Stable and Scalable Bundle Adjustment of Holistic 3D Structures

Shaohui Liu, Rémi Pautrat, Daniel Barath, Richard Hartley, Viktor Larsson, Marc Pollefeys

ETH zürich · Lund University · Australian National University · Microsoft

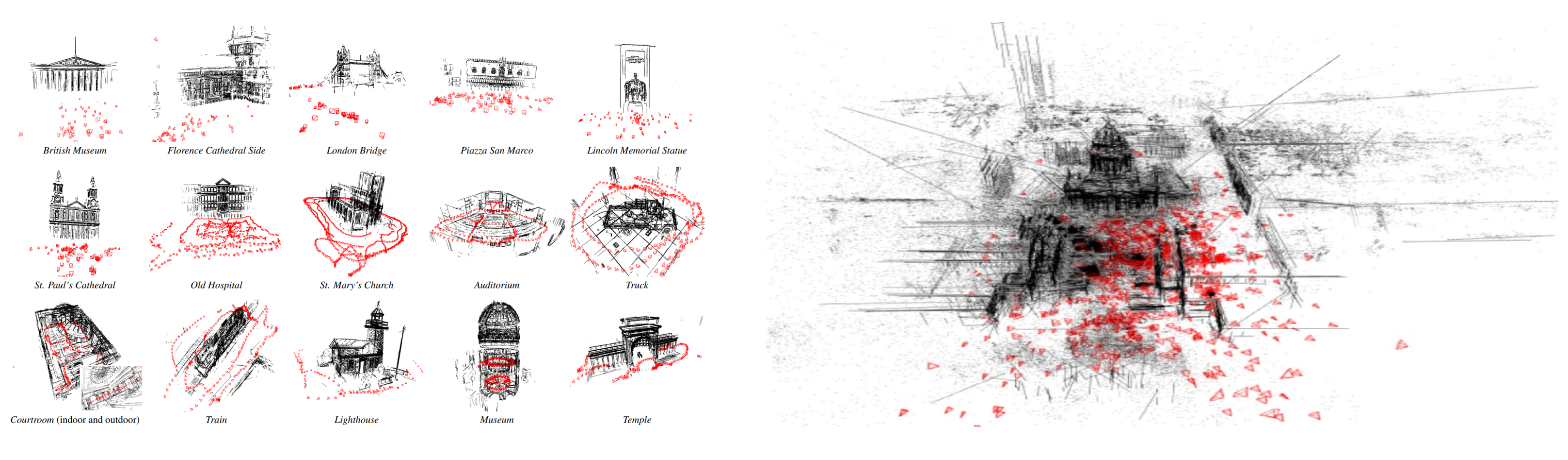
ECCV 2026 | MALMÖ | SEPT 8-12

Code available

Instead of individual set of 3D points, man-made structured environments are full of mid-level 3D structures and relations:

- **Feature units:** points, lines
- **Primitives:** VPs, planes, spheres, cylinders, cones, etc.
- **Relational constraints:** parallelism, orthogonality, coplanarity, wireframes, etc.

