Robust Dense Mapping for Large-Scale Dynamic Environments

[1]ETH Zurich, Switzerland, [2]University of Toronto, Canada, [3]Uber ATG, Toronto, Canada
[4]Microsoft, Redmond, USA, [5]MPI IS, Tübingen, Germany
Ioan Andrei Bârsan[1,2,3], Peidong Liu[1], Marc Pollefeys[1,4], Andreas Geiger[1,5]

ETH zürich
CVG Computer Vision and Geometry Lab
Max Planck Institute for Intelligent Systems

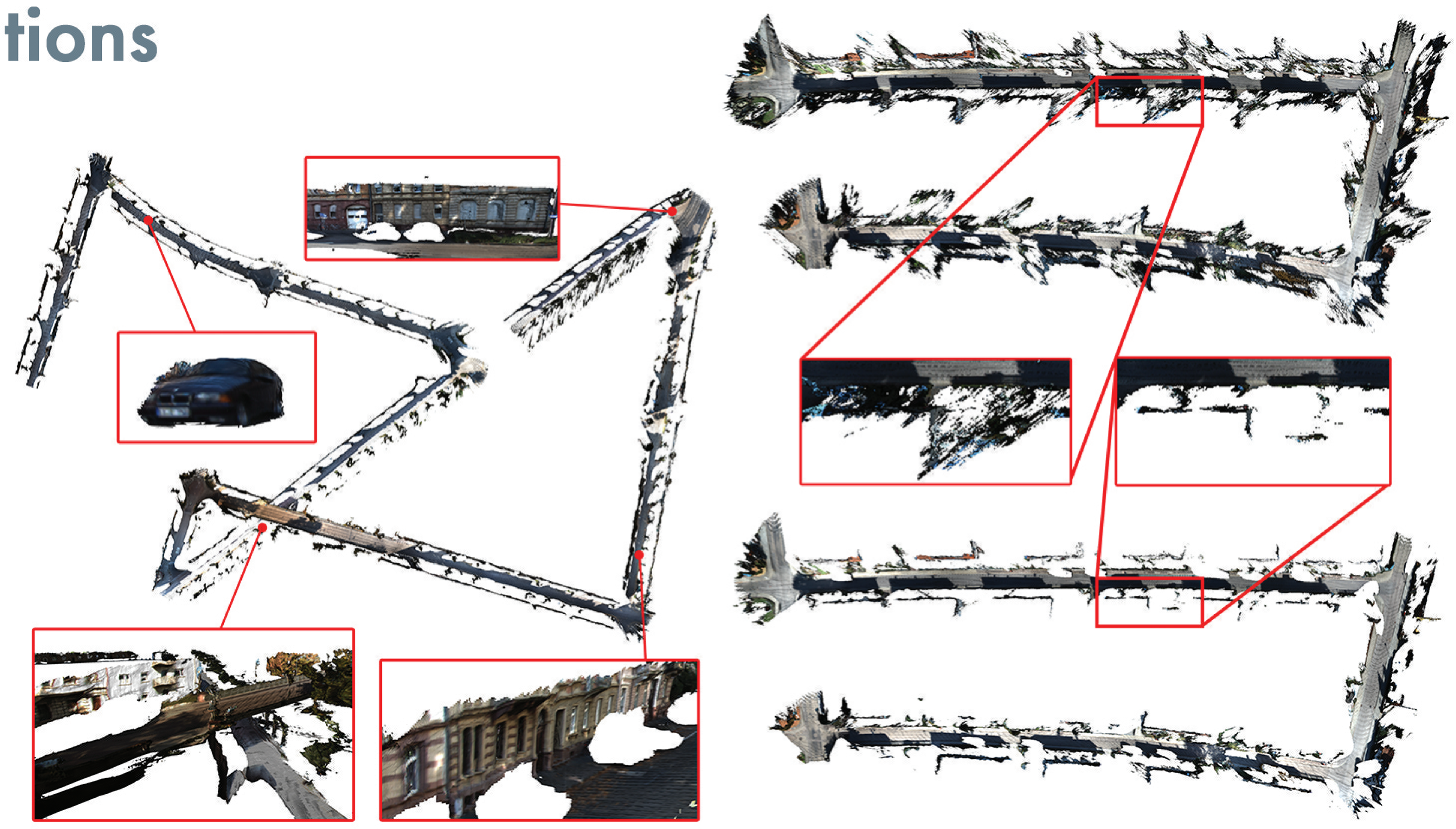
1. Problem and Contributions

Problem

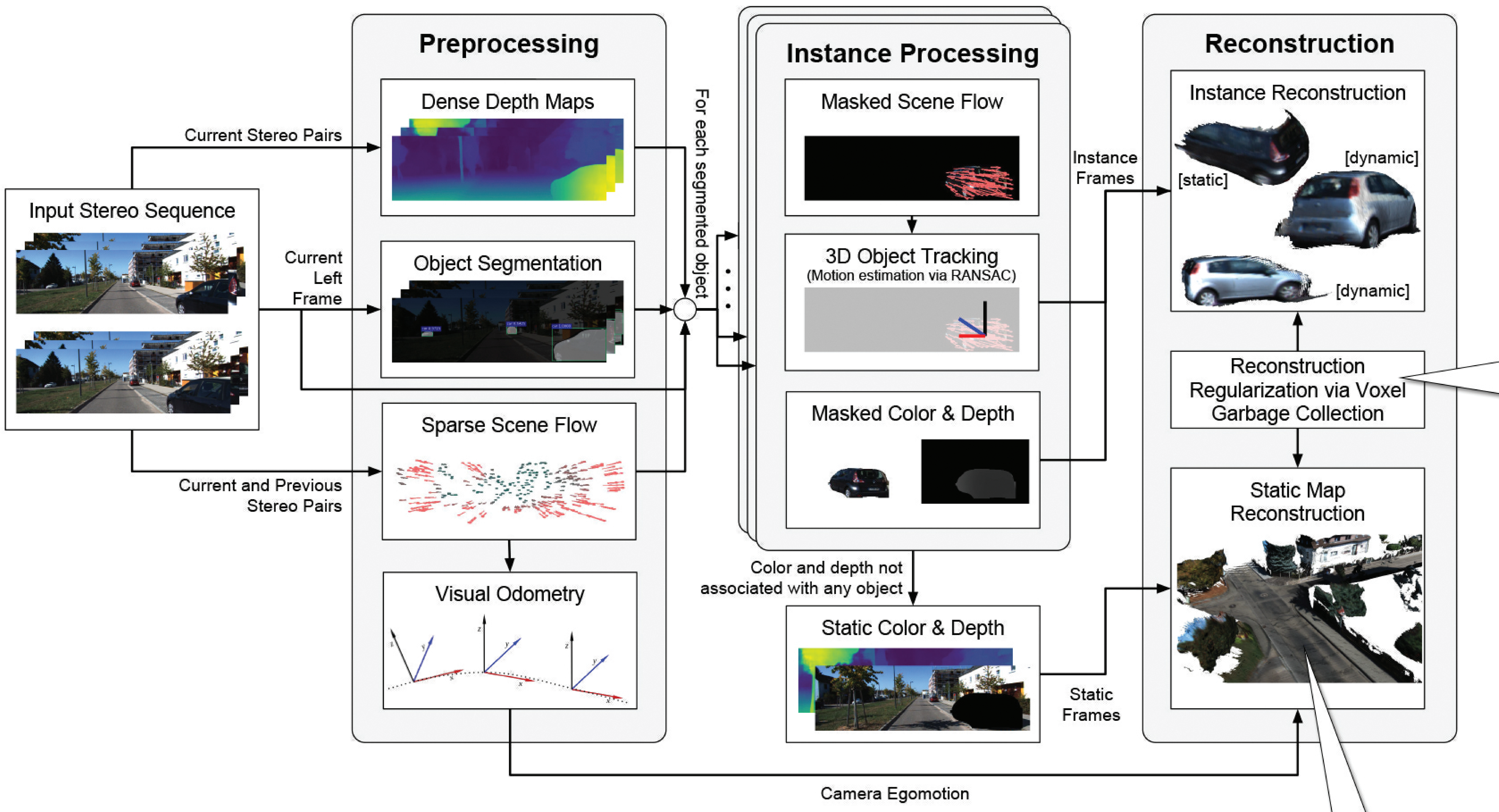
- Most mapping systems assume the environment is static or treat dynamic entities as noise.
- Dense mapping typically requires large amounts of GPU memory.

