Object Scene Flow for Autonomous Vehicles

Moritz Menze[1], Andreas Geiger[2]
[1]Leibniz Universität Hannover. [2]MPI Tübingen.

This paper proposes a novel model and dataset for 3D scene flow estimation with an application to autonomous driving. Taking advantage of the fact that outdoor scenes often decompose into a small number of independently moving objects, we represent each element in the scene by its rigid motion parameters and each superpixel by a 3D plane as well as an index to the corresponding object. This minimal representation increases robustness and leads to a discrete-continuous CRF where the data term decomposes into pairwise potentials between superpixels and objects. Moreover, our model intrinsically segments the scene into its constituting dynamic components. We demonstrate the performance of our model on existing benchmarks as well as a novel realistic dataset with scene flow ground truth. We obtain this dataset by annotating 400 dynamic scenes from the KITTI raw data collection using detailed 3D CAD models for all vehicles in motion. Our experiments also reveal novel challenges which cannot be handled by existing methods.

Model: We focus on the classical scene flow setting where the input is given by two consecutive stereo image pairs of a calibrated camera and the task is to determine the 3D location and 3D flow of each pixel in a reference frame. Following the idea of piecewise-rigid shape and motion [10], a slanted-plane model is used, i.e., we assume that the 3D structure of the scene can be approximated by a set of piecewise planar superpixels. As opposed to existing works, we assume a finite number of ridigly moving objects in the scene. Our goal is to recover this decomposition jointly with the shape of the superpixels and the motion of the objects. We specify the problem in terms of a discrete-continuous CRF

$$E(\mathbf{s},\mathbf{o}) = \sum_{i\in\mathcal{S}} \underbrace{\varphi_i(\mathbf{s}_i,\mathbf{o})}_{\text{data}} + \sum_{i\sim j} \underbrace{\psi_{ij}(\mathbf{s}_i,\mathbf{s}_j)}_{\text{smoothness}} \qquad (1)$$

where $\mathbf{s}_i$ determines the shape and associated object of superpixel $i \in \mathcal{S}$, and $\mathbf{o}$ are the rigid motion parameters of all objects in the scene.

Dataset: Currently there exists no realistic benchmark dataset providing dynamic objects and ground truth for the evaluation of scene flow or optical flow. Therefore, in this paper, we take advantage of the KITTI raw data [5] to create a realistic and challenging scene flow benchmark with independently moving objects and annotated ground truth, comprising 200 training and 200 test scenes in total. The process of ground truth generation is especially challenging in the presence of individually moving objects since they cannot be easily recovered from laser scanner data alone due to the rolling shutter of the Velodyne and the low framerate (10 fps). Our annotation work-flow consists of two major steps: First, we recover the static background of the scene by removing all dynamic objects and compensating for the vehicle's egomotion. Second, we re-insert the dynamic objects by fitting detailed CAD models to the point clouds in each frame. We make our dataset and evaluation available as part of the KITTI benchmark.

Results: As input to our method, we leverage sparse optical flow from feature matches [3] and SGM disparity maps [6] for both rectified frames. We obtain superpixels using StereoSLIC [11] and initialize the rigid motion parameters of all objects in the scene by greedily extracting rigid body motion hypotheses using the 3-point RANSAC algorithm. By reasoning jointly about the decomposition of the scene into its constituent objects as well as the geometry and motion of all objects in the scene, the proposed model is able to produce accurate dense 3D scene flow estimates, comparing favorably with respect to current state-of-the-art as illustrated in Table 1. Our method also performs on par with the leading entries on the original static KITTI dataset [4] as well as on the Sphere sequence of Huguet et al. [8].

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

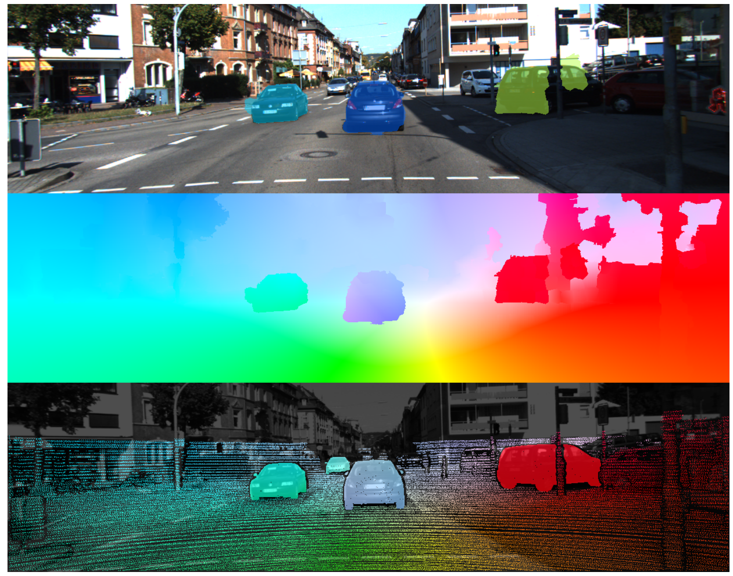

Figure 1: **Scene Flow Results on the proposed Scene Flow Dataset.** Top-to-bottom: Estimated segmentation into moving objects with background in transparent, optical flow results and proposed optical flow ground truth.

	bg	*fg*	*bg+fg*
Huguet [8]	67.69	64.03	67.08
GCSF [2]	52.92	59.11	53.95
SGM [6] + LDOF [1]	43.99	44.77	44.12
SGM [6] + Sun [9]	38.21	53.03	40.68
SGM [6] + Sphere Flow [7]	23.09	37.11	25.42
PRSF [10]	13.49	33.71	16.85
Our full model	**7.01**	**28.76**	**10.63**

Table 1: **Quantitative Results on the Proposed Scene Flow Dataset.** This table shows scene flow errors in %, averaged over all 200 test images. We provide separate numbers for the background region (bg), all foreground objects (fg) as well as all pixels in the image (bg+fg).

[1] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *PAMI*, 33:500–513, March 2011.

[2] Jan Cech, Jordi Sanchez-Riera, and Radu P. Horaud. Scene flow estimation by growing correspondence seeds. In *CVPR*, 2011.

[3] Andreas Geiger, Julius Ziegler, and Christoph Stiller. StereoScan: Dense 3D reconstruction in real-time. In *IV*, 2011.

[4] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *CVPR*, 2012.

[5] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *IJRR*, 32(11):1231–1237, September 2013.

[6] Heiko Hirschmueller. Stereo processing by semiglobal matching and mutual information. *PAMI*, 30(2):328–341, 2008.

[7] Michael Hornacek, Andrew Fitzgibbon, and Carsten Rother. SphereFlow: 6 DoF scene flow from RGB-D pairs. In *CVPR*, 2014.

[8] Frédéric Huguet and Frédéric Devernay. A variational method for scene flow estimation from stereo sequences. In *ICCV*, 2007.

[9] Deqing Sun, Stefan Roth, and Michael J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *IJCV*, 106(2):115–137, 2013.

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

[11] K. Yamaguchi, D. McAllester, and R. Urtasun. Robust monocular epipolar flow estimation. In *CVPR*, 2013.