Massively Parallel Multiview Stereopsis by Surface Normal Diffusion

Silvano Galliani Katrin Lasinger Konrad Schindler

Photogrammetry and Remote Sensing, ETH Zurich

1. Additional KITTI results

We show additional results for the stereo dataset of the *KITTI Vision Benchmark Suite* [2]. The KITTI Vision Benchmark Suite [2] consists of 194 training and 195 test image pairs (rectified) with a resolution of 1240×376 pixels. It is well known to provide a challenging, realistic testbed due to outdoor lighting conditions. The experiment thus emphasizes that our method is not limited to (or even particularly tuned to) laboratory settings, but rather can deal with general stereo and multiview problems, also under uncontrolled lighting.

The dataset consists of outdoor images captured with a stereo rig from a moving car. Images include rural areas as well as streets around the city of Karlsruhe. Semi-dense ground truth disparity maps are provided for non-occluded areas as well as for the complete image (including regions that are occluded in the second view).

Multi-view KITTI In addition to computing results with the synchronous stereo pairs, we also process the data from three consecutive time steps together to obtain a 6-view reconstruction. This is possible, since the KITTI scenes are largely static. The relative rotation and translation of the stereo rig between consecutive frames were kindly provided by the authors of [3], who have in their work already computed the ego-motion of the stereo camera system. A similar ego-motion estimate was also described in [1].

Qualitative results Exemplary reconstruction results for both 2 and 6 views are shown, to highlight the improvement obtained by collecting evidence from multiple images, see Fig. 1, 2, 3. As for two views, we show disparity maps from the left to the right image of the rig, at the second of the three adjacent time steps (reprojected from the estimated 3D points). We point out that the results were obtained without explicit smoothness assumptions, and without any post-processing. To reconstruct a single depthmap the computation takes ≈ 2.1 seconds for a stereo pair, respectively ≈ 6.5 seconds for six views.

References

[1] H. Badino and T. Kanade. A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion. *IAPR MVA 2011*.

[2] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. *CVPR 2012*.

[3] C. Vogel, K. Schindler, and S. Roth. Piecewise rigid scene flow. *ICCV 2013*.

