# 3D Modeling on the Go: Interactive 3D Reconstruction of Large-Scale Scenes on Mobile Devices

Thomas Schöps, Torsten Sattler, Christian Häne, Marc Pollefeys

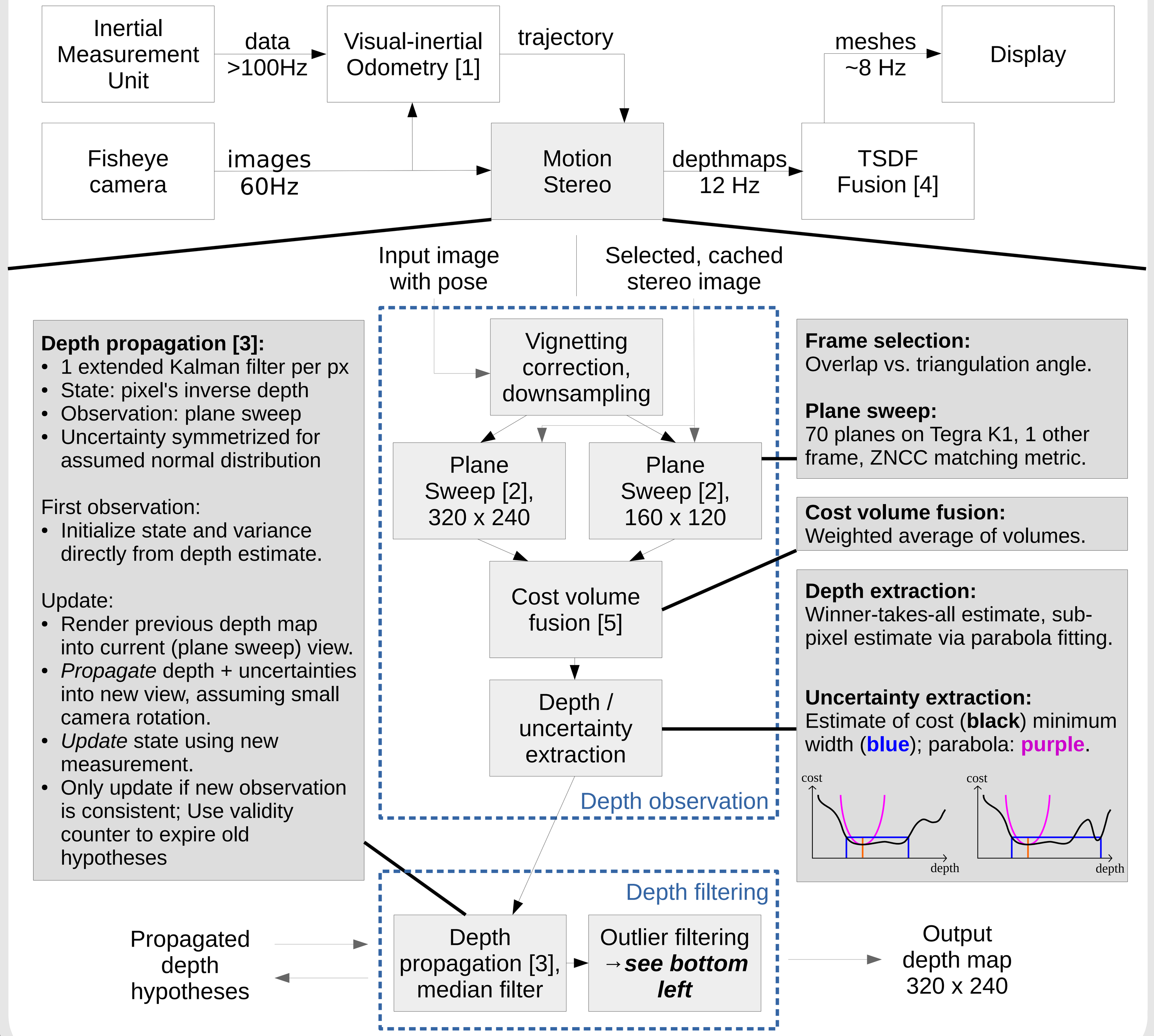
## Goals & key insights

- **Large-scale dense 3D reconstruction in real-time on mobile devices (Google Project Tango tablets)**
  - Suitable for AR applications
  - Good quality preview for offline reconstruction
- **Input: Monocular camera, IMU**
  - Unlimited reconstruction range
  - Can operate in sunlight