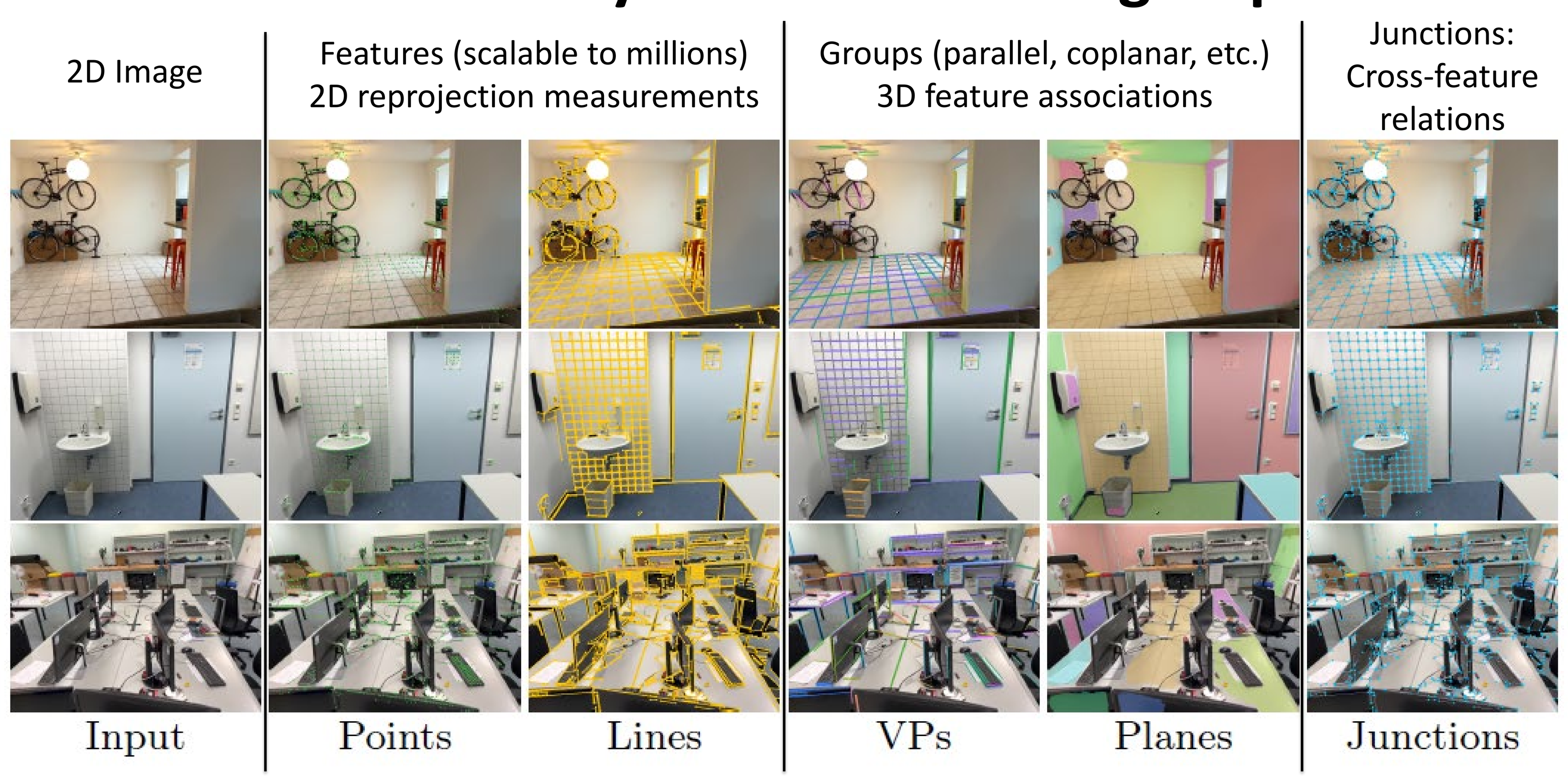
Our previous works on modeling geometric structures



LIMAP at CVPR'23 LIMAP-SFM at ECCV'24

Motive: Can we model structures while maintaining efficiency?

A taxonomy on features and groups

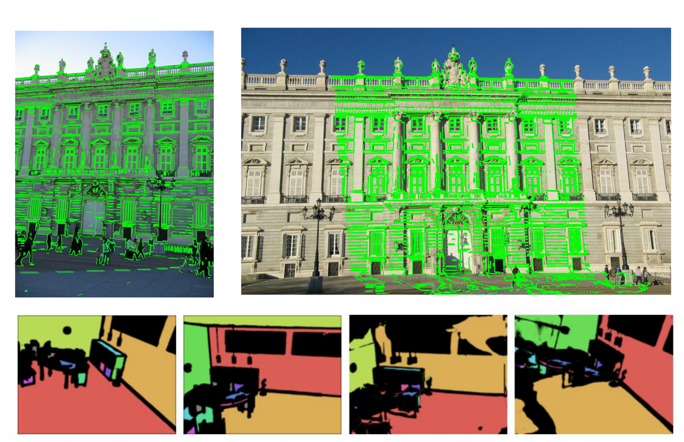


Some remarks:

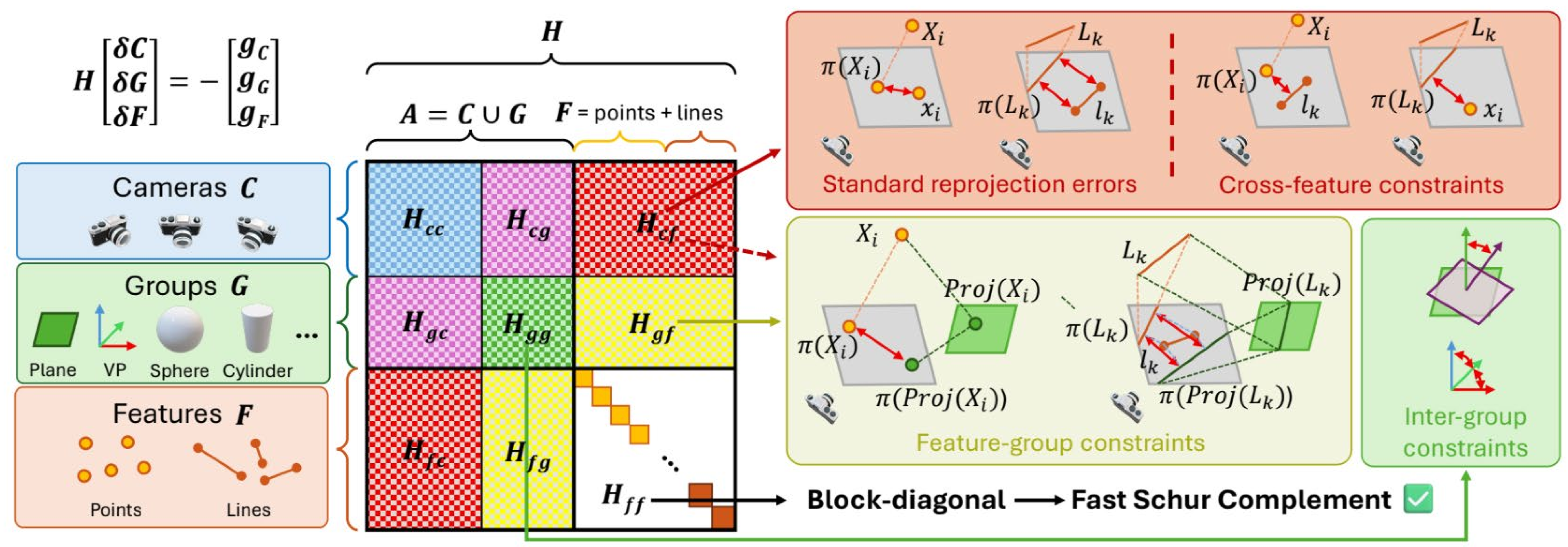
- The only scalable compact features are points and lines.
- Groups are general concepts modeling feature associations, which can be VPs, planes, spheres, cones, and even monocular depth or CAD priors.

Building the constraints

- Off-the-shelf modular point and line detection and matching.
- Custom plane detection module based on MoGe **(you are welcome to check poster #73 in the same session)**
- Structure association via dense matching



Holistic Bundle Adjustment: A unified framework for optimizing holistic 3D structure



Summary of key ideas:

- **Groups as auxiliary variables in the reduced system**
- **Wireframe constraints as cross-feature 2D errors**

→ Fast Schur Complement

- **Group constraints as pixel-wise reprojection errors:** Stable optimization
- **Natural inter-group constraints:** Manhattan World, Atlanta World, etc.

Brief Analysis of group reprojection error

$$r_g^{(p)} = \pi_p(C_i, \mathbf{X}_j) - \pi_p\left(C_i, \mathrm{Proj}_{G_t}(\mathbf{X}_j)\right) \in \mathbb{R}^2$$

