Supplementary Material: Object Scene Flow for Autonomous Vehicles

Moritz Menze
Leibniz Universität Hannover
menze@ipi.uni-hannover.de

Andreas Geiger
MPI Tübingen
andreas.geiger@tue.mpg.de

Abstract

This **supplementary document** *provides additional descriptions, visualizations and experiments. We start by visualizing the 3D CAD models used for generating our scene flow ground truth. Next, we provide a detailed description of the initialization procedure for object hypotheses in our model. For reproducibility, we also detail all (estimated) model parameters together with plots illustrating the sensitivity of our model with respect to the choice of parameters. We further demonstrate that the loss in performance is small when limiting the runtime of our method to two minutes per scene on a single core. We also provide additional qualitative results of our method on the sphere sequence and quantitatively compare our method to state-of-the-art stereo, optical flow and scene flow approaches on the well-established KITTI stereo and optical flow benchmarks. Finally, we show additional quantitative and qualitative results on the novel scene flow dataset.*

1. 3D CAD Models for Ground Truth Annotation

Figure 1 depicts the 3D CAD models we used to generate ground truth for the proposed scene flow dataset. In order to adjust the shape of the models to the image observations/laser scans, we estimate all 9 parameters of a 3D similarity transformation as explained in section 4.2 of the paper.

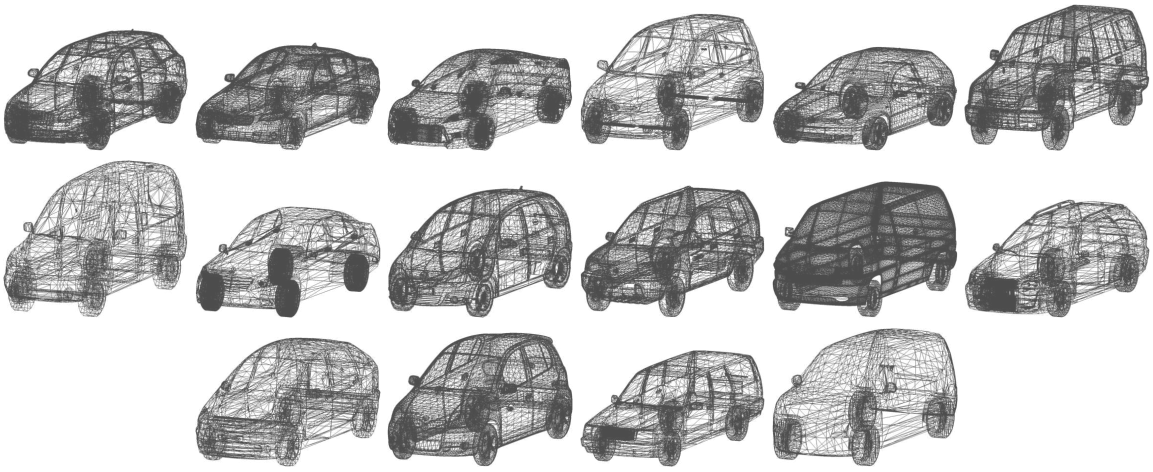


Figure 1: **3D CAD Models.** This figure shows the 3D CAD models we used to create ground truth for the proposed dataset.

978-1-4673-6964-0/15/$31.00 ©2015 IEEE

2. Initialization of Object Hypotheses

In this section we provide a more detailed description of the procedure we use to establish initial object hypotheses. We start by extracting sparse scene flow correspondences and robustly estimating the dominant motion in the scene (i.e., the background/camera motion) using the approach described in [3]. In order to reduce the number of outliers in the correspondence set, we mask all (background) feature points in regions which leave the image domain according to the estimated dominant motion and the initial SGM disparity estimate as illustrated in Fig. 2.

Figure 2: **Visibility Mask.** Image regions visible in the first frame but not in the second frame are shown in white.

Figure 3: **Object Hypotheses Generation.** This figure depicts the optical flow field induced by the camera motion and the remaining sparse scene flow features grouped into individual motion hypotheses using color-coded flow vectors.

Given the scene flow correspondences which do not agree with the background motion (deviate by more than 5 pixels), we extract object hypotheses as follows: We randomly sample 50 correspondences and fit a rigid motion model to all correspondences within a ball of radius 2.5 meters in 3D using the 3-point RANSAC algorithm. The resulting proposals are sorted with respect to the number of respective inliers. Furthermore, we apply non-maxima suppression to avoid multiple overlapping proposals on the same object. The process is illustrated in Fig. 3, where 3 motion hypotheses have been recovered.

3. Training of Parameters

In order to obtain the model parameters $\{\theta\}$ and $\{\tau\}$, we perform block coordinate descent on a subset of 30 randomly selected training images. The learned weights and truncation thresholds are shown in Table 1. Following the notation in the paper, we denote the truncation thresholds for sparse features by $\tau_{1,\mathrm{x}}$ where $\mathrm{x} \in \{\text{stereo}, \text{flow}, \text{cross}\}$. The truncation value of all dense Census features is denoted $C_{\max}$. The provided values refer to the normalized Hamming distance. In contrast to the description in the paper the outlier penalty C of the dense Census features does not equal the truncation value $C_{\max}$ but is estimated from training data according to Table 1. We will correct this in the final version.

	$\theta_{1,\text{stereo}}$	$\theta_{1,\text{flow}}$	$\theta_{1,\text{cross}}$	$\theta_{2,\text{stereo}}$	$\theta_{2,\text{flow}}$	$\theta_{2,\text{cross}}$	θ_3	θ_4	θ_5
Scene Flow	1.0	1.0	1.0	0.0235	0.5791	0.5791	0.3750	17.3949	166.2633
KITTI Stereo/Flow	1.0	1.0	1.0	1.4916	1.8727	1.8727	1.2000	538.6560	0.2633

	$C_{\max}$	C	$\tau_{1,\text{stereo}}$	$\tau_{1,\text{flow}}$	$\tau_{1,\text{cross}}$	τ_2	τ_3	λ
Scene Flow	0.8361	0.3590	2.4278	3.3947	3.3947	2.5559	0.2513	0.1324
KITTI Stereo/Flow	0.9096	0.0546	2.5252	9.5897	9.5897	5.0700	0.1844	0.0001

Table 1: **Trained parameters.** This table shows the values of the model parameters after training.

