162.2	194033.6
HGE	-	8.3	48.4	38.9	-	2.9	22.2	18.0	-	9.4	50.3	46.9	-	1182.4	12587.2	249771.1
LIN*	-	18.8	87.3	44.2	-	7.0	46.7	17.7	-	38.5	85.6	52.3	-	2097.9	18466.6	620176.4
Average	-	11.0	49.1	32.0	-	4.4	24.5	13.8	-	19.1	50.9	39.4	-	1542.0	12405.3	354660.4

Experiments – results IMC 2023

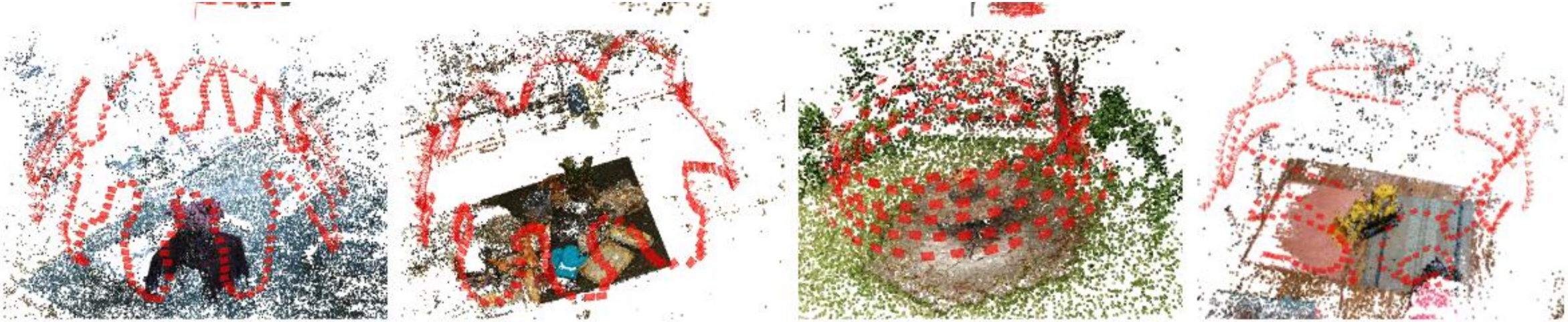
- unordered image collections over complex scenes
- various sources and often lack prior camera intrinsics
- ground truth of the dataset is built by COLMAP

Experiments – results IMC 2023

Table 5: Results on IMC 2023 [16]. Our GLOMAP method comes close to COLMAP generated ground truth while outperforming global SfM by a large margin.

	AUC @ 3°				AUC @ 5°				AUC @ 10°				Time (s)			
	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP
bike	-	0.0	35.0	77.9	-	0.0	38.9	86.7	-	0.0	41.9	93.4	-	1.4	1.5	1.1
chairs	-	0.0	82.6	0.8	-	0.0	89.6	0.8	-	0.0	94.8	0.8	-	1.7	1.4	0.6
fountain	22.1	57.1	91.2	91.3	24.2	61.6	94.7	94.8	25.7	64.9	97.4	97.4	0.8	1.4	3.4	5.9
cyprus	1.6	17.3	67.1	45.7	1.7	21.8	73.8	48.9	1.7	28.0	80.5	51.4	0.3	1.3	2.6	11.1
dioscuri	0.4	1.7	59.4	58.7	0.4	2.5	61.9	61.4	0.5	4.5	64.4	63.9	41.9	24.7	115.6	156.3
wall	57.1	84.6	95.3	88.6	73.9	88.9	97.2	93.2	87.0	92.1	98.6	96.6	23.8	28.1	77.8	63.5
kyiv-puppet-theater	0.7	1.0	10.0	0.3	0.7	1.1	12.0	0.3	0.9	1.5	19.5	0.3	0.3	2.5	2.6	0.3
brandenburg_gate	21.4	42.3	68.7	70.1	35.0	52.8	75.2	77.2	53.6	65.2	81.5	83.9	1171.8	199.9	368.4	1472.1
british_museum	18.5	34.9	62.0	61.6	32.0	47.4	72.7	72.7	51.3	63.8	83.5	83.9	78.1	84.6	117.6	318.0
buckingham_palace	4.1	26.0	85.9	80.5	12.5	37.2	89.1	86.2	34.0	53.0	92.1	91.1	429.7	173.3	484.1	4948.2
colosseum_exterior	37.3	69.0	80.5	80.7	52.8	77.1	85.8	86.1	69.2	84.7	90.5	90.8	542.8	489.2	767.9	3561.7
grand_place_brussels	18.3	34.7	71.8	68.7	32.4	50.8	78.5	76.6	50.5	67.6	84.6	84.2	124.2	87.6	220.7	520.3
lincoln_..statue	1.0	30.7	68.4	67.2	4.4	36.3	72.5	71.5	23.8	41.7	76.0	75.3	99.9	46.1	103.9	362.3
notre_dame_..facade	32.5	52.3	67.2	69.1	43.0	59.1	70.8	72.5	55.2	65.5	74.3	75.5	2592.1	2393.9	4146.9	52135.4
pantheon_exterior	49.6	68.3	77.0	79.4	62.3	74.9	81.8	83.5	74.4	81.2	86.2	87.2	285.7	169.3	488.8	1454.6
piazza_san_marco	2.6	48.6	72.7	58.0	6.4	62.0	82.3	71.0	23.7	75.9	90.6	83.7	16.7	12.6	43.4	74.2
sacre_coeur	37.1	73.7	78.8	78.5	50.8	77.5	81.1	80.9	64.8	80.9	83.0	82.8	196.1	130.4	266.9	682.2
sagrada_familia	30.3	50.7	53.8	54.3	39.2	55.5	58.7	58.9	48.9	60.0	62.9	62.9	42.6	48.6	140.9	237.0
st_pauls_cathedral	1.9	60.5	74.2	72.7	6.5	70.4	80.2	78.9	25.8	79.7	85.8	85.0	61.9	45.0	101.0	241.4
st_peters_square	31.1	56.3	79.5	83.6	46.0	66.8	84.2	87.8	64.0	77.5	88.8	91.6	961.0	621.1	1177.9	6051.9
taj_mahal	38.6	58.6	72.1	68.9	51.0	65.7	77.3	75.5	64.9	73.3	82.4	81.7	380.1	379.0	630.7	4528.9
trevi_fountain	43.4	64.6	79.0	80.5	56.2	72.7	82.8	84.4	69.3	80.3	86.4	87.8	1202.9	669.6	1676.0	12294.4
Average	20.4	42.4	69.6	65.3	28.7	49.2	74.6	70.4	40.4	56.4	79.4	75.1	412.6	255.1	497.3	4051.0