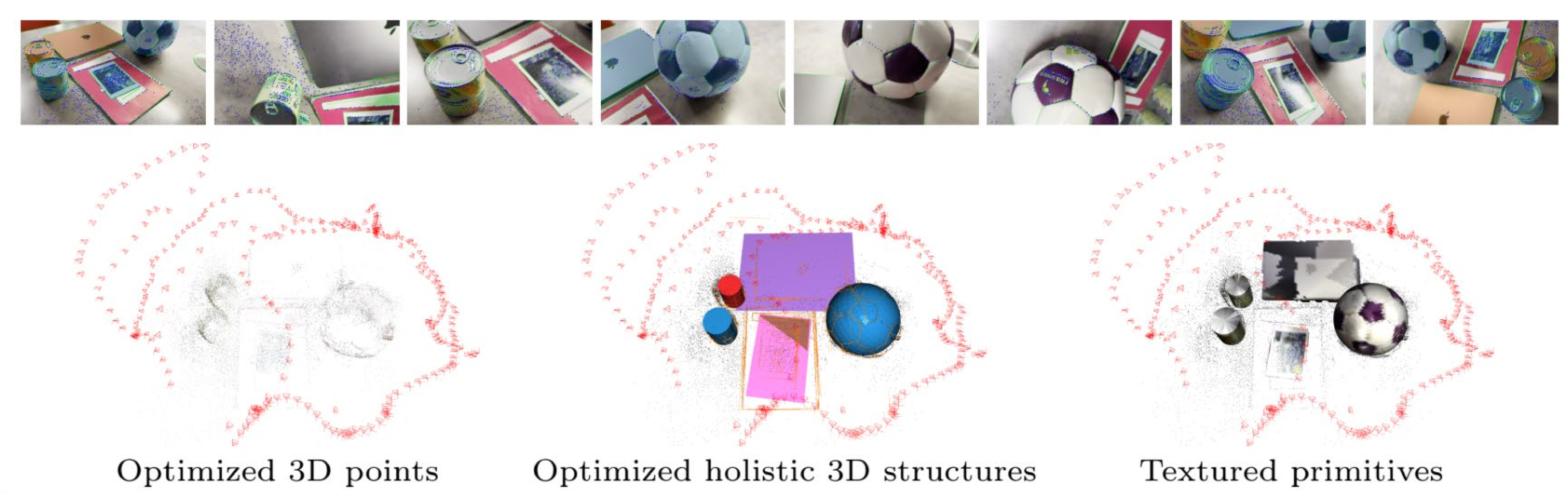
The projection difference of the original point and its feet in the group.

Across multiple views, this is equivalent to 3D error with its information calculated via the relinearized Jacobians as:

$$E_g \approx \sum_i \|J_i \delta\|^2 = \delta^\top \left(\sum_i J_i^\top J_i\right) \delta$$

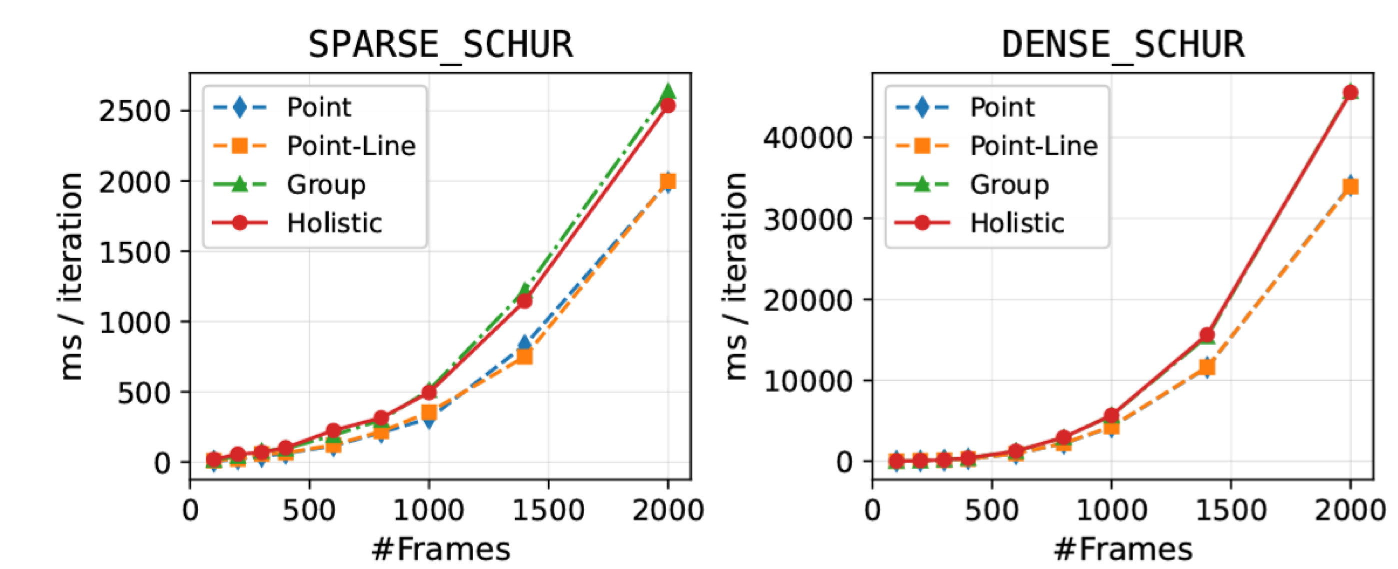
The cost is therefore statistically principled while sharing a same unit as the original reprojection error, leading to stable optimization and natural selection of termination criteria.