**Key insights:**

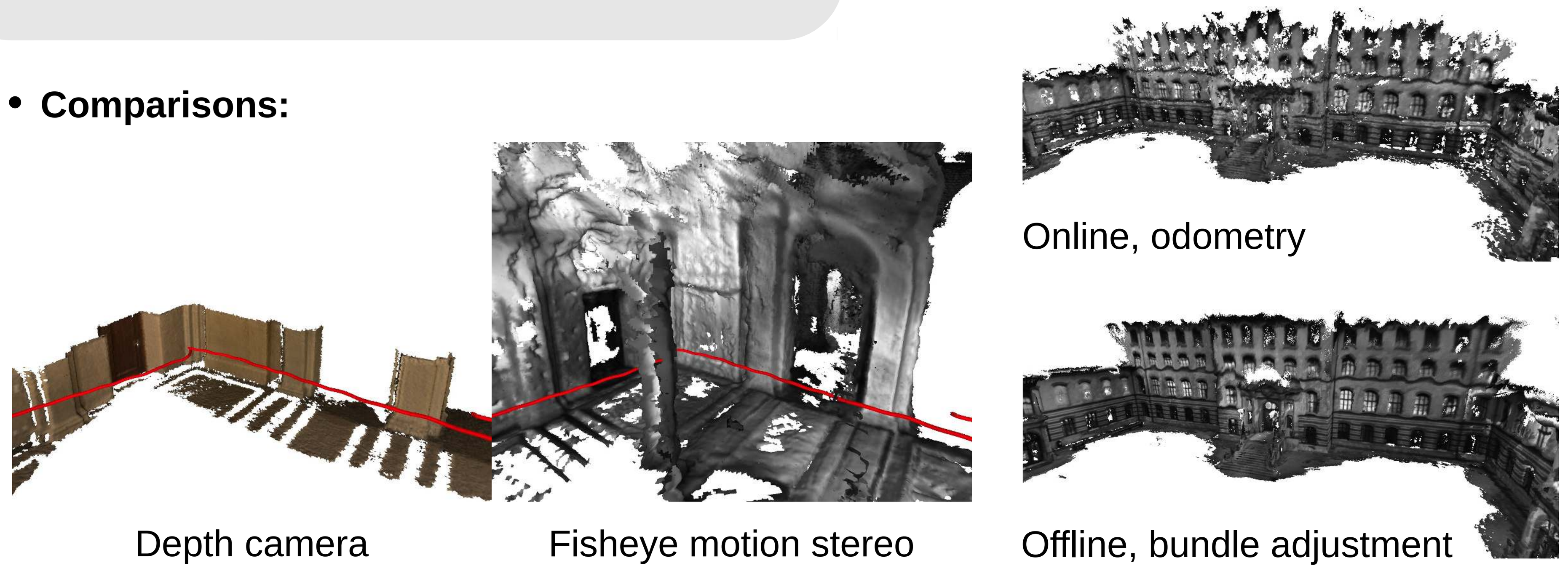
- Motion stereo based, large scale, dense 3D reconstruction is feasible on mobile devices with interactive frame rates.
  → ***See our processing pipeline and results achieved in real-time.***
- In contrast to object reconstruction, large-scale reconstruction with passive sensing needs strong filtering.
- Aggressive filtering is preferable to complete but noisy depth maps for interactive systems.
  → ***See our filtering pipeline.***

