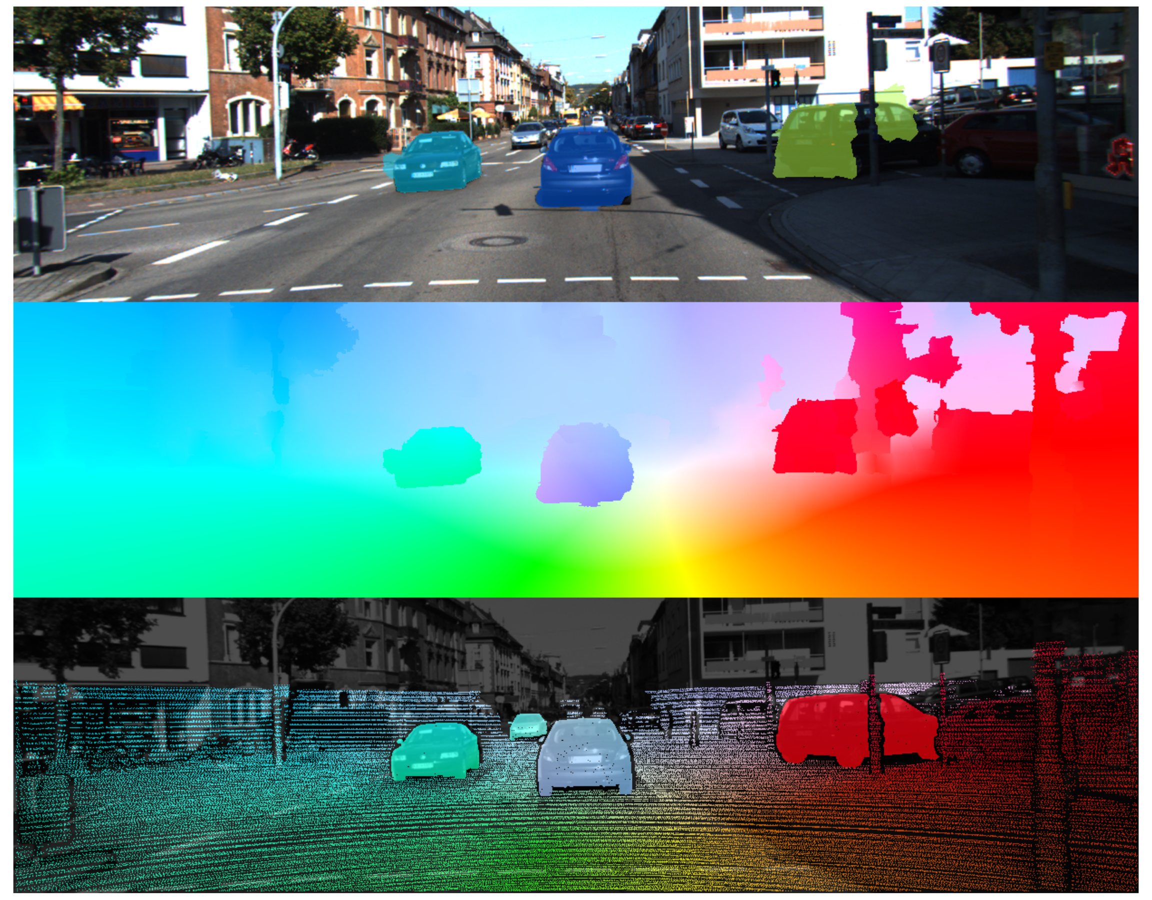
Object Scene Flow for Autonomous Vehicles

Moritz Menze[1], Andreas Geiger[2]

[1]Leibniz Universität Hannover, [2]MPI for Intelligent Systems, Tübingen

Abstract

We propose a novel model and dataset for 3D scene flow estimation with an application to autonomous driving. As 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 and an index to the corresponding object. This minimal representation increases robustness and is formalized in a discrete-continuous CRF. Our model intrinsically segments the scene into its constituting dynamic components. We demonstrate the performance of our model on existing benchmarks and 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 3D CAD models for all vehicles in motion. Our experiments reveal novel challenges which cannot be handled by existing methods.