Contributions

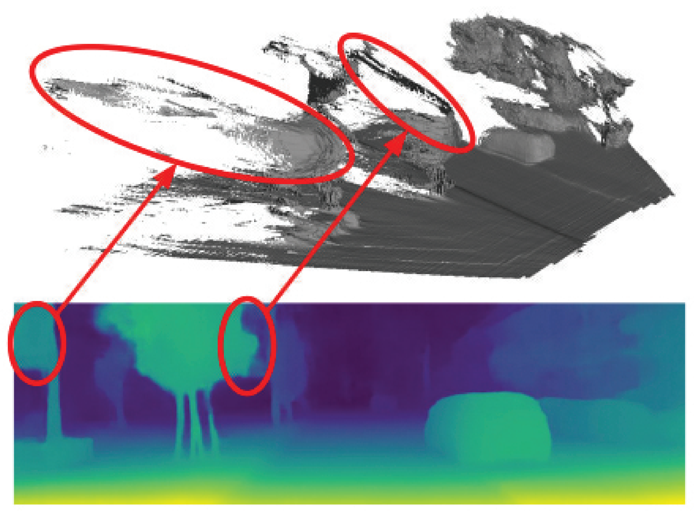
- Online dense mapping system that reconstructs:
  - Environment map
  - Moving objects
  - Potentially moving objects
- Performs low-overhead map pruning to significantly reduce memory footprint.
- Scales to large environments, such as entire neighborhoods.