## Processing pipeline

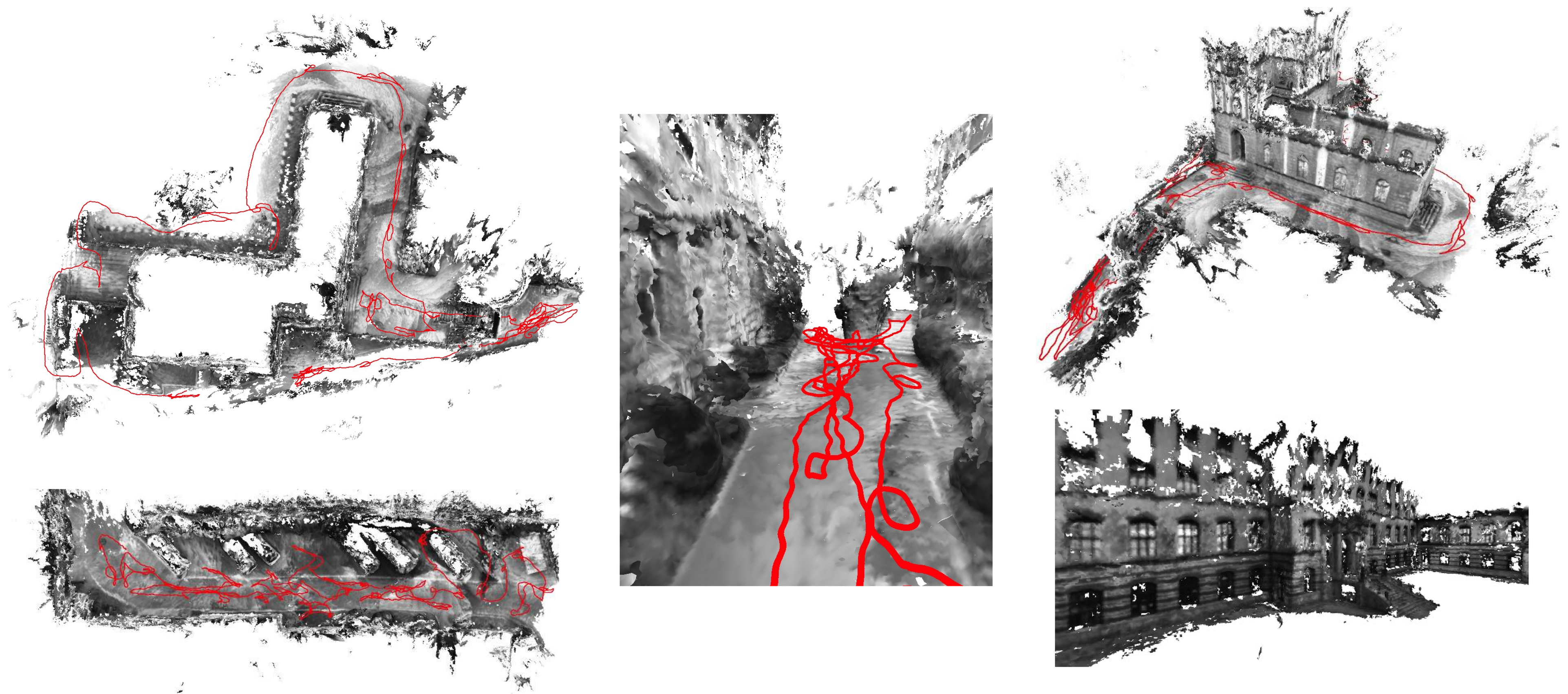



## Results

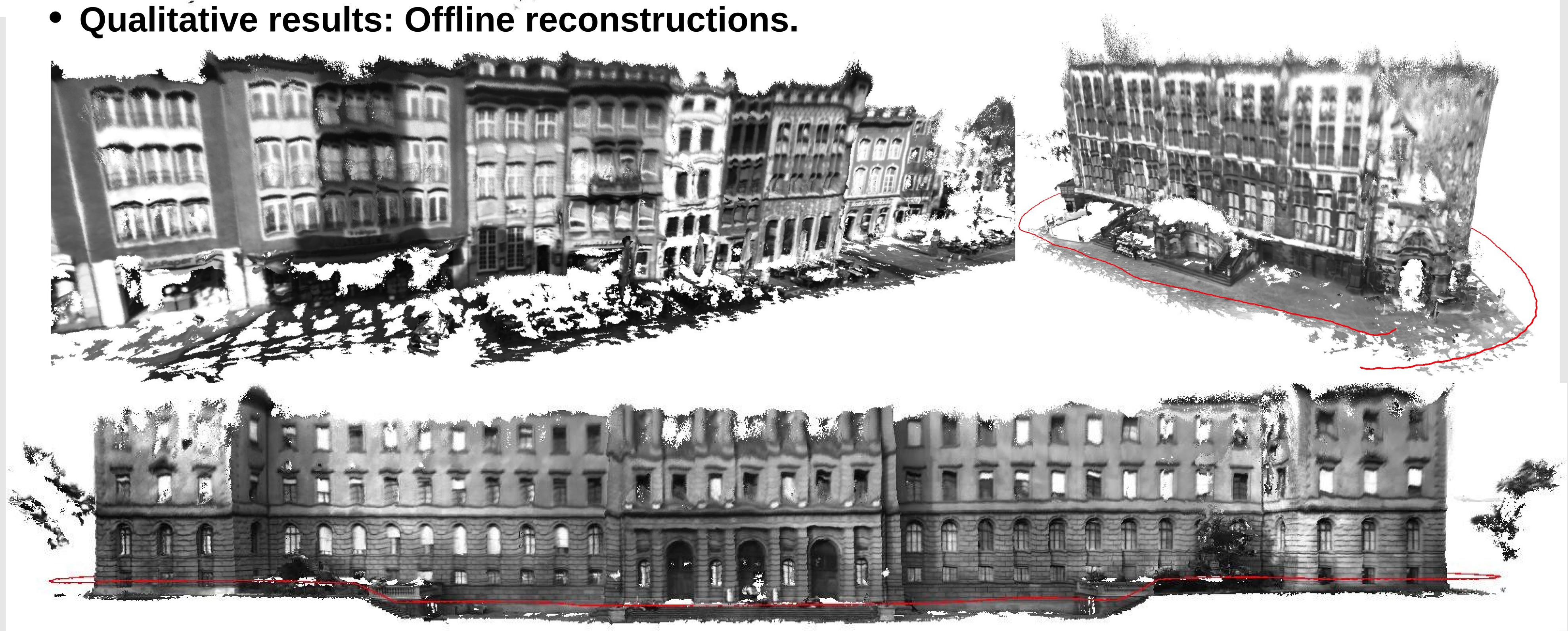
- **Comparisons:**




- **Qualitative results: Real-time reconstructions.**



- **Qualitative results: Offline reconstructions.**