Results and Ground Truth. Estimated segmentation into moving objects (top), optical flow (center) and proposed optical flow ground truth (bottom).

Representation

- Following the idea of piecewise-rigid shape and motion [8], the 3D scene is approximated by planar superpixels from StereoSLIC

$$\mathbf{s} = \{\mathbf{s}_i | i \in \mathcal{S}\}$$

- As opposed to existing works, we assume a finite number of rigidly moving objects in the scene $\mathbf{o} = \{\mathbf{o}_i | i \in \mathcal{O}\}$
- Superpixels are parameterized by a plane $\mathbf{n}$ and an object index k:

$$\mathbf{s}_i = (\mathbf{n}_i, k_i)^T$$

- Objects store their rigid motion parameters $\mathbf{o}_j = (\mathbf{R}_j, \mathbf{t}_j)^T$
- The superpixels inherit motion parameters from the assigned object

Model

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

Data Term

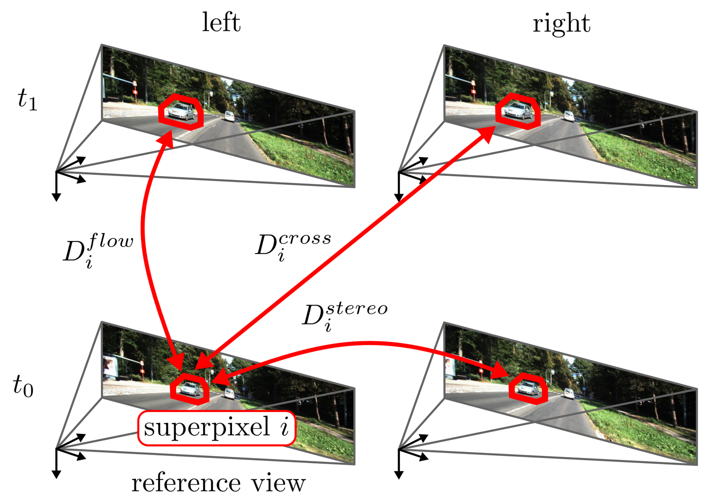
The data term φ consists of pairwise potentials which are evaluated for 3 pairs of images (see figure below)

$$\varphi_i(\mathbf{s}_i, \mathbf{o}) = \sum_{j \in \mathcal{O}} [k_i = j] \cdot D_i(\mathbf{n}_i, \mathbf{o}_j)$$

where the Iverson bracket restricts φ to the selected object.

To compute $D_i(\mathbf{n}_i, \mathbf{o}_j)$ we leverage:

- Dense Census features
- Sparse optical flow from feature matches
- SGM disparity maps for both rectified pairs



Smoothness Term

Our smoothness potential $\psi_{ij}(\mathbf{s}_i, \mathbf{s}_j)$ decomposes as:

$$\psi_{ij}(\mathbf{s}_i, \mathbf{s}_j) = \theta_3 \, \psi_{ij}^{\text{depth}}(\mathbf{n}_i, \mathbf{n}_j) + \theta_4 \, \psi_{ij}^{\text{orient}}(\mathbf{n}_i, \mathbf{n}_j) + \theta_5 \, \psi_{ij}^{\text{motion}}(\mathbf{s}_i, \mathbf{s}_j)$$

