Real-Time Dense Mapping for Self-Driving Vehicles using Fisheye Cameras

Zhaopeng Cui[1], Lionel Heng[2], Ye Chuan Yeo[2], Andreas Geiger[3], Marc Pollefeys[1,4], Torsten Sattler[5]

[1]Department of Computer Science, ETH Zürich [2]DSO National Laboratories
[3]MPI-IS and University of Tübingen [4]Microsoft, Switzerland
[5]Chalmers University of Technology

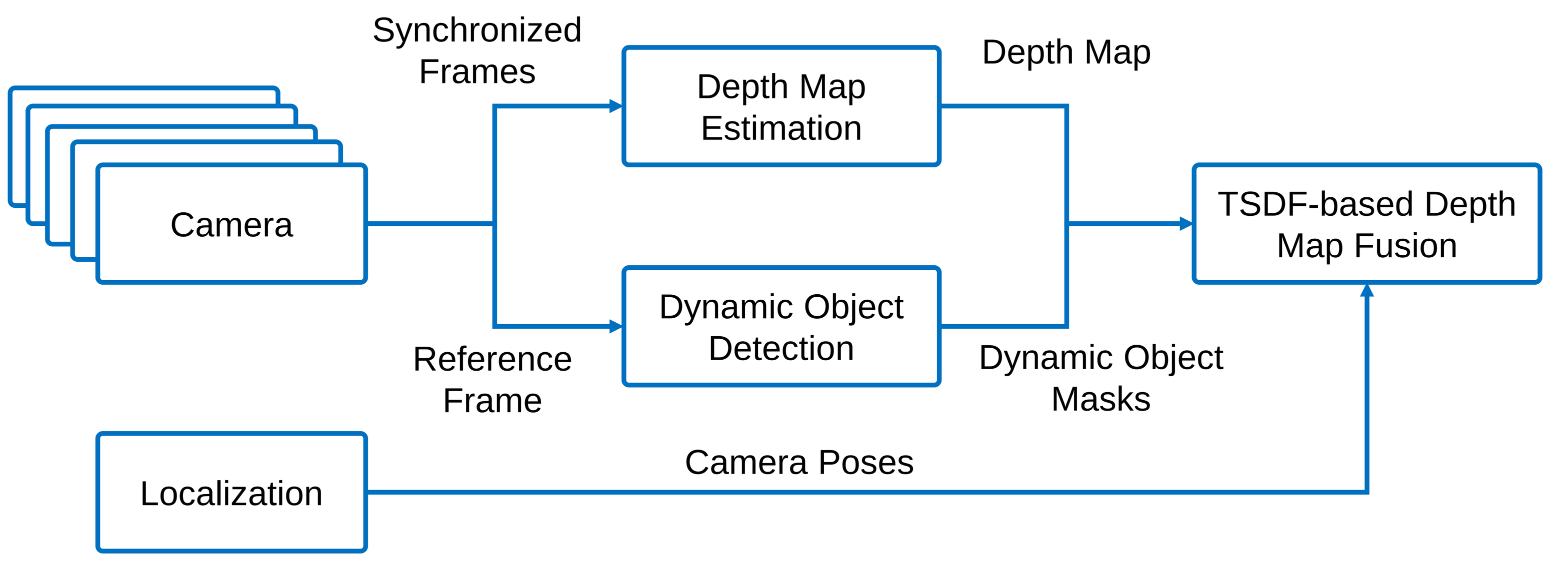
Motivation

- Real-time 3D mapping required to perceive and thus navigate in complex environments.
- Higher-resolution 3D maps obtained from images compared to LiDAR.
- More well-developed scene understanding techniques for images than those for LiDAR point clouds.