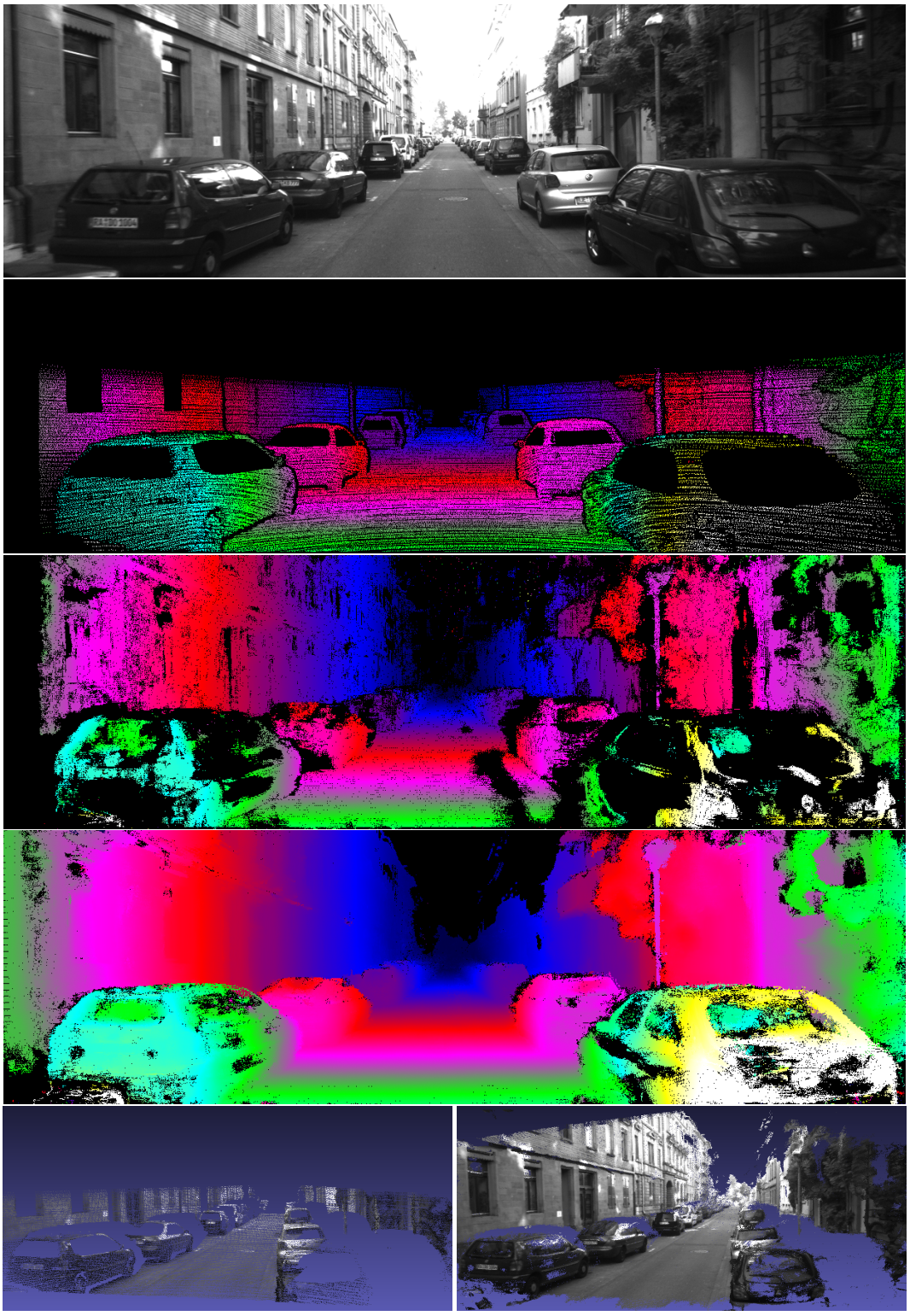

Figure 1: *From top to bottom:* Input image number 63 of KITTI training dataset. Ground truth disparity. Disparity with a stereo pair. Disparity with six input images. *Bottom left:* Ground truth point cloud. *Bottom right:* Reconstructed point cloud.

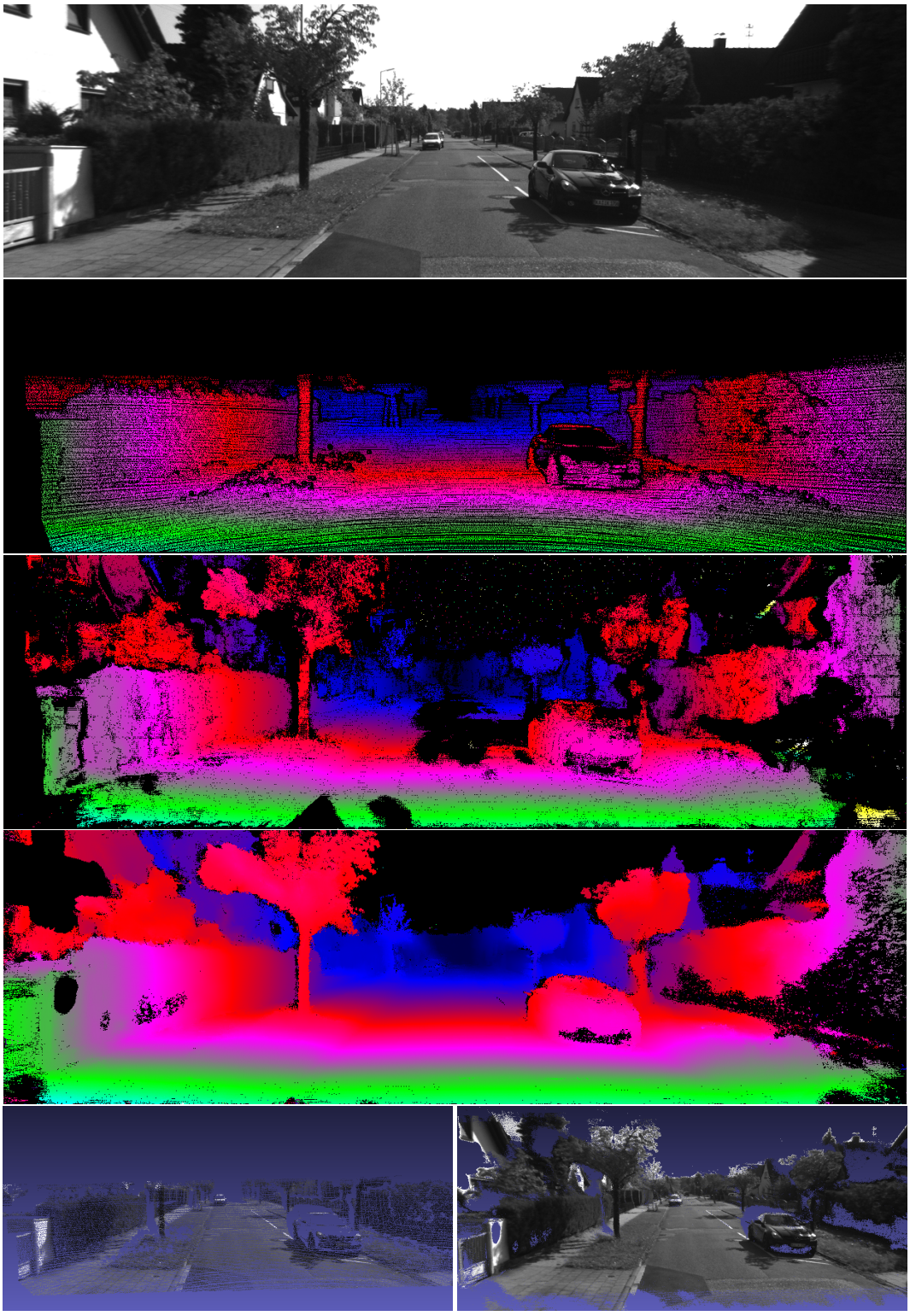

Figure 2: *From top to bottom:* Input image number 76 of KITTI training dataset. Ground truth disparity. Disparity with a stereo pair. Disparity with six input images. *Bottom left:* Ground truth point cloud. *Bottom right:* Reconstructed point cloud.