4. Sensitivity to Choice of Parameters

In this section we provide plots illustrating the sensitivity of our model to the choice of parameters. For each weight in Table 1 we investigate a range from 0.0 to 20 times the estimated optimal value. The first row of Fig. 4 shows that the weight of the dense census features has a significant impact on the performance while the weights of the sparse features are not that critical. Regarding the smoothness terms depicted in the second row, the weight of the boundary term has a dominant influence while the sensitivity with respect to the weights of normal and motion smoothness terms is relatively low. These results are in accordance with the results of the ablation studies described in Section 5 of the paper submission.

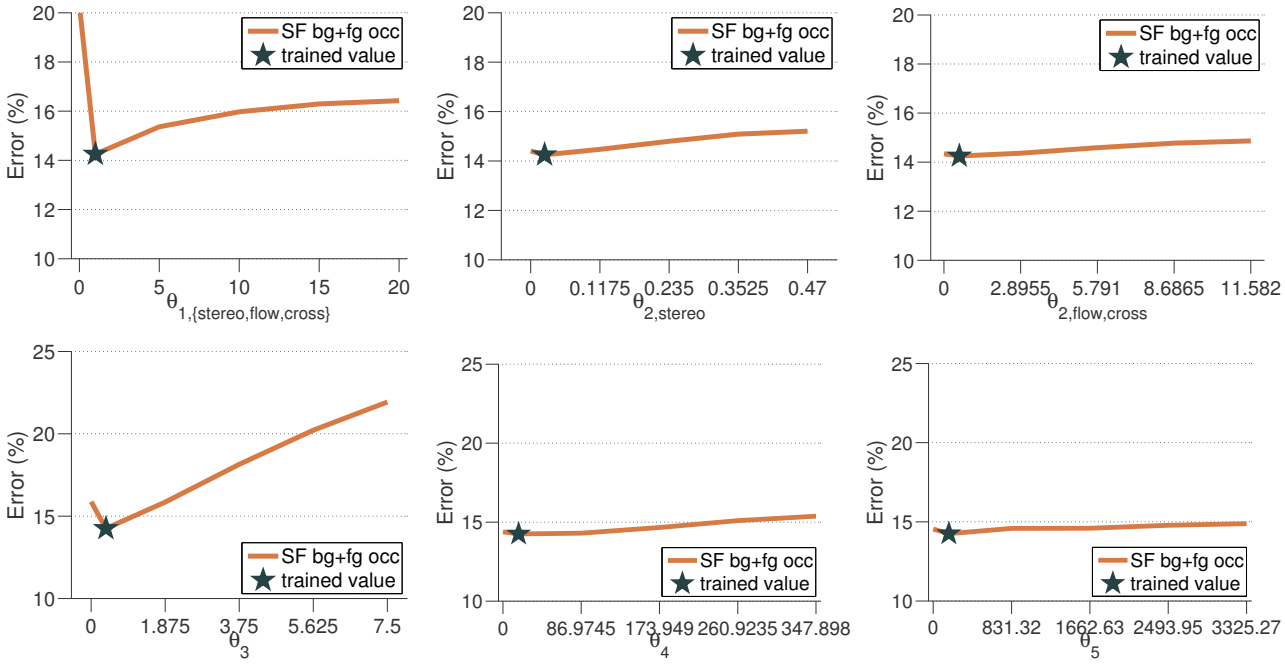


Figure 4: **Sensitivity with respect to Choice of Parameters.** The first row shows the impact of the data term weights. The second row shows the impact of smoothness term weights.

5. Results on the Proposed Dataset for Non-Occluded Regions

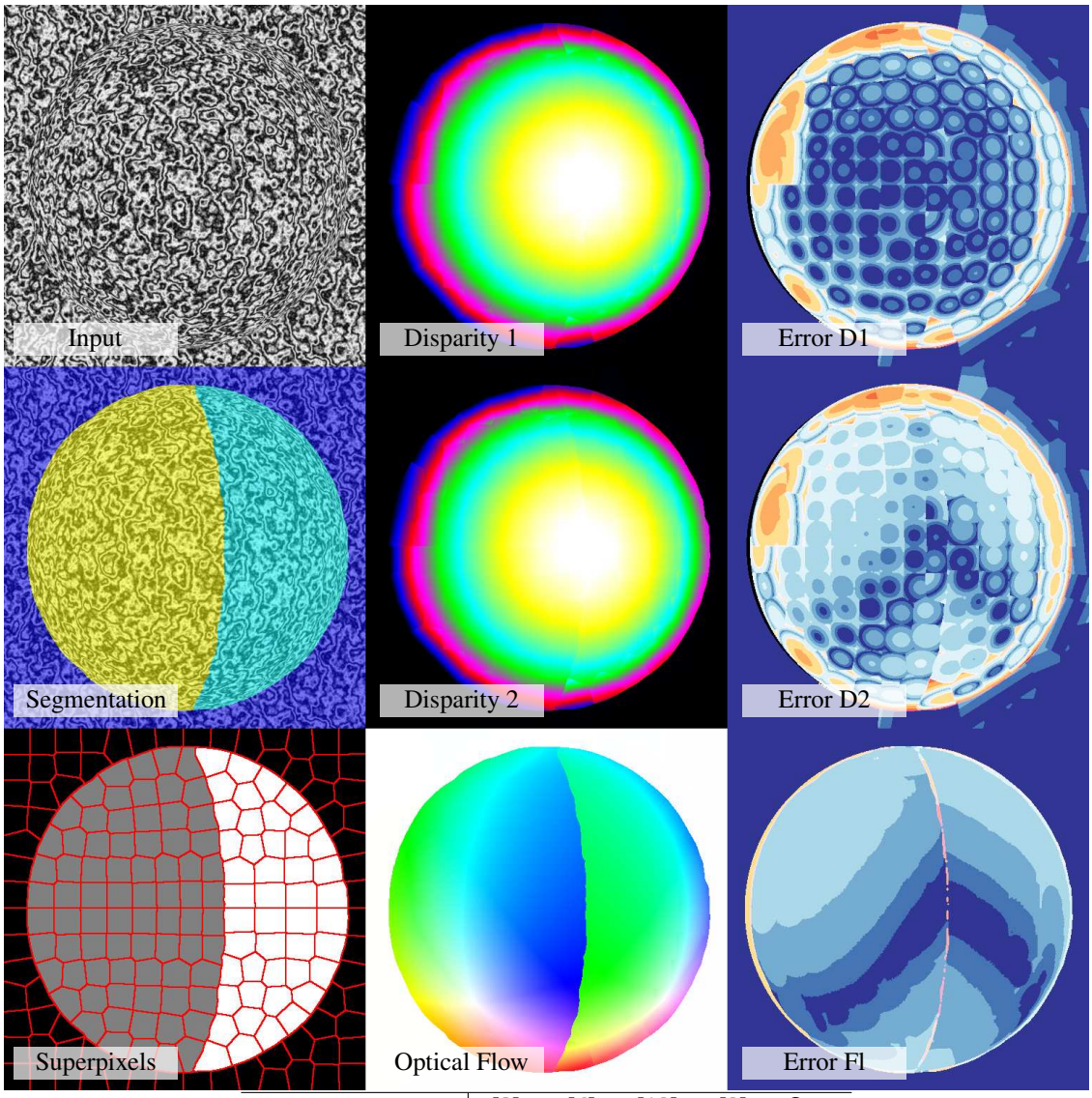
Table 2 provides additional quantitative results of our method and the baselines on the proposed dataset when excluding all half-occluded regions in the image (i.e., we do not evaluate pixels leaving the image domain in any of the four views). The proposed approach significantly outperforms all baselines with respect to scene flow estimation performance.