Contributions

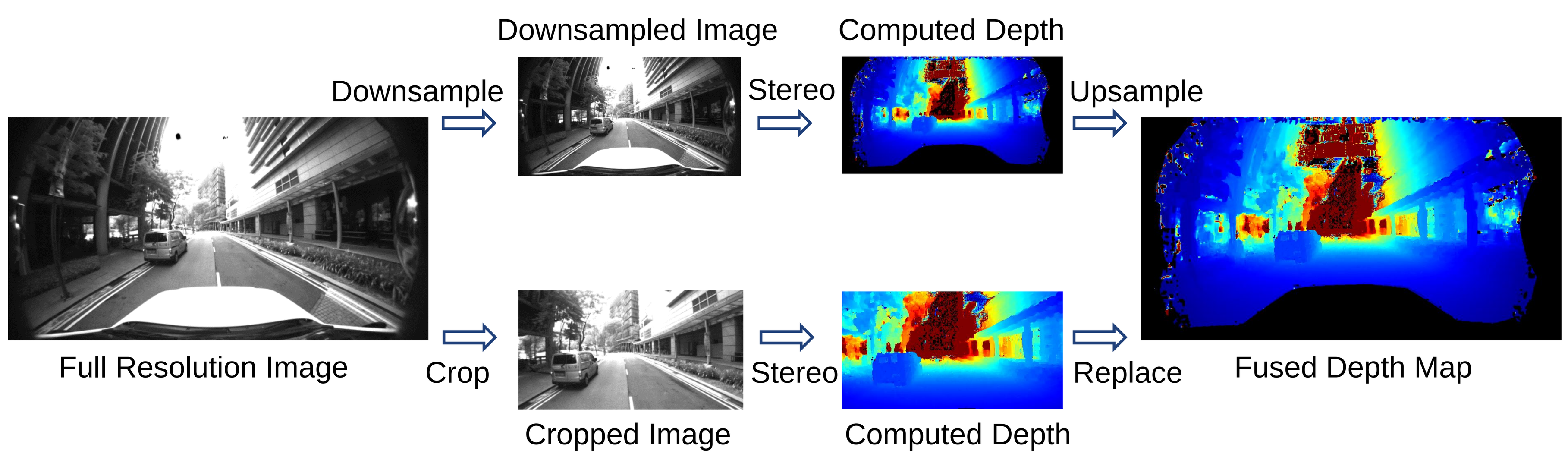
- A practical system for real-time dense mapping purely using fisheye cameras.
- A new multi-scale strategy for fisheye depth map estimation to maintain both accuracy and efficiency.
- Evaluation of multiple depth filtering and local map pruning techniques with LiDAR data.

System Overview



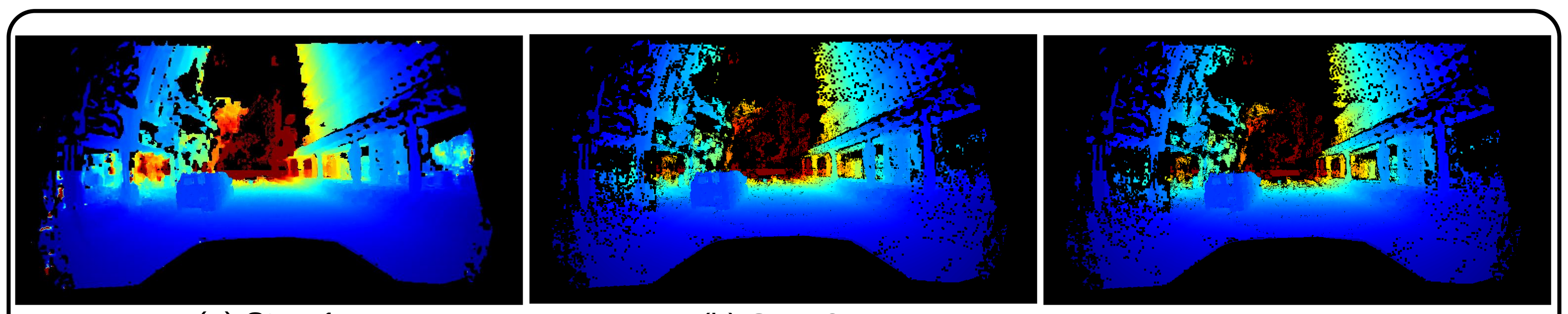
- Step 1: Compute the depth map given multiple images captured at the same time.
- Step 2: Detect dynamic objects using YOLOv3 [1] with the finetuned model.
- Step 3: Integrate depth maps over time into a truncated signed distance function volume using camera poses provided by a localization system.

Depth Map Estimation at Multiple Scales



- Two scales: 1) the downsampled fisheye image; 2) the cropped central area of the fisheye image (close-to-pinhole camera).
- Use the plane-sweeping stereo algorithm [2] for the depth map estimation.
- Reduce the running time by about 28% compared to processing original images.