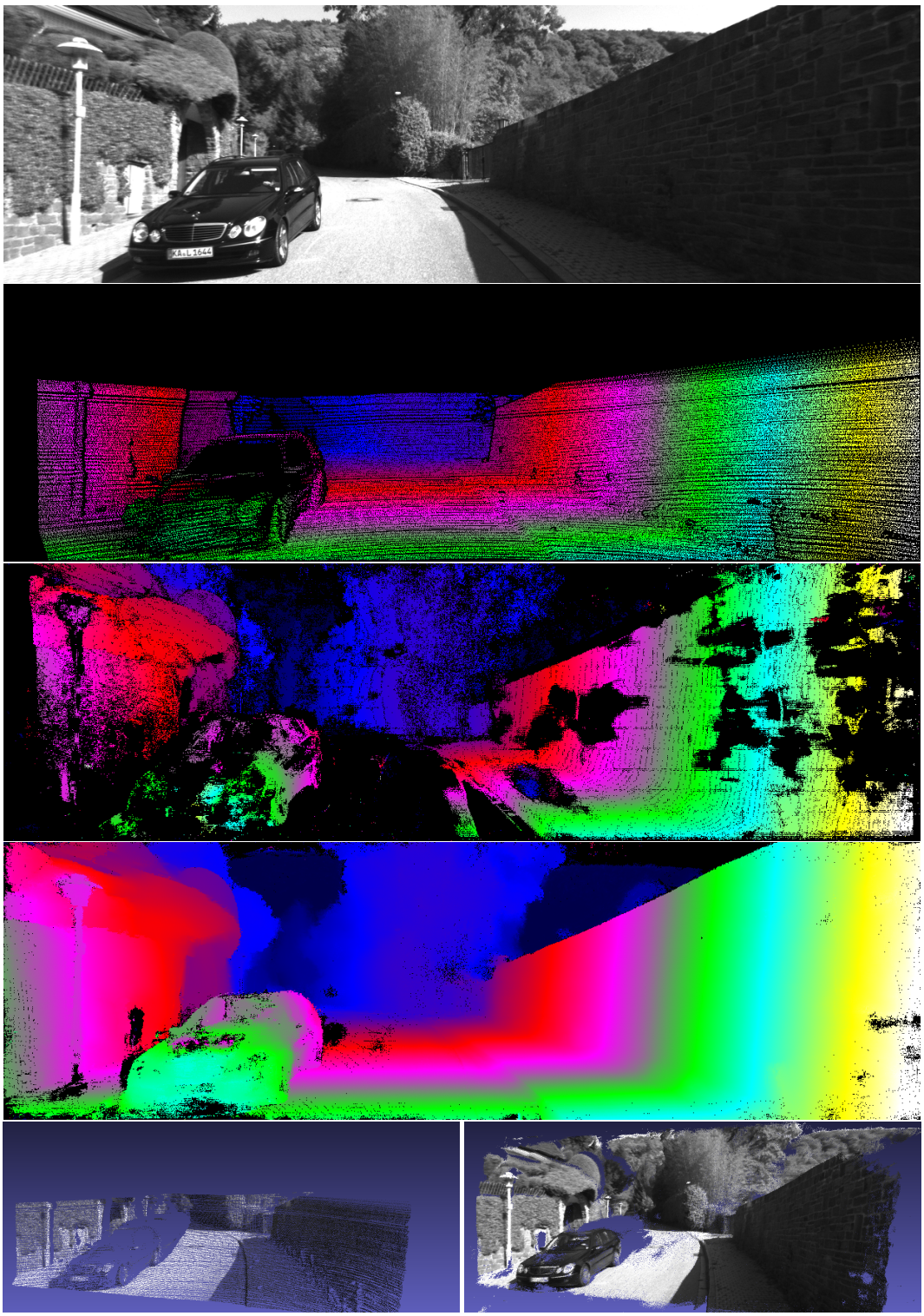

Figure 3: *From top to bottom:* Input image number 126 of KITTI training dataset. Ground truth disparity. Disparity with a stereo pair. Disparity with six input images. *Bottom left:* Ground truth point cloud. *Bottom right:* Reconstructed point cloud.