

# Establishing the Gold Standard for 360° Visual-Inertial Reconstruction

Gilbert Tanner<sup>1</sup>, Giorgos Evangelou<sup>1</sup>, Julian Lechner<sup>2</sup>

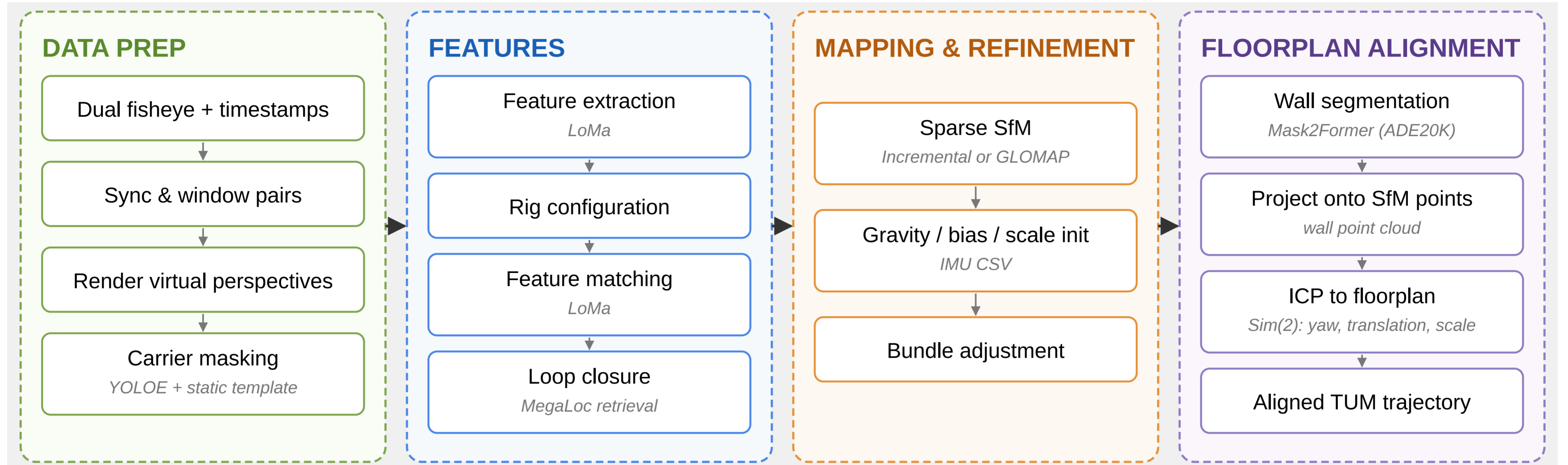
Supervisors: Zador Pataki<sup>1</sup>, Xudong Jiang<sup>1</sup>, Paul-Edouard Sarlin<sup>3</sup>, Shaohui Liu<sup>1</sup>