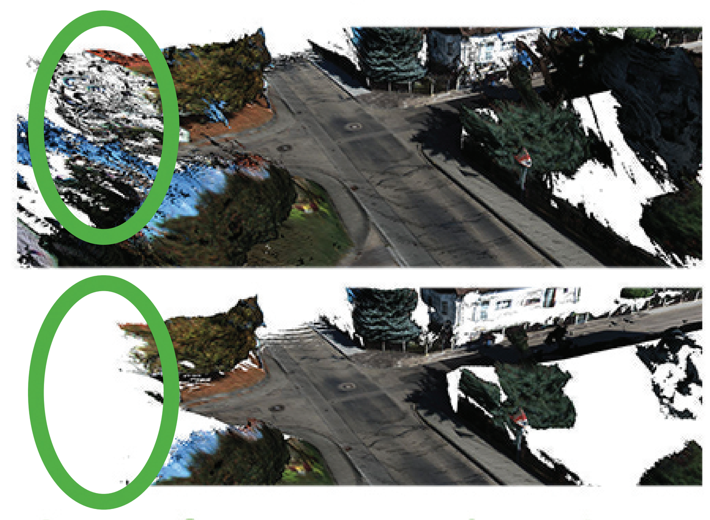
2. Method



Saving Memory: Map Pruning



Problem: 3D from stereo produces streak-like artifacts near object boundaries

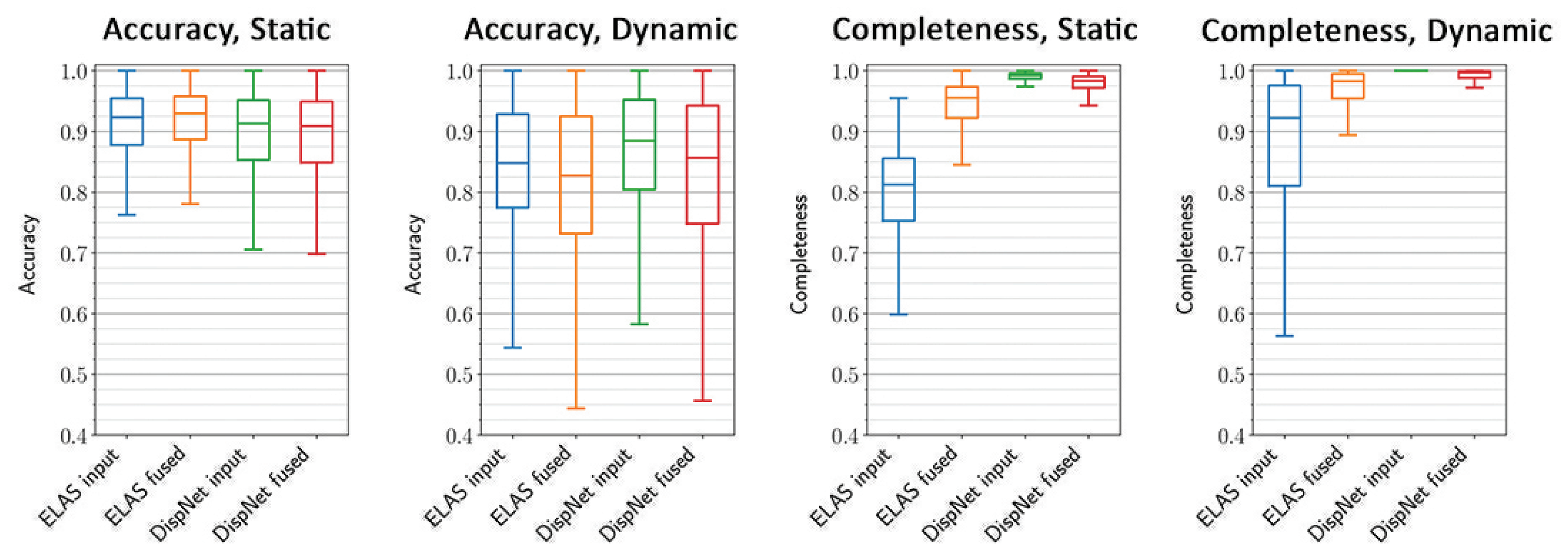


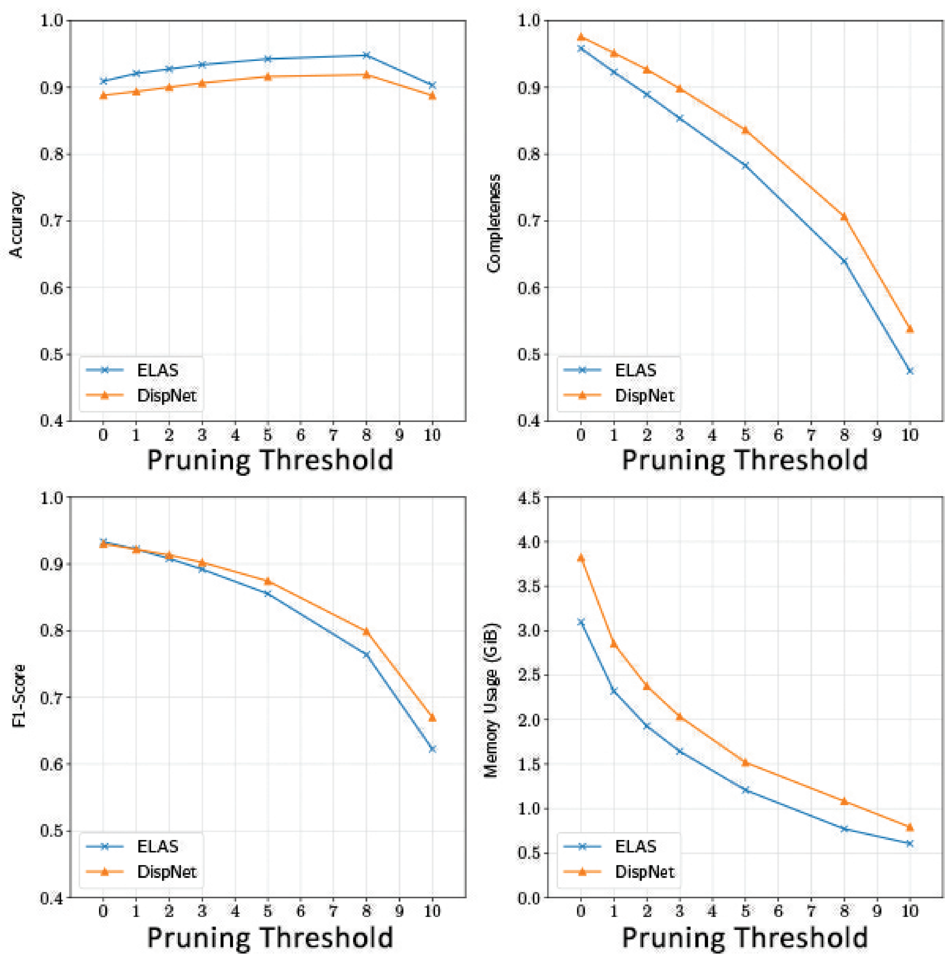
Solution: enforce temporal consistency and prune spurious structures in the map.

Reconstruction performed using the InfiniTAM system, leveraging voxel block hashing for efficient measurement fusion.

3. Results

- Evaluation performed on the KITTI dataset.
- Use LIDAR as depth ground truth.
- Compare two depth from stereo methods:
  - ELAS [Geiger et al., "Efficient large-scale stereo matching." ACCV, 2010.]
  - DispNet [Mayer, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." CVPR 2016.]
- We also show that map pruning can substantially reduce memory usage with only a small loss in map quality.





Additional Information

- Supplementary results as well as the video and source code are available on the project website: andreibarsan.github.io/dynslam
- The experiments use both ELAS and DispNet to compute depth from stereo because they leverage very different approaches: ELAS is geometry-focused, and DispNet is learning-focused.
- The visual odometry and the sparse scene flow are computed using libviso2.
- The semantic instance segmentation is computed using the multi-task network cascade (MNC) architecture.
- Directions for future work include:
  - Improved speed
  - Global consistency (loop closures)
  - More robust vehicle tracking in 3D

Acknowledgements

The authors of this paper would like to thank Torsten Sattler for his valuable support and feedback during the development of the paper and poster.