	D1			D2			Fl			SF		
	bg	*fg*	*bg+fg*	*bg*	*fg*	*bg+fg*	*bg*	*fg*	*bg+fg*	*bg*	*fg*	*bg+fg*
Huguet [6]	26.38	19.88	25.31	59.04	43.70	56.51	41.15	44.15	41.70	61.73	61.21	61.64
GCSF [2]	11.24	26.25	13.72	32.28	34.75	32.69	38.12	41.52	38.74	44.44	55.91	46.50
SGM [4] + LDOF [1]	4.75	14.21	6.31	28.67	22.37	27.63	30.41	31.33	30.57	33.78	40.57	35.00
SGM [4] + Sun [7]	4.75	14.21	6.31	27.79	24.43	27.23	23.04	41.91	26.46	26.96	49.52	31.01
SGM [4] + Sphere Flow [5]	4.75	14.21	6.31	13.45	21.99	14.86	13.36	25.20	15.51	15.64	33.12	18.78
PRSF [9]	4.41	13.09	5.84	10.57	19.55	12.05	6.94	23.64	9.97	8.78	29.43	12.50
Unary (SGM+SpF)	4.50	12.98	5.90	5.62	27.01	9.15	5.70	32.98	10.65	7.51	38.10	13.01
Unary (Census)	6.06	18.24	8.07	7.09	24.30	9.93	5.63	23.01	8.78	7.65	31.64	11.96
Unary (All)	5.23	15.64	6.95	6.22	22.62	8.93	5.15	22.29	8.26	6.90	30.02	11.06
Unary (All) + Pair (Boundary)	4.17	**9.87**	**5.11**	5.08	18.43	7.28	4.41	20.29	7.29	5.84	26.28	9.51
Unary (All) + Pair (Normal)	4.98	15.10	6.65	5.96	22.35	8.67	4.97	22.05	8.07	6.64	29.72	10.79
Unary (All) + Pair (Object)	5.37	16.79	7.26	6.30	22.86	9.04	5.15	21.26	8.07	6.92	29.31	10.95
Unary (SGM+SpF) + Pair (All)	6.19	24.82	9.26	7.33	32.56	11.49	7.00	34.34	11.96	9.11	40.82	14.81
Unary (Census) + Pair (All)	4.26	11.75	5.50	5.09	19.01	7.39	4.25	19.09	6.94	5.70	25.89	9.33
Unary (All) + Pair (All) Fast	**4.01**	11.76	5.29	**4.87**	19.08	7.21	**4.20**	20.02	7.07	5.64	26.71	9.43
Unary (All) + Pair (All)	4.14	11.11	5.29	4.95	**18.41**	**7.17**	4.21	**18.65**	**6.83**	**5.63**	**25.03**	**9.11**

Table 2: **Quantitative Results on the Proposed Scene Flow Dataset for all Non-Occluded Regions.** This table shows the disparity (D1/D2), flow (Fl) and scene flow (SF) errors averaged over all 200 test images. For each modality we separately provide the errors for the background region (bg), all foreground objects (fg) as well as all pixels in the image (bg+fg).

6. Sphere Sequence

We also provide results of our method on the synthetic "Sphere" dataset by Huguet et al. [6]. As described in the paper, we modify the StereoSLIC algorithm to consider dense optical flow instead of disparity for superpixel estimation and provide the Horn-Schunck results of Sun et al. [7] as input. Fig. 5 shows our results in terms of disparity, optical flow and color-coded error maps. Our method performs surprisingly well despite the fact that we restrict the scene to only 200 planar superpixels.



	[8]	[6]	[10]	[9]	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**

Figure 5: **Illustration of our results for the synthetic "Sphere" sequence [6].** This figure shows (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.

7. KITTI Benchmark

We also compare our results with respect to the state-of-the-art on the popular KITTI stereo and optical flow benchmarks[1]. Despite the fact that this dataset only contains static scenes our method performs excellent, yielding rank 5 in the stereo benchmark (Table 3) and rank 4 in the optical flow evaluation (Table 4) including anonymous submissions, motion stereo methods and approaches that make use of more than four images. When excluding all submissions in review and methods that make use of additional information or constraints, our method ranks first in both benchmarks. Table 3 and Table 4 show the 50 leading entries out of 65 listed stereo and 55 listed optical flow approaches at the time of submission. The third column "Setting" specifies additional information/constraints used by the individual methods:

- fl - Flow: Method uses optical flow (2 temporally adjacent images)
- mv - Multiview: Method uses more than 2 temporally adjacent images
- ms - Motion stereo: Method uses epipolar geometry for computing optical flow

Fig. 6 shows qualitative results for one test image as provided on the benchmark homepage.