Experiments – results MIP360



	AUC @ 3°				AUC @ 5°				AUC @ 10°				Time (s)			
	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP	OpenMVG	Theia	GLOMAP	COLMAP
bicycle	89.2	12.0	95.8	95.8	92.7	14.7	97.5	97.5	95.3	17.9	98.7	98.7	123.1	28.6	66.9	120.6
bonsai	9.4	87.0	98.5	92.6	27.9	91.8	99.1	95.6	61.2	95.6	99.5	97.8	176.0	194.5	467.5	662.5
counter	96.9	98.9	99.3	99.2	98.1	99.4	99.6	99.5	99.1	99.7	99.8	99.8	46.6	71.0	203.9	270.5
garden	95.0	36.8	97.3	97.3	97.0	37.9	98.4	98.4	98.5	38.9	99.2	99.2	40.2	39.6	128.9	291.8
kitchen	88.5	93.5	94.8	94.9	93.1	95.8	96.9	97.0	96.5	97.6	98.4	98.5	127.0	187.3	426.9	619.1
room	39.6	26.0	97.7	96.2	42.2	26.8	98.6	97.7	44.9	27.6	99.3	98.9	96.3	85.6	216.3	371.6
stump	95.3	7.1	99.1	99.1	97.1	7.5	99.5	99.5	98.5	8.0	99.7	99.7	38.7	10.9	36.6	83.9
Average	73.4	51.6	97.5	96.5	78.3	53.4	98.5	97.9	84.9	55.0	99.2	98.9	92.6	88.2	221.0	345.7

4.4 Limitations

Though generally achieving satisfying performance, there still remain some failure cases. The major cause is a failure of rotation averaging, *e.g.*, due to symmetric structures (see *Exhibition_Hall* in Table 3). In such a case, our method could be combined with existing approaches like Doppelganger [11]. Also, since we rely on traditional correspondence search, incorrectly estimated two-view geometries or the inability to match image pairs altogether (*e.g.*, due to drastic appearance or viewpoint changes) will lead to degraded results or, in the worst case, catastrophic failures.

GLOMAP – main take aways

- The algorithm discard separate step for translation averaging and 3D reconstruction
- Replaced with the single global positioning step
- Result on-par (or superior) to COLMAP
- (random initialization -> not deterministic)
- 10-100x faster
- Code is open-source under a commercially friendly license

New SfM in town – different approach (global)

MASt3R-SfM: a Fully-Integrated Solution for Unconstrained Structure-from-Motion

Scenes	COLMAP [46]		ACE-Zero [9]		FlowMap [50]		VGGSfM [62]		DF-SfM [20]		**MASt3R-SfM**	
	RRA@5	RTA@5	RRA@5	RTA@5	RRA@5	RTA@5	RRA@5	RTA@5	RRA@5	RTA@5	RRA@5	RTA@5
courtyard	56.3	60.0	4.0	1.9	7.5	3.6	50.5	51.2	80.7	74.8	89.8	64.4
delivery area	34.0	28.1	27.4	1.9	29.4	23.8	22.0	19.6	82.5	82.0	83.1	81.8
electro	53.3	48.5	16.9	7.9	2.5	1.2	79.9	58.6	82.8	81.2	100.0	95.5
facade	92.2	90.0	74.5	64.1	15.7	16.8	57.5	48.7	80.9	82.6	74.3	75.3
kicker	87.3	86.2	26.2	16.8	1.5	1.5	100.0	97.8	93.5	91.0	100.0	100.0
meadow	0.9	0.9	3.8	0.9	3.8	2.9	100.0	96.2	56.2	58.1	58.1	58.1
office	36.9	32.3	0.9	0.0	0.9	1.5	64.9	42.1	71.1	54.5	100.0	98.5
pipes	30.8	28.6	9.9	1.1	6.6	12.1	100.0	97.8	72.5	61.5	100.0	100.0
playground	17.2	18.1	3.8	2.6	2.6	2.8	37.3	40.8	70.5	70.1	100.0	93.6
relief	16.8	16.8	16.8	17.0	6.9	7.7	59.6	57.9	32.9	32.9	34.2	40.2
relief 2	11.8	11.8	7.3	5.6	8.4	2.8	69.9	70.3	40.9	39.1	57.4	76.1
terrace	100.0	100.0	5.5	2.0	33.2	24.1	38.7	29.6	100.0	99.6	100.0	100.0
terrains	100.0	99.5	15.8	4.5	12.3	13.8	70.4	54.9	100.0	91.9	58.2	52.5
Average	49.0	47.8	16.4	9.7	10.1	8.8	65.4	58.9	74.2	70.7	81.2	79.7

Table 3: **Detailed per-scene translation and rotation accuracies (↑) on ETH-3D.** For clarity, we color-code results with a linear gradient between the worst and best result for a given scene.

DEMO