<sup>1</sup>ETH Zurich <sup>2</sup>AAU Klagenfurt & DLR <sup>3</sup>Google

## 1. Introduction

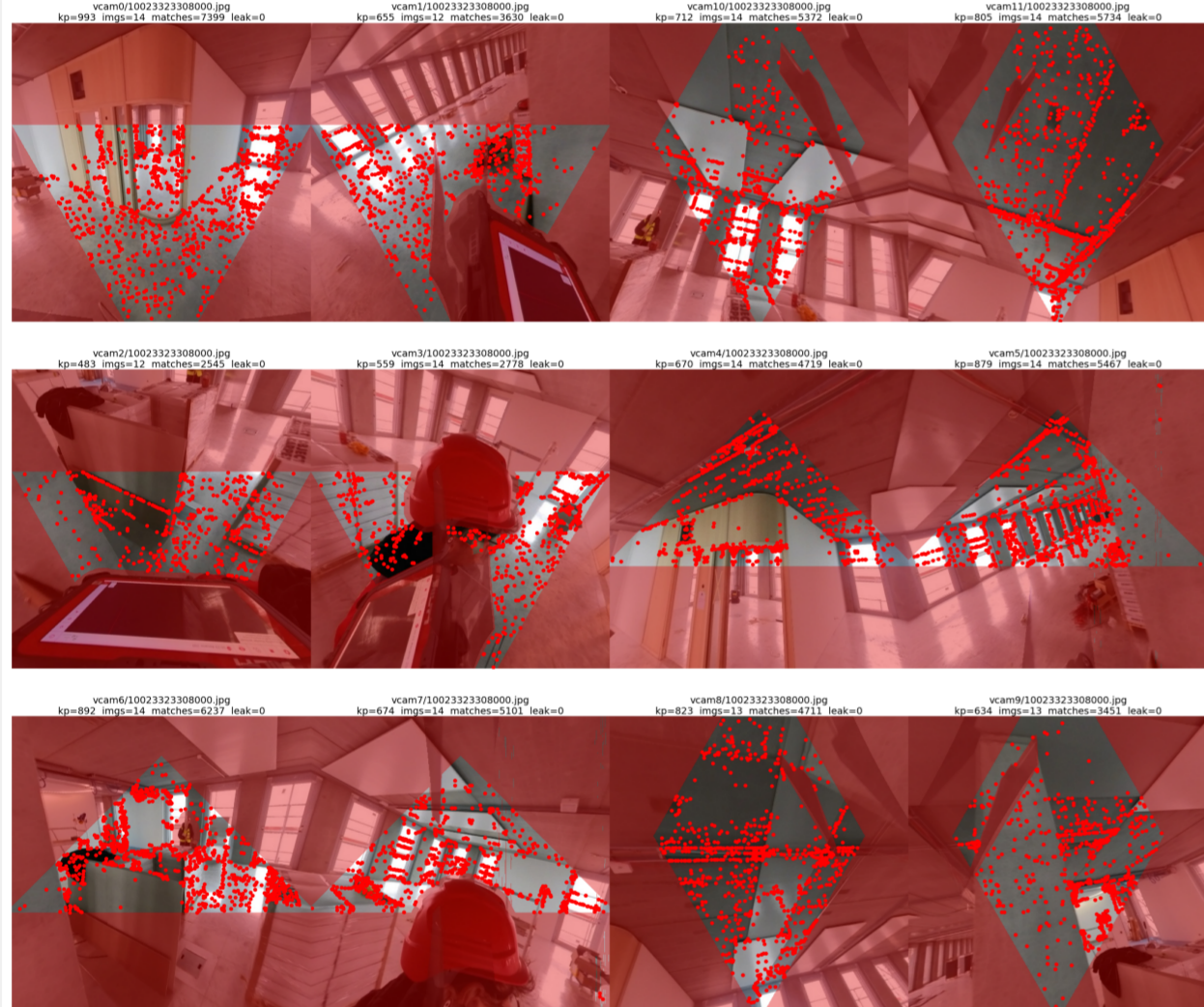
**Task.** Metric, gravity-aligned, floorplan-consistent trajectory and 3D map on the Hilti × Trimble SLAM Challenge 2026 indoor dual-fisheye + IMU dataset. **Problem.** OpenVINS [5] drifts on long indoor traverses with repetitive textures. **Our solution** combines two pipelines on top of the VIO trajectory — a lightweight COLMAP BA refinement and a full panorama SfM from scratch with tightly-coupled VI bundle adjustment — plus a unified 2D ICP post-processing step that aligns both outputs to the building floorplan.