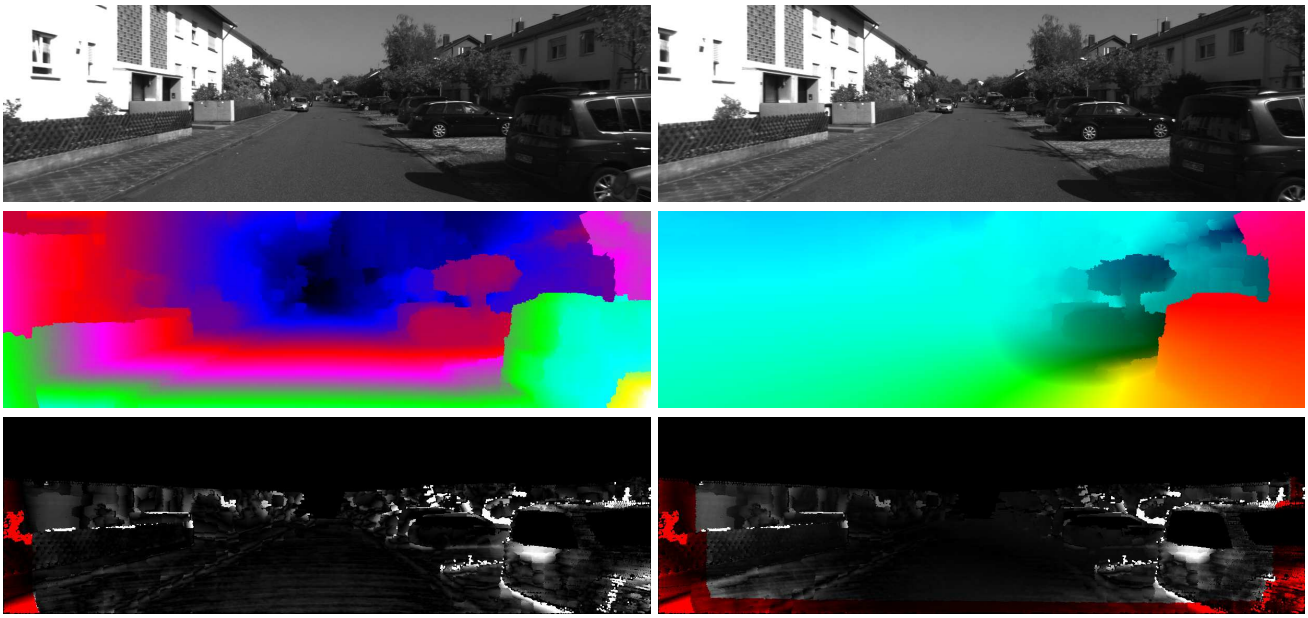


Figure 6: **Results on KITTI Stereo Benchmark.** This figure shows two subsequent input images (top), the disparity and flow maps estimated by our method (center), as well as the corresponding error maps (bottom, errors ≥ 5 pixels in white).

[1] http://www.cvlibs.net/datasets/kitti

Rank	Method	Setting	Out-Noc	Out-All	Avg-Noc	Avg-All	Runtime
1	CVPR 1186		2.47 %	3.27 %	0.7 px	0.9 px	265 s
2	MC-CNN		2.61 %	3.84 %	0.8 px	1.0 px	100 s
3	SPS-StFl	fl ms	2.83 %	3.64 %	0.8 px	0.9 px	35 s
4	VC-SF	fl mv	3.05 %	3.31 %	0.8 px	0.8 px	300 s
5	**Our method**	fl	3.28 %	4.07 %	0.8 px	0.9 px	50 min
6	CoR		3.30 %	4.10 %	0.8 px	0.9 px	6 s
7	SPS-St		3.39 %	4.41 %	0.9 px	1.0 px	2 s
8	PCBP-SS		3.40 %	4.72 %	0.8 px	1.0 px	5 min
9	DDS-SS		3.83 %	4.59 %	0.9 px	1.0 px	1 min
10	StereoSLIC		3.92 %	5.11 %	0.9 px	1.0 px	2.3 s
11	PR-Sf+E	fl	4.02 %	4.87 %	0.9 px	1.0 px	200 s
12	PCBP		4.04 %	5.37 %	0.9 px	1.1 px	5 min
13	PR-Sceneflow	fl	4.36 %	5.22 %	0.9 px	1.1 px	150 sec
14	CoR-Conf		4.49 %	5.26 %	1.0 px	1.2 px	6 s
15	OF-MPV		4.77 %	5.88 %	1.0 px	1.2 px	300 s
16	AARBM		4.86 %	5.94 %	1.0 px	1.2 px	0.25 s
17	rdSGM		4.91 %	6.07 %	1.2 px	1.3 px	10 s
18	wSGM		4.97 %	6.18 %	1.3 px	1.6 px	6s
19	ATGV		5.02 %	6.88 %	1.0 px	1.6 px	6 min
20	rSGM		5.03 %	6.60 %	1.1 px	1.5 px	0.2 s
21	iSGM		5.11 %	7.15 %	1.2 px	2.1 px	8 s
22	RBM		5.18 %	6.21 %	1.1 px	1.3 px	0.2 s
23	DLP		5.32 %	7.22 %	1.2 px	2.0 px	60 s
24	ALTGV		5.36 %	6.49 %	1.1 px	1.2 px	20 s
25	OCV-SGBM2		5.38 %	6.50 %	1.0 px	1.2 px	2 s
26	SNCC		5.40 %	6.44 %	1.2 px	1.3 px	0.11 s
27	AABM		5.42 %	6.52 %	1.1 px	1.3 px	0.43 s
28	GCA		5.45 %	6.54 %	1.1 px	1.2 px	10 s
29	Ensemble		5.48 %	7.00 %	1.6 px	2.1 px	135 s
30	SGM		5.76 %	7.00 %	1.2 px	1.3 px	3.7 s
31	TGV2ADC		5.87 %	6.79 %	1.1 px	1.2 px	8 s
32	ADCSGM		5.94 %	7.68 %	1.4 px	1.9 px	0.9 s
33	mSGM-LDE		6.01 %	8.22 %	1.4 px	2.4 px	55 s
34	CD		6.15 %	7.47 %	1.3 px	1.4 px	5 s
35	Toast2	st	6.16 %	7.42 %	1.2 px	1.4 px	0.03 s
36	ITGV		6.20 %	7.30 %	1.3 px	1.5 px	7 s
37	RWR		6.25 %	7.41 %	1.2 px	1.4 px	1 min
38	RWR+Gradient		6.40 %	7.49 %	1.4 px	1.6 px	5 s
39	LDE		6.73 %	8.85 %	1.8 px	2.5 px	14 s
40	BSSM		7.39 %	8.79 %	1.4 px	1.6 px	20.7 s
41	OCV-SGBM		7.64 %	9.13 %	1.8 px	2.0 px	1.1 s
42	SR-TMP	mv	8.00 %	8.65 %	1.9 px	2.1 px	50 s
43	MSMW	st	8.01 %	9.24 %	1.6 px	1.7 px	3 min
44	TV-WL1+ELAS	mv	8.11 %	8.43 %	1.5 px	1.6 px	1 min
45	ELAS		8.24 %	9.96 %	1.4 px	1.6 px	0.3 s
46	linBP		8.56 %	10.70 %	1.7 px	2.7 px	1.6 min
47	ELSE		8.72 %	9.90 %	1.7 px	1.9 px	1 s
48	S+GF (Cen)		9.03 %	11.21 %	2.1 px	3.4 px	140 s
49	SM_GPTM		9.79 %	11.38 %	2.1 px	2.6 px	6.5 s
50	LAMC-DS		9.82 %	11.49 %	2.1 px	2.7 px	10.8 min

Table 3: **Results on the KITTI stereo benchmark.** This table shows errors with respect to the (default) 3 px threshold.