Some Experimental Results



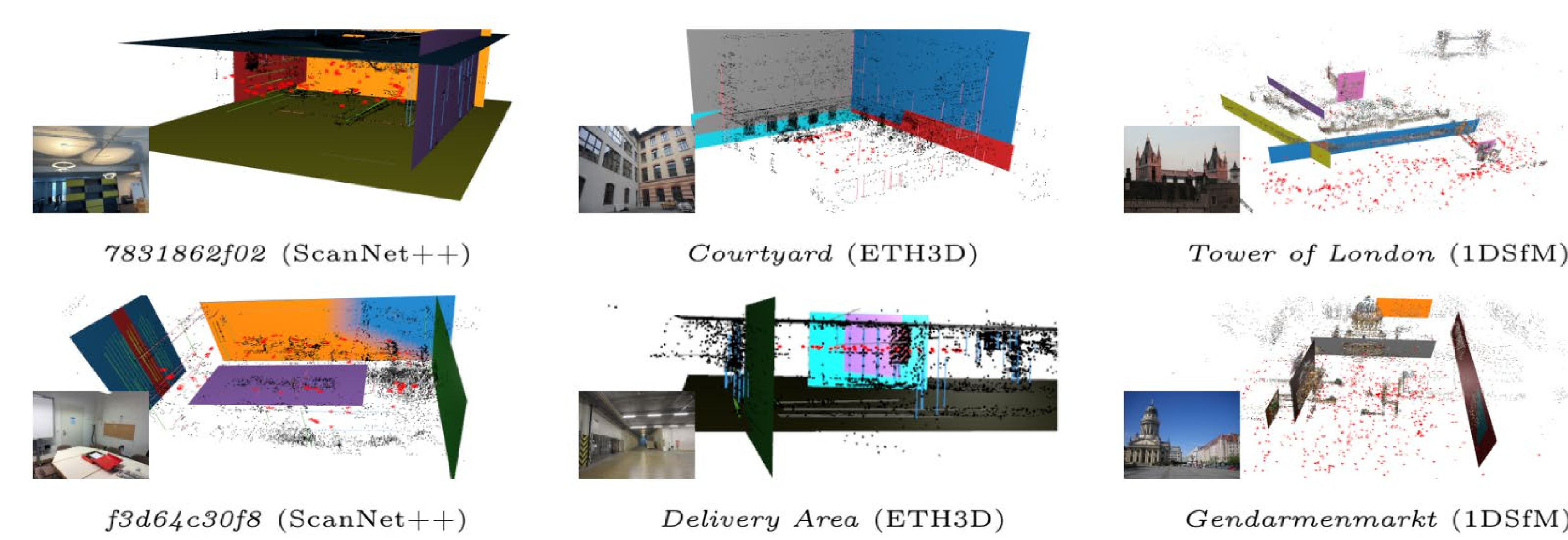
Optimized 3D points Optimized holistic 3D structures Textured primitives

Example holistic reconstructions on a daily smartphone video
SAM3 based primitive detection: football -> sphere, cans -> cylinder



Similar scaling complexity as classical bundle adjustment with only points

	Hypersim (8 scenes)			ScanNet++ (20 scenes)		
	AUC@3°	AUC@5°	AUC@10°	AUC@3°	AUC@5°	AUC@10°
COLMAP	89.3	90.4	92.2	84.0	86.0	87.5
Point-Line SfM	92.2	93.2	94.0	84.7	86.7	89.2
Holistic SfM (Ours)	**92.5**	**93.4**	**94.2**	**87.4**	**89.4**	**90.6**
	ETH3D (11 scenes)			**7Scenes (7 scenes)**		
	AUC@3°	AUC@5°	AUC@10°	AUC@3°	AUC@5°	AUC@10°
COLMAP	50.2	66.4	77.6	22.9	48.7	75.5
Point-Line SfM	49.4	67.1	79.6	23.1	48.9	76.8
Holistic SfM (Ours)	**50.4**	**68.1**	**80.4**	**24.3**	**50.8**	**78.4**



7831862f02 (ScanNet++) *Courtyard (ETH3D)* *Tower of London (1DSfM)*
f3d64c30f8 (ScanNet++) *Delivery Area (ETH3D)* *Gendarmenmarkt (1DSfM)*

Fully integrated in a principled SfM pipeline, compatible with the COLMAP ecosystem.
Therefore naturally benefits from new features such as multi-camera rig support and recent advances in two-view geometry.