- **Quantitative results: see our paper.**

## Filtering steps

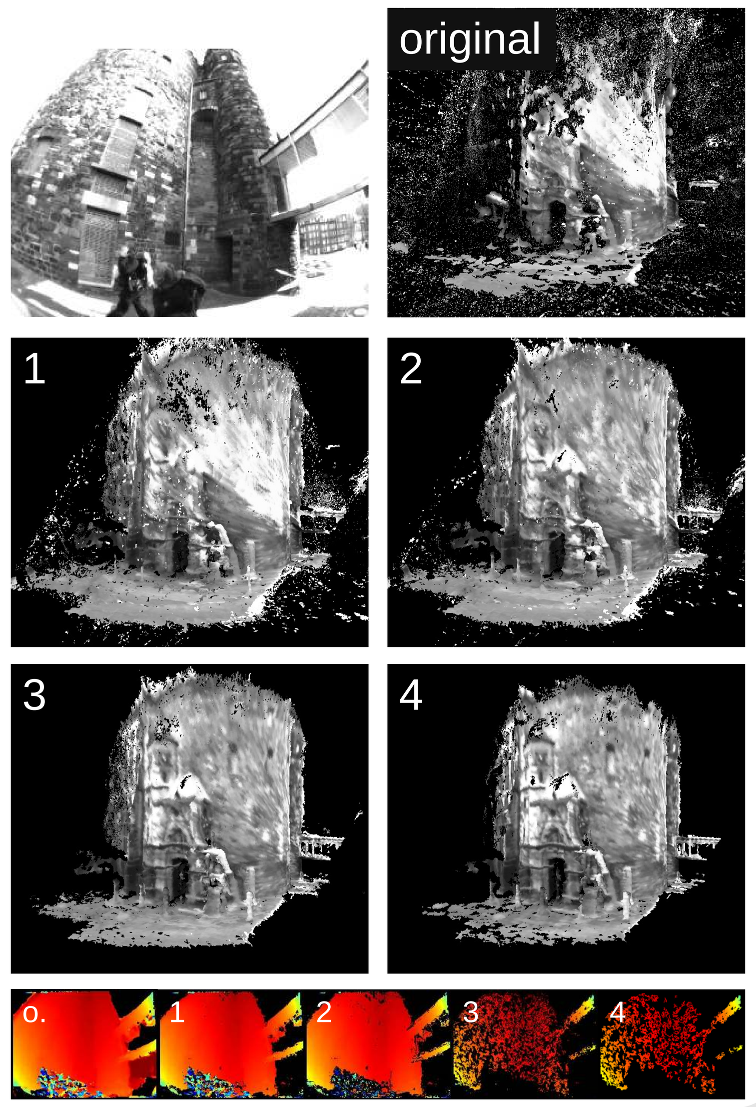
**Improving accuracy: Kalman filtering [3]**
Per-pixel ext. Kalman filters for inverse depth. Integrate only points that are stable.