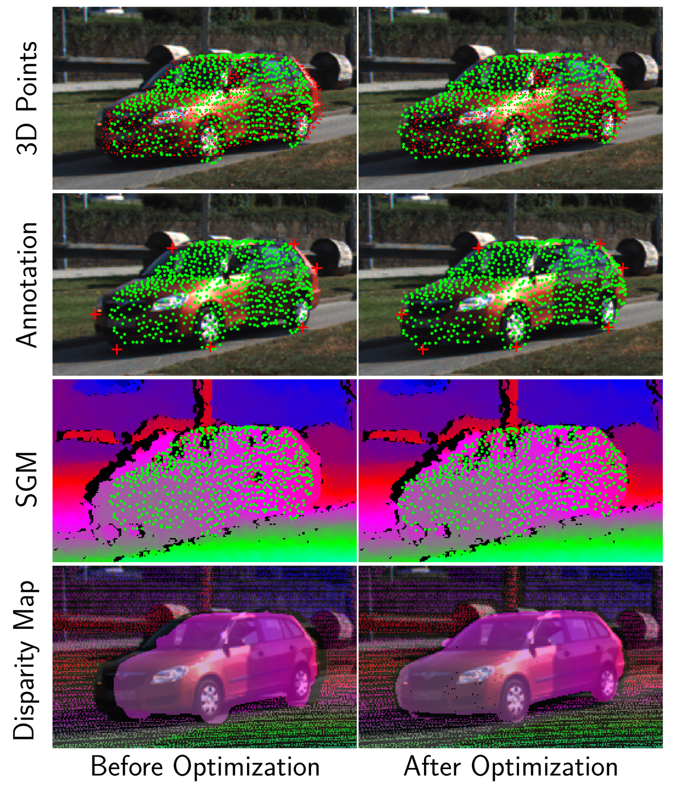
with weights θ controlling the individual terms:

- Regularization of depth is achieved by penalizing differences in disparity at shared boundary pixels (ψ_{ij}^{depth})
- We encourage the orientation of neighboring planes to be similar by evaluating the difference of plane normals $\mathbf{n}$ ($\psi_{ij}^{\text{orient}}$)
- Coherence of the assigned object indices is enforced by an orientation-sensitive Potts model ($\psi_{ij}^{\text{motion}}$)

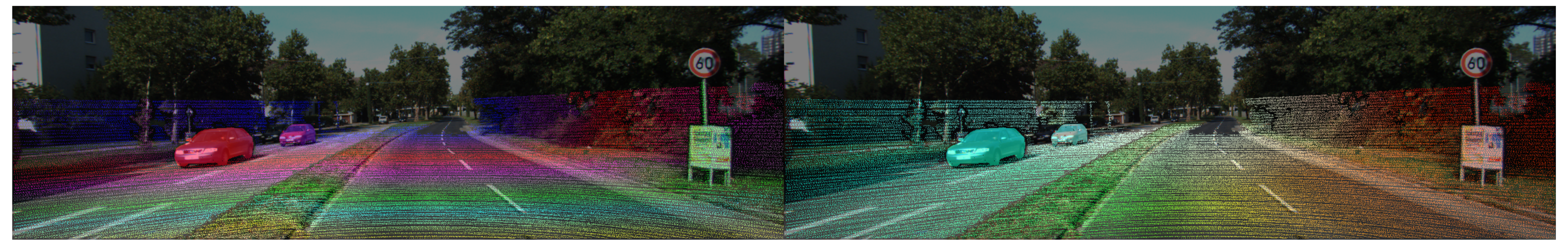
Inference

- We use max-product particle BP to jointly infer shape and motion parameters with TRW-S for the inner loop
- Particles are drawn from normal distributions around the current MAP solution and from neighboring superpixels