Rank	Method	Setting	Out-Noc	Out-All	Avg-Noc	Avg-All	Runtime
1	VC-SF	st mv	2.72 %	4.84 %	0.8 px	1.3 px	300 s
2	SPS-StFl	st ms	2.82 %	5.61 %	0.8 px	1.3 px	35 s
3	SPS-Fl	ms	3.38 %	10.06 %	0.9 px	2.9 px	11 s
4	**Our method**	st	3.47 %	6.34 %	1.0 px	1.5 px	50 min
5	PR-Sf+E	st	3.57 %	7.07 %	0.9 px	1.6 px	200 s
6	PCBP-Flow	ms	3.64 %	8.28 %	0.9 px	2.2 px	3 min
7	PR-Sceneflow	st	3.76 %	7.39 %	1.2 px	2.8 px	150 sec
8	MotionSLIC	ms	3.91 %	10.56 %	0.9 px	2.7 px	11 s
9	PPR-Flow		5.76 %	10.57 %	1.3 px	2.9 px	800 s
10	NLTGV-SC		5.93 %	11.96 %	1.6 px	3.8 px	16 s
11	DDS-DF		6.03 %	13.08 %	1.6 px	4.2 px	1 min
12	TGV2ADCSIFT		6.20 %	15.15 %	1.5 px	4.5 px	12s
13	AnyFlow		6.37 %	15.80 %	1.5 px	4.3 px	15 s
14	BTF-ILLUM		6.52 %	11.03 %	1.5 px	2.8 px	80 seconds
15	CRT-TGV		6.71 %	12.09 %	2.0 px	3.9 px	10.5 min
16	Data-Flow		7.11 %	14.57 %	1.9 px	5.5 px	3 min
17	DeepFlow		7.22 %	17.79 %	1.5 px	5.8 px	17 s
18	EpicFlow		7.88 %	17.08 %	1.5 px	3.8 px	15 s
19	TVL1-HOG		7.91 %	18.90 %	2.0 px	6.1 px	180 s
20	MLDP-OF		8.67 %	18.78 %	2.4 px	6.7 px	160 s
21	DescFlow		8.76 %	19.45 %	2.1 px	5.7 px	9.0 s
22	SparseFlow		9.09 %	19.32 %	2.6 px	7.6 px	10 s
23	CRTflow		9.43 %	18.72 %	2.7 px	6.5 px	18 s
24	C++		10.04 %	20.26 %	2.6 px	7.1 px	8.5 min
25	TF+OFM	mv	10.22 %	18.46 %	2.0 px	5.0 px	350 s
26	C+NL		10.49 %	20.64 %	2.8 px	7.2 px	14.8 min
27	NNF-Local		10.68 %	21.09 %	2.7 px	7.4 px	1073 s
28	fSGM		10.74 %	22.66 %	3.2 px	12.2 px	60 s
29	TGV2CENSUS		11.03 %	18.37 %	2.9 px	6.6 px	4 s
30	C+NL-fast		12.36 %	22.28 %	3.2 px	7.9 px	2.9 min
31	EPPM		12.75 %	23.55 %	2.5 px	9.2 px	0.25 s
32	CVPR-738b		13.01 %	20.67 %	2.8 px	6.6 px	3.6 s
33	HS		14.75 %	24.11 %	4.0 px	9.0 px	2.6 min
34	CVPR-738a		16.04 %	24.60 %	3.0 px	6.5 px	0.28 s
35	GC-BM-Bino	st ms	18.83 %	29.30 %	5.0 px	12.1 px	1.3 s
36	IQFlow		18.84 %	28.25 %	3.6 px	8.8 px	60 s
37	C+NL-M		19.19 %	26.36 %	7.4 px	14.5 px	5 min
38	eFolki		19.31 %	28.79 %	5.2 px	10.9 px	0.026 s
39	GC-BM-Mono	ms	19.38 %	29.80 %	5.0 px	12.1 px	1.3 s
40	RSRS-Flow		20.78 %	29.75 %	6.2 px	12.1 px	4 min
41	ALD		21.37 %	30.71 %	10.9 px	16.0 px	110 s
42	LDOF		21.93 %	31.39 %	5.6 px	12.4 px	1 min
43	2Bit-BM-tele		24.10 %	33.59 %	7.1 px	15.2 px	6 min
44	HMM		24.78 %	34.19 %	7.2 px	15.1 px	10 min
45	DB-TV-L1		30.87 %	39.25 %	7.9 px	14.6 px	16 s
46	GCSF	st	33.17 %	41.71 %	7.0 px	15.3 px	2.4 s
47	HAOF		35.87 %	43.46 %	11.1 px	18.3 px	16.2 s
48	BERLOF		37.66 %	45.27 %	8.5 px	16.2 px	0.231 s
49	RLOF		38.60 %	46.13 %	8.7 px	16.5 px	0.488 s
50	SpaGloM		41.91 %	48.59 %	8.6 px	15.3 px	50 s

Table 4: **Results on the KITTI optical flow benchmark.** This table shows errors with respect to the (default) 3 px threshold.

8. Qualitative Results on the Proposed Dataset

In this section, we provide additional qualitative results of our method on the proposed dataset using 38 *randomly selected* scenes. For each scene, we show the ground truth disparity and optical flow map overlayed with the input images, the estimated disparity and optical flow maps and the corresponding error images using a logarithmic color coding where red shades represent errors above 3 pixels and blue shades denote errors below 3 pixels.