**Spatial regularization: 3x3 median filtering**
Suppress outlier noise, smooth the depth map.

**Outlier suppression**

1. Variance thresholding:
   Discard pixels with high estimated variance.
2. Angle thresholding:
   Discard surfaces with slanted angles.
3. Consistency over time:
   Current depth map is compared to (warped) previous one 0.25 seconds ago.
   Keep consistent estimates only.
4. Connected component analysis:
   Discard small connected components.

**Outlier suppression visualization**




## Timings

**Tango Tablet**
- Nvidia Tegra K1 (quad core) chipset
- 70 planes, 7.5cm voxel resolution

**Desktop PC**
- Nvidia GeForce 780 GTX, Intel Core i7-4770K
- 200 planes, 4cm voxel resolution

| | PC | Tango Tablet |
|---|---|---|
| **Addition of a reference frame** | $1.5 \pm 0.2$ | $6.2 \pm 3.3$ |
| Vignetting correction | $1.0 \pm 0.1$ | $1.2 \pm 0.7$ |
| Downsampling, transfer to GPU | $0.5 \pm 0.2$ | $4.9 \pm 3.0$ |
| **Depth estimation** | $8.1 \pm 0.9$ | $83.0 \pm 7.6$ |
| Plane sweep ($320 \times 240$) | $3.9 \pm 0.6$ | $35.2 \pm 2.1$ |
| Plane sweep ($160 \times 120$) | $1.3 \pm 0.2$ | $13.4 \pm 2.6$ |
| Multires. fusion, depth extraction | $1.6 \pm 0.4$ | $15.2 \pm 2.6$ |
| Depth propagation | $1.3 \pm 1.7$ | $10.3 \pm 5.8$ |
| Filtering, transfer to CPU | $1.3 \pm 0.4$ | $13.5 \pm 5.9$ |
| **TSDF integration, remeshing** (CPU) | $30.2 \pm 23.7$ | $122.1 \pm 74.6$ |

Timings in milliseconds. Components run in parallel.
→ **Depth maps estimated with 12 FPS on Tango tablet.**

## References

[1] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis. Camera-IMU-based localization: Observability analysis and consistency improvement. *The International Journal of Robotics Research*, 2013.
[2] C. Häne, L. Heng, G. H. Lee, A. Sizov, and M. Pollefeys. Real-time direct dense matching on fisheye images using plane-sweeping stereo. In *3DV*, 2014.
[3] J. Engel, J. Sturm, and D. Cremers. Semi-dense visual odometry for a monocular camera. In *ICCV*, 2013.
[4] M. Klingensmith, I. Dryanovski, S. Srinivasa, and J. Xiao. Chisel: Real time large scale 3D reconstruction onboard a mobile device. In *RSS*, 2015.
[5] K. Zhang, Y. Fang, D. Min, L. Sun, S. Yang, S. Yan, and Q. Tian. Cross-scale cost aggregation for stereo matching. In *CVPR*, 2014.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
Computer Vision and Geometry Lab

The research leading to these results has received funding from Google's Project Tango, from the Swiss National Science Foundation under the project Nr. 156973, and from Qualcomm.