Dataset

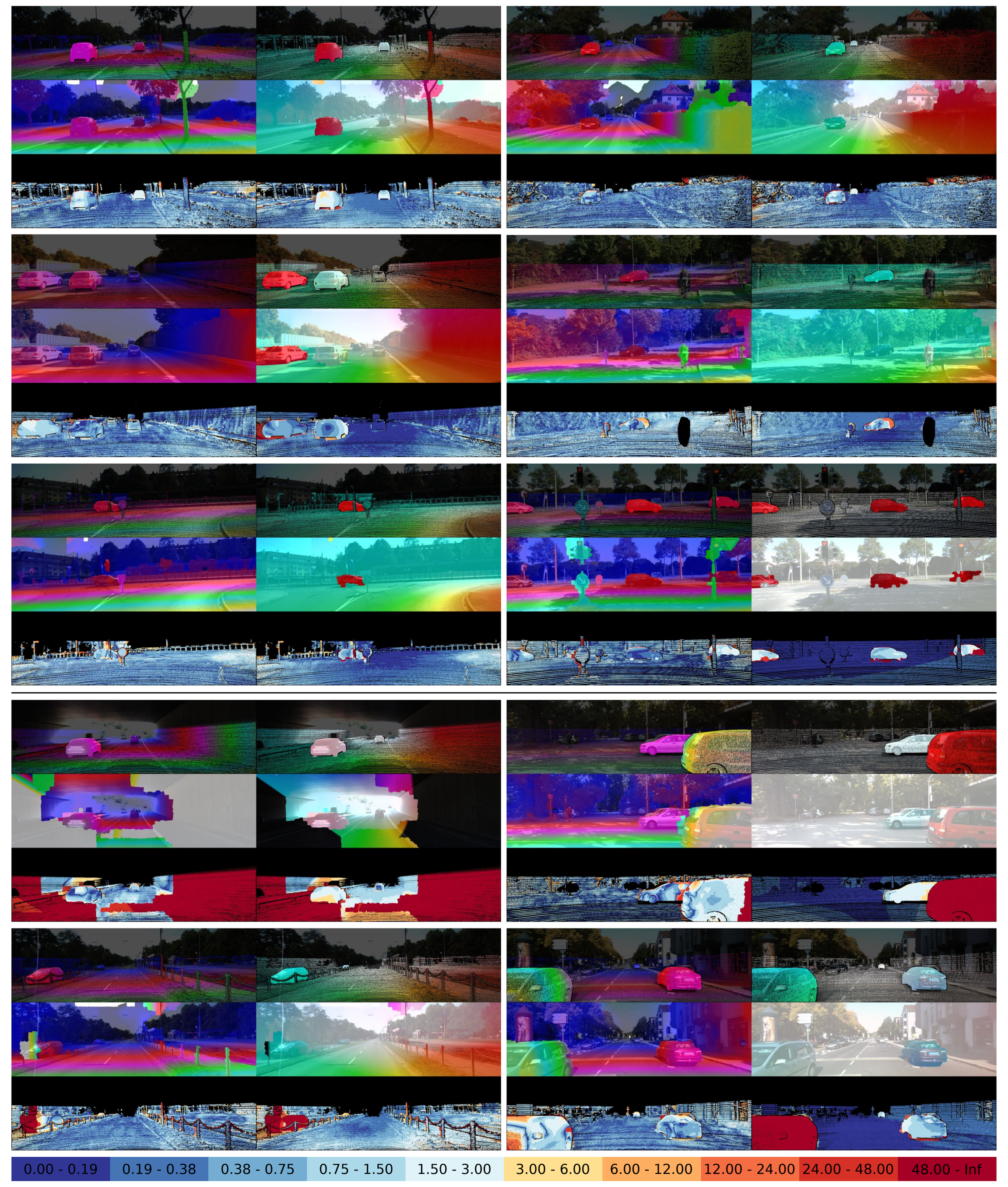


- We propose a novel realistic scene flow dataset which includes dynamic objects
- We annotated 200 training and 200 test scenes based on KITTI raw data
- The static background of the scenes is recovered from laser scanner data by removing all dynamic objects and compensating for the vehicle's ego-motion
- Dynamic objects are captured by fitting detailed CAD models to
  - accumulated 3D laser point clouds in each frame
  - manually annotated 2D control points
  - SGM disparity maps