## 2. Pipeline Overview

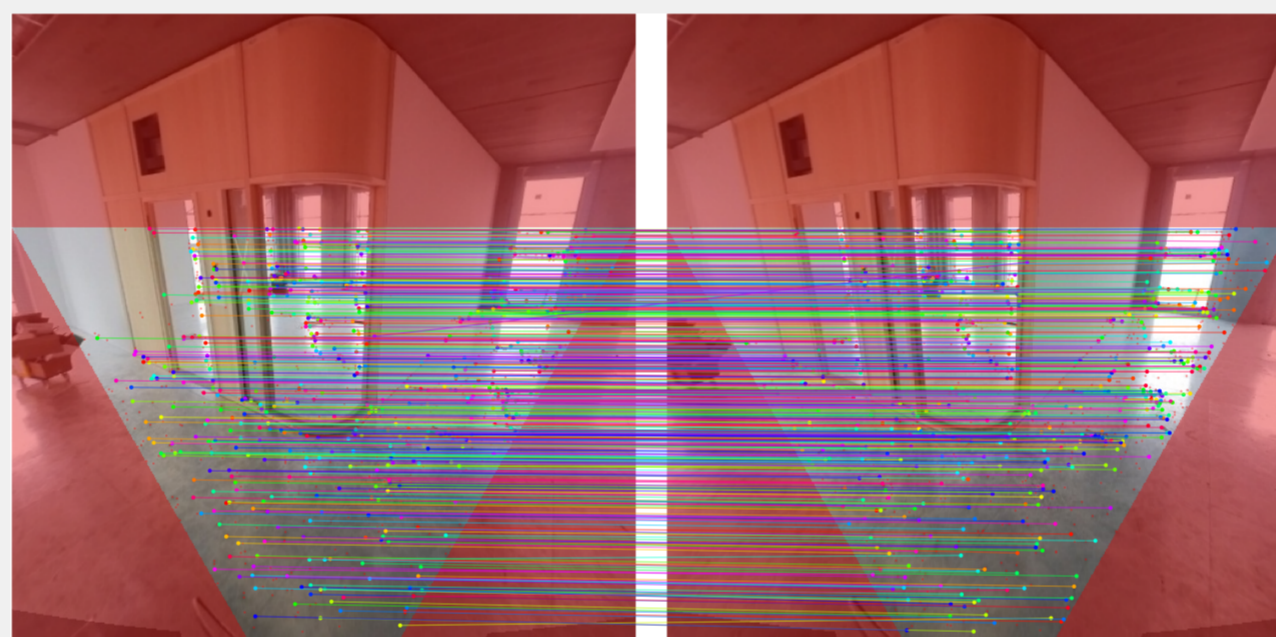


## 3. Method Steps

**Virtual perspective views.** Each fisheye pair is remapped into a 4×3 rig of 90° perspective cameras with a 1/θ blend and Voronoi seams.



**LoMa feature matching.** LoMa [7] delivers dense correspondences on the dark, repetitive indoor frames that classical matchers struggle with.



**Floorplan alignment.** Mask2Former [3] on the equirectangular pair isolates wall pixels (top); the wall vote is back-projected to the SfM cloud and 2D ICP [2] aligns it to the floorplan wall mask (bottom).