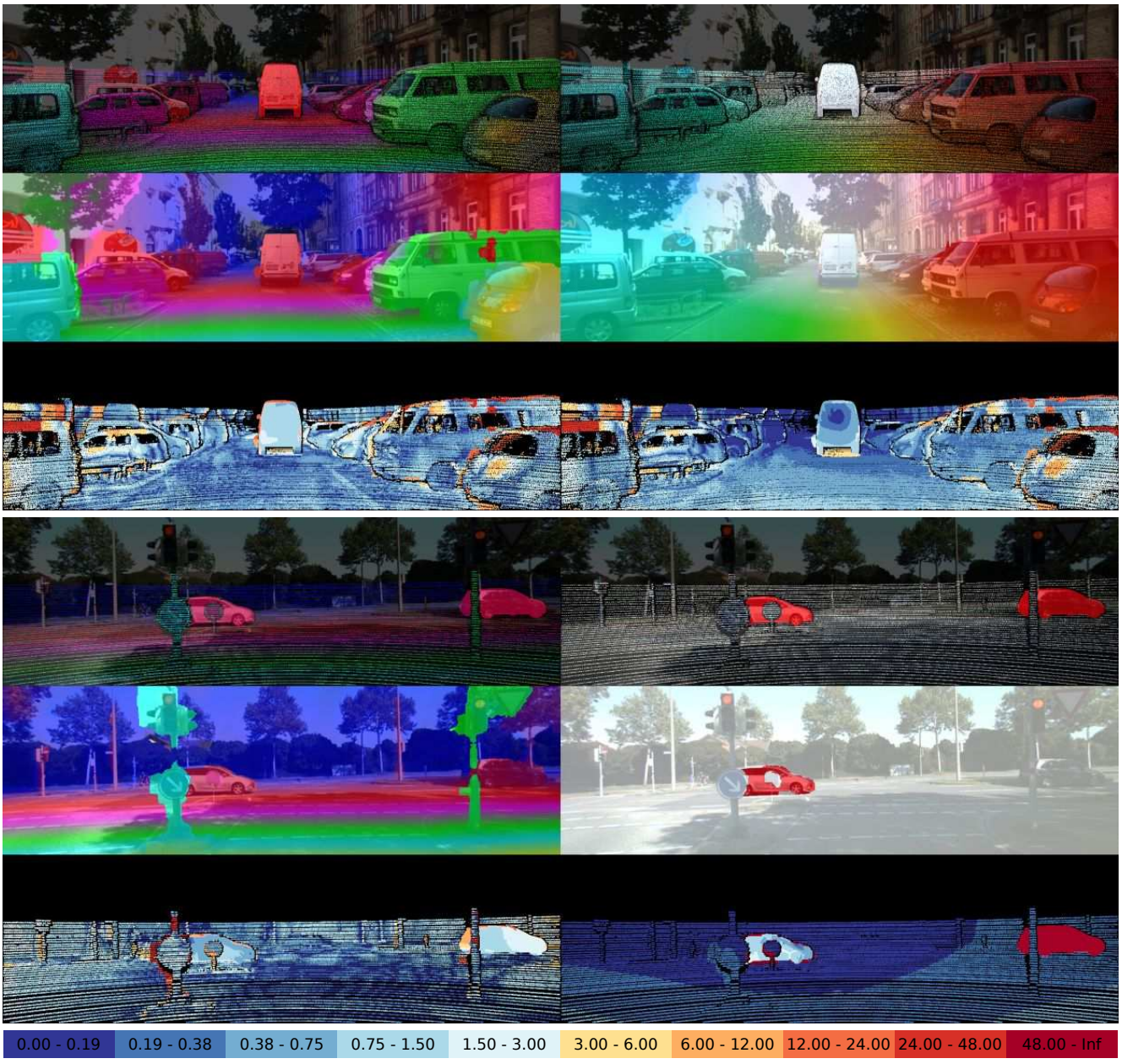


Figure 7: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

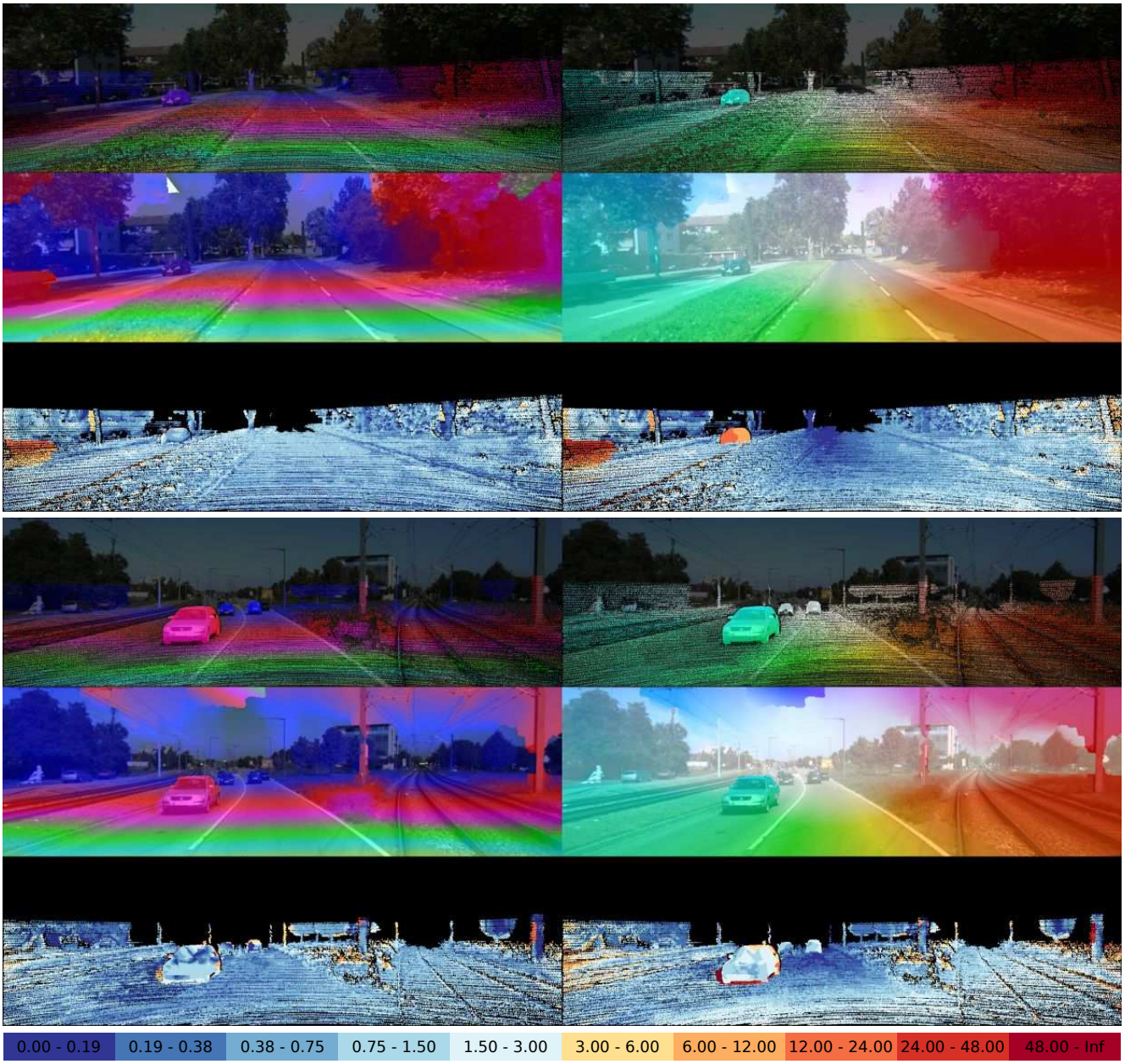


Figure 8: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

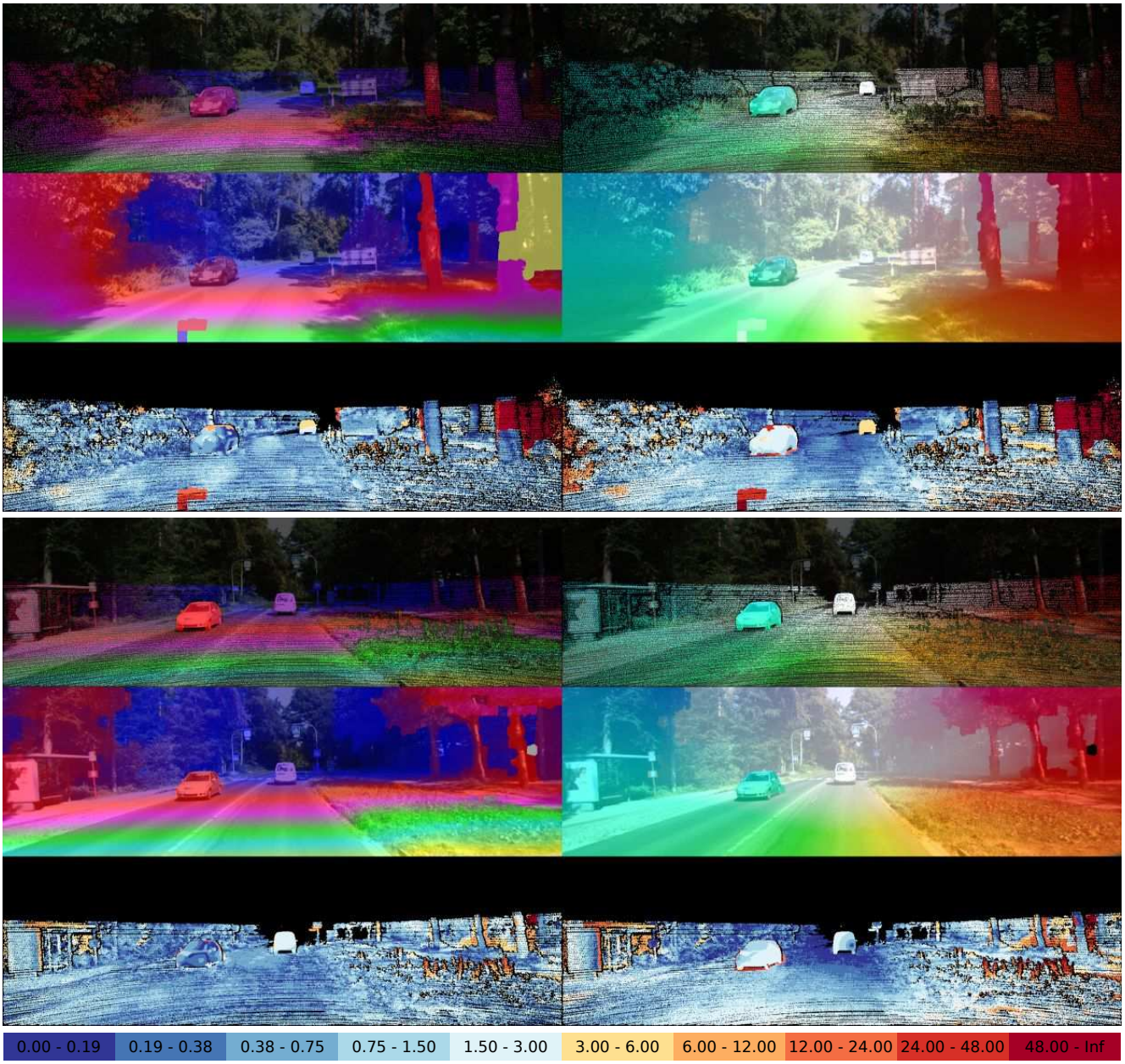


Figure 9: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

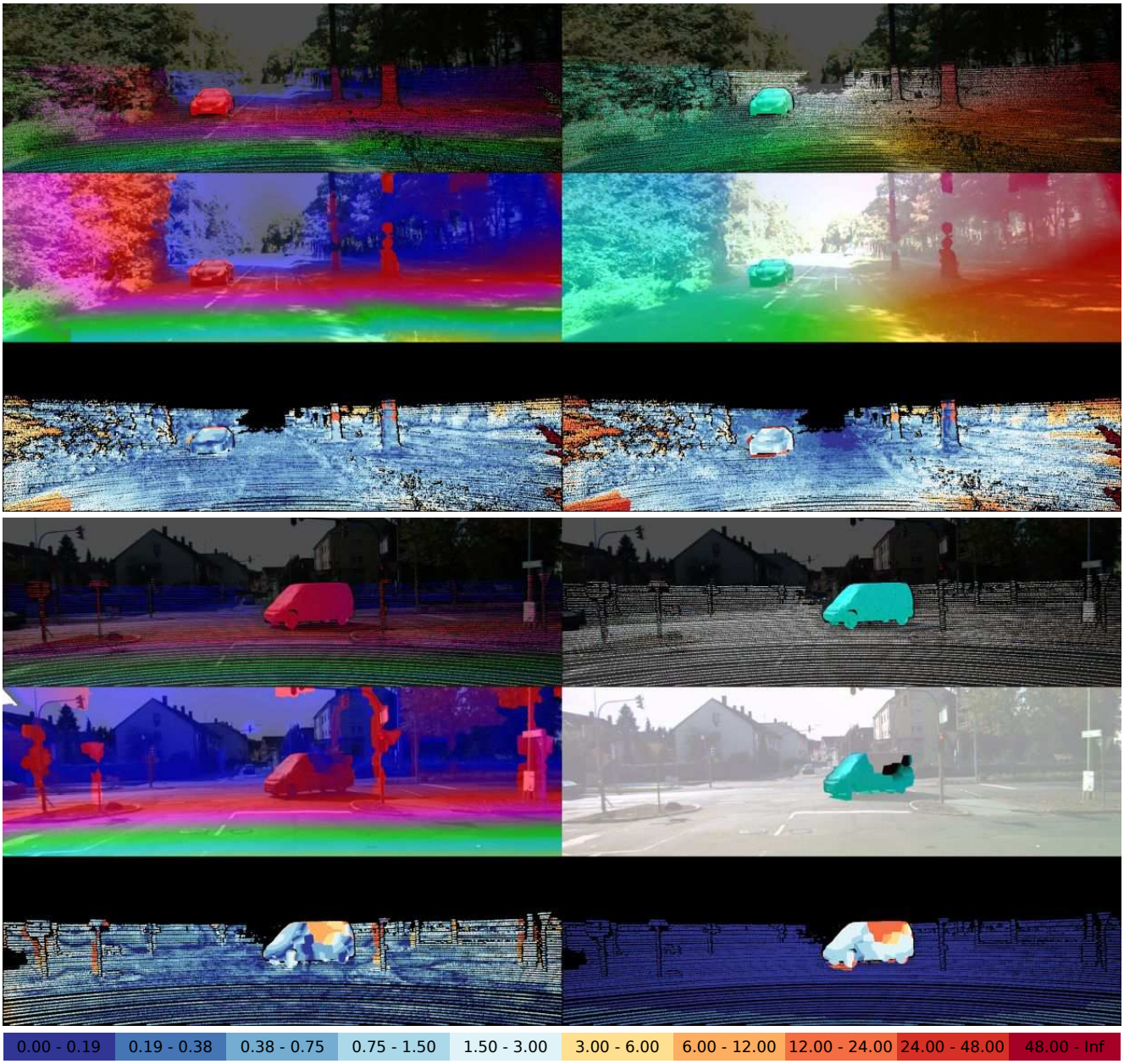


Figure 10: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

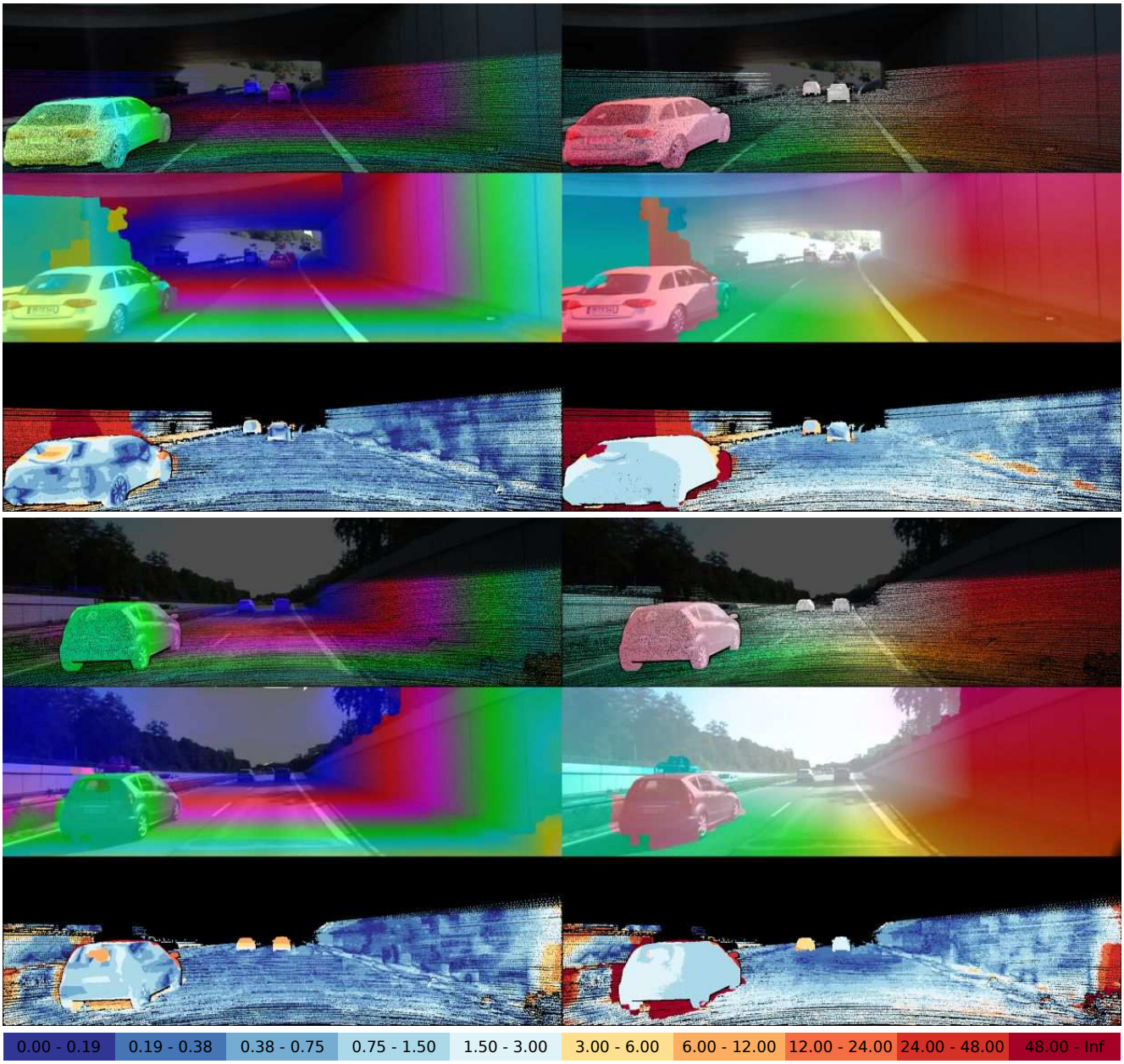


Figure 11: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