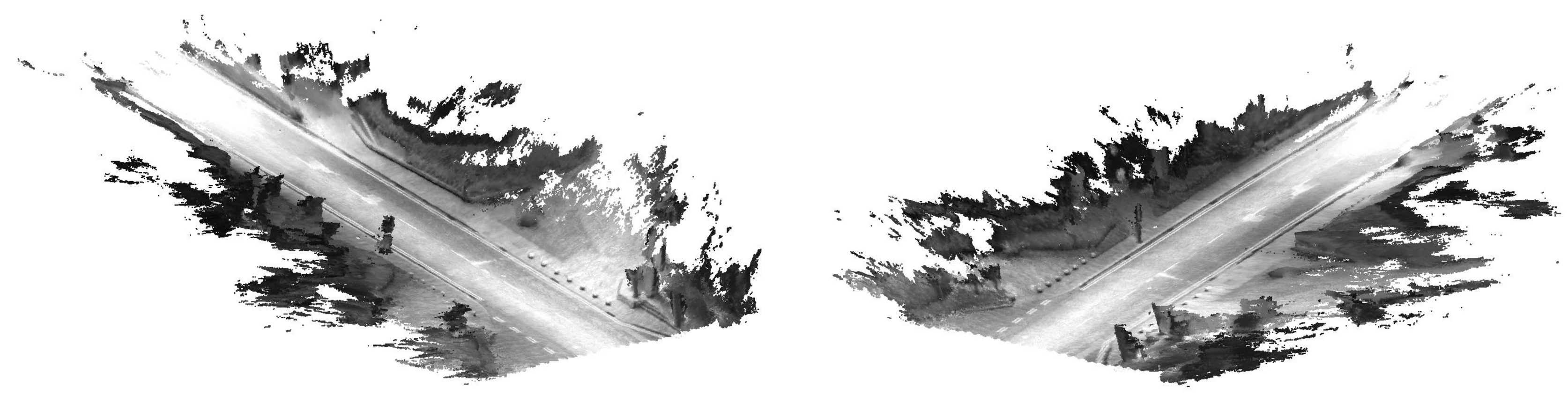
Depth Map Filtering



(a) Step 1 (b) Step 2 (c) Step 3

- Filter Step 1: Filter with the matching cost value of the best depth candidate for a pixel.
- Filter Step 2: Filter with the ratio between the first and second best cost values.
- Filter Step 3: Filter with the local depth continuity checking.

TSDF-based Depth Map Fusion



- Adapt the depth map fusion pipeline [3] to support the fisheye camera model.
- Maintain a local map with a size of $60\text{m} \times 60\text{m} \times 3\text{m}$ centered at the current vehicle position for online mapping.
- Consider voxel blocks with at least 3 observations only.