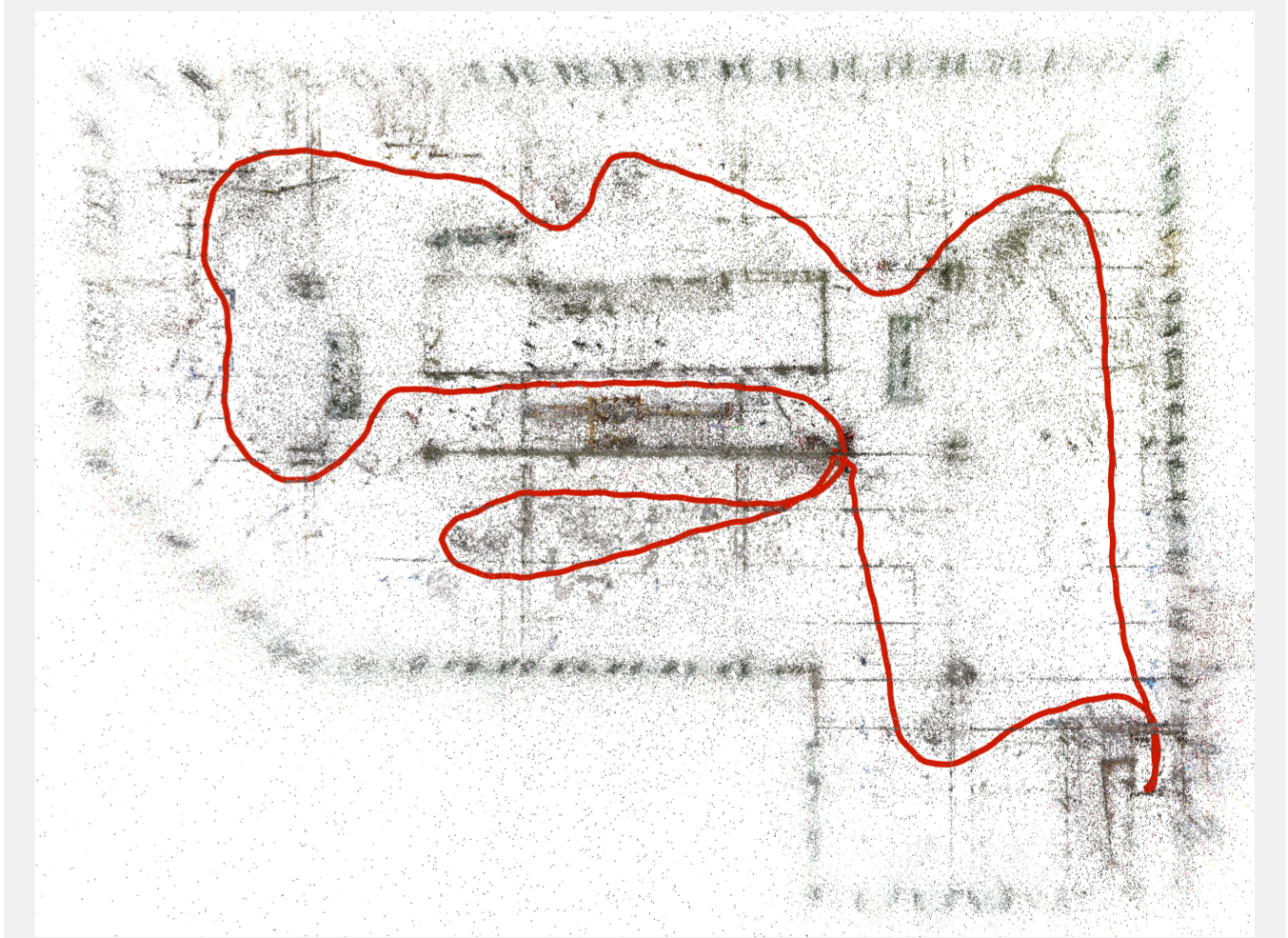
**Loop closure via MegaLoc.** MegaLoc [1] descriptors averaged over the twelve virtual cameras yield heading-invariant frame descriptors, catching start/end revisits.



**Carrier masking.** YOLOE [8] segments the operator + helmet + tablet; the dilated mask (yellow) is unioned with a static carrier template (green).



**Final aligned reconstruction.** The full SfM point cloud (gray) fills the building outline, with the recovered trajectory (red) tracing the operator's path.



## 4. Takeaways

- 360° rig + LoMa + MegaLoc provide matches and loops that classical features miss on construction-site textures.
- Tightly-coupled VI BA (closed-form init [6] + preintegration [4]) recovers metric scale and gravity.
- Wall segmentation + 2D ICP closes the residual yaw / translation gap to the floorplan.

## 5. References

- [1] Gabriele Berton, Gabriele Trivigno, Barbara Caputo, and Carlo Masone. MegaLoc: One retrieval to place them all, 2025. arXiv:2502.17237.
- [2] Paul J. Besl and Neil D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [3] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [4] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2017.
- [5] Patrick Geneva, Kevin Eickenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. OpenVINS: A research platform for visual-inertial estimation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [6] Raúl Mur-Artal and Juan D. Tardós. Visual-inertial monocular SLAM with map reuse. *IEEE Robotics and Automation Letters*, 2(2):796–803, 2017.
- [7] David Nordström, Johan Edstedt, Georg Bökman, Jonathan Astermark, Anders Heyden, Viktor Larsson, Märten Wadenbäck, Michael Felsberg, and Fredrik Kahl. LoMa: Local feature matching revisited, 2026. arXiv:2604.04931.
- [8] Ao Wang, Lihao Liu, Hui Chen, Zijia Lin, Jungong Han, and Guiguang Ding. YOLOE: Real-time seeing anything. In *arXiv:2503.07465*, 2025.