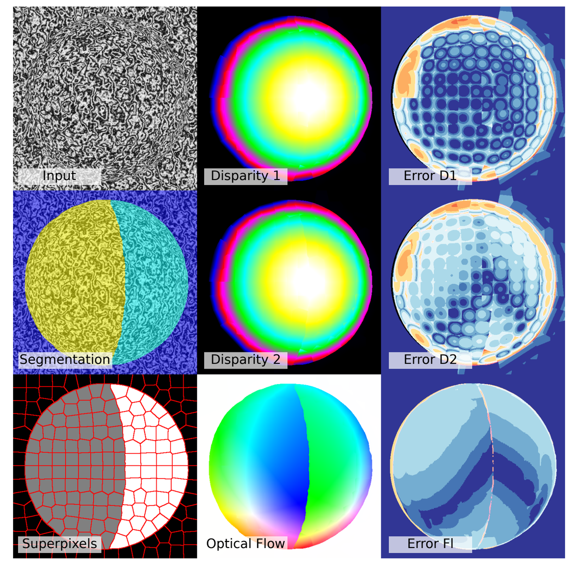
Ground Truth. Disparity at t_0 (left) and optical flow (right).

Qualitative Results



Qualitative Results. Each subfigure shows from top-to-bottom: disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our algorithm, and the respective error images using the color scheme depicted in the legend. The four scenes below the horizontal line are failure cases.

Results on the "Sphere" sequence



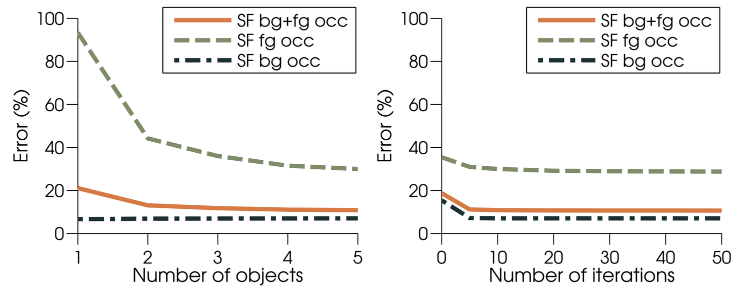
	[7]	[5]	[9]	[8]	Ours
RMSE 2D Flow	0.63	0.69	0.77	0.63	**0.55**
RMSE Disparity	3.8	3.8	10.9	2.84	**2.58**
RMSE Scene Flow	1.76	2.51	2.55	1.73	**0.75**

Illustration of our results for the synthetic "Sphere" sequence [5]. Top-left to bottom-right: The left input image of the first frame, our first disparity/error map, the obtained segmentation into different rigid body motions, the second disparity/error map, the superpixels we use, the recovered optical flow/error map.

Quantitative Results

	bg	*fg*	*bg+fg*
Huguet [5]	67.69	64.03	67.08
GCSF [2]	52.92	59.11	53.95
SGM [3] + LDOF [1]	43.99	44.77	44.12
SGM [3] + Sun [6]	38.21	53.03	40.68
SGM [3] + Sphere Flow [4]	23.09	37.11	25.42
PRSF [8]	13.49	33.71	16.85
Our full model	**7.01**	**28.76**	**10.63**

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). Outliers are defined as those values exceeding ground truth by at least 3 px *and* 5%.



Parameter Variation. This figure shows the scene flow errors of our method on the proposed dataset with respect to the number of object proposals and MP-PBP iterations.

References

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

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

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

[6] 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.

[7] Levi Valgaerts, Andres Bruhn, Henning Zimmer, Joachim Weickert, Carsten Stoll, and Christian Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In ECCV, 2010.

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

[9] Andreas Wedel, Clemens Rabe, Tobi Vaudrey, Thomas Brox, Uwe Franke, and Daniel Cremers. Efficient dense scene flow from sparse or dense stereo data. In ECCV, 2008.

Contact Information

Web http://www.cvlibs.net/projects/objectsceneflow

Email menze@ipi.uni-hannover.de, andreas.geiger@tue.mpg.de