Experimental Evaluation

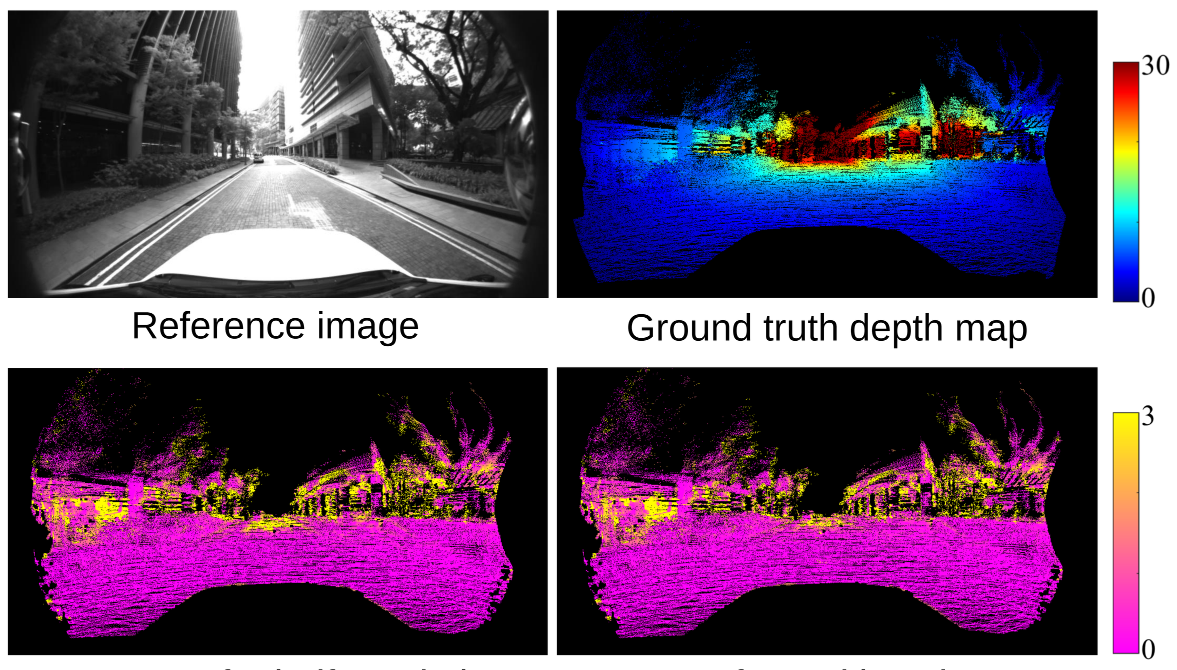
Experimental Setup

AutoVision [4] vehicle platform.

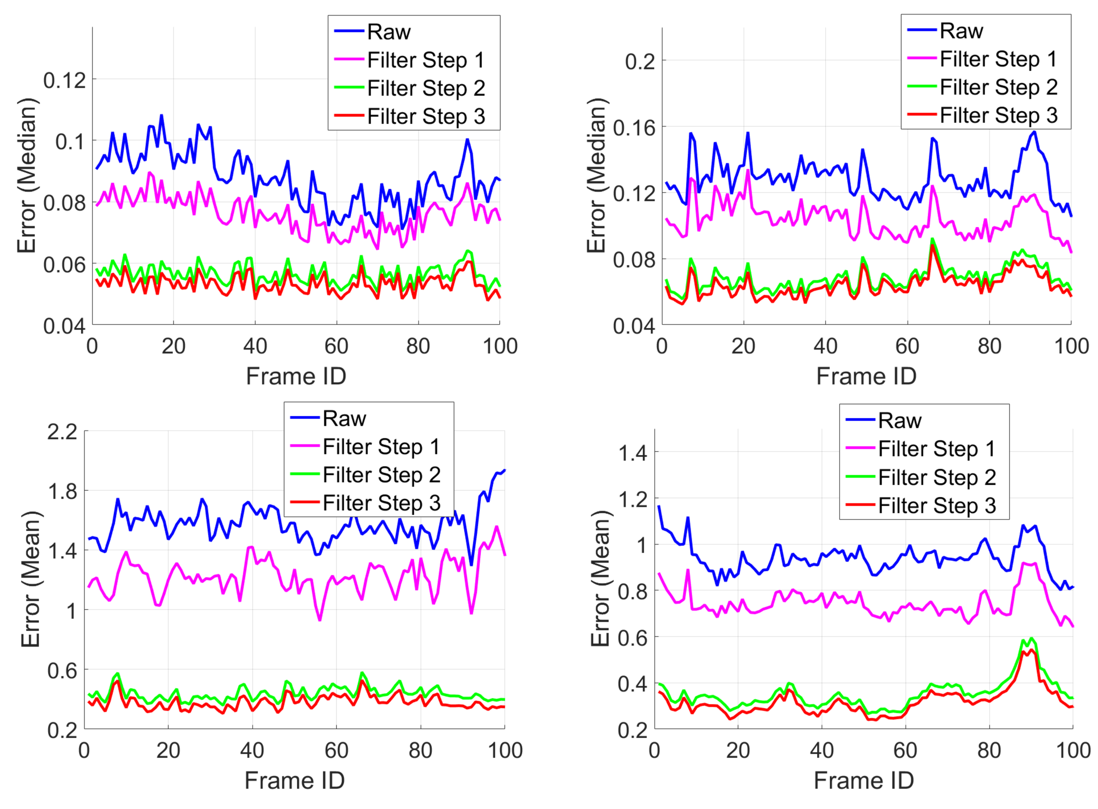
Component	Runtime
Depth map estimation (5 cameras)	60 ms
Dynamic object detection	40 ms
Depth fusion	20 ms

Average runtime.

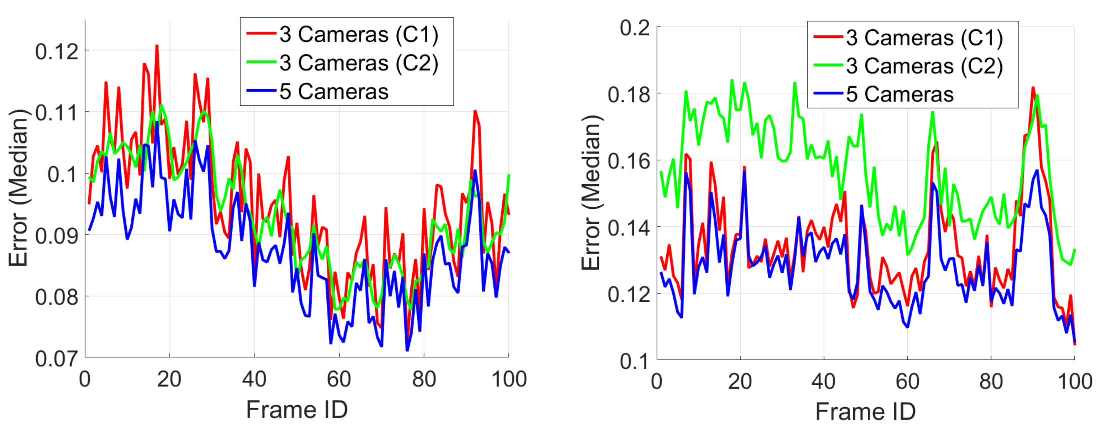
Evaluation of the Depth Estimation Stage



Reference image | Ground truth depth map
Error map for half resolution | Error map for multi-scale strategy

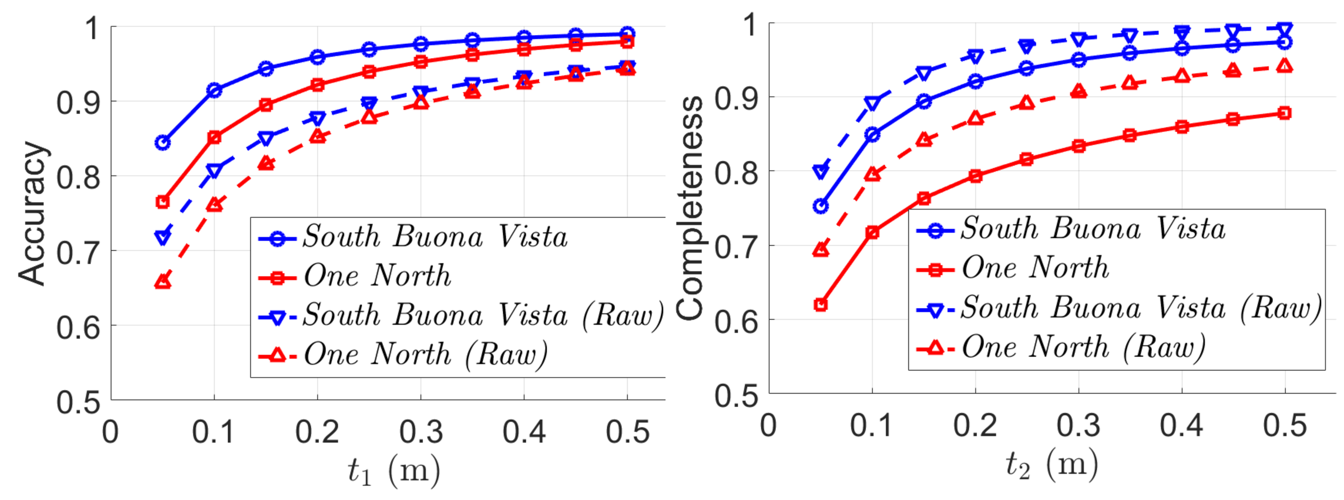


Impact of the filtering stages on the depth error (in meter).



Performance with different camera configurations.

Evaluation of the 3D Mapping Stage



Evaluation of mapping results with and without unreliable voxels removal.

Evaluation of Object Detection

Recovered 3D points (left) without and (right) with moving object detection.

References

[1] J. Redmon and A. Farhadi. YOLOv3: An incremental improvement. *CoRR*, abs/1804.02767, 2018.
[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 *International Conference on 3D Vision (3DV)*, 2014.
[3] O. Kähler, V. A. Prisacariu, C. Y. Ren, X. Sun, P. Torr, and D. Murray. Very high frame rate volumetric integration of depth images on mobile devices. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 21(11):1241–1250, 2015.
[4] L. Heng, B. Choi, Z. Cui, M. Geppert, S. Hu, B. Kuan, P. Liu, R. Nguyen, Y. C. Yeo, A. Geiger, G. H. Lee, M. Pollefeys, and T. Sattler. Project autovision: Localization and 3d scene perception for an autonomous vehicle with a multi-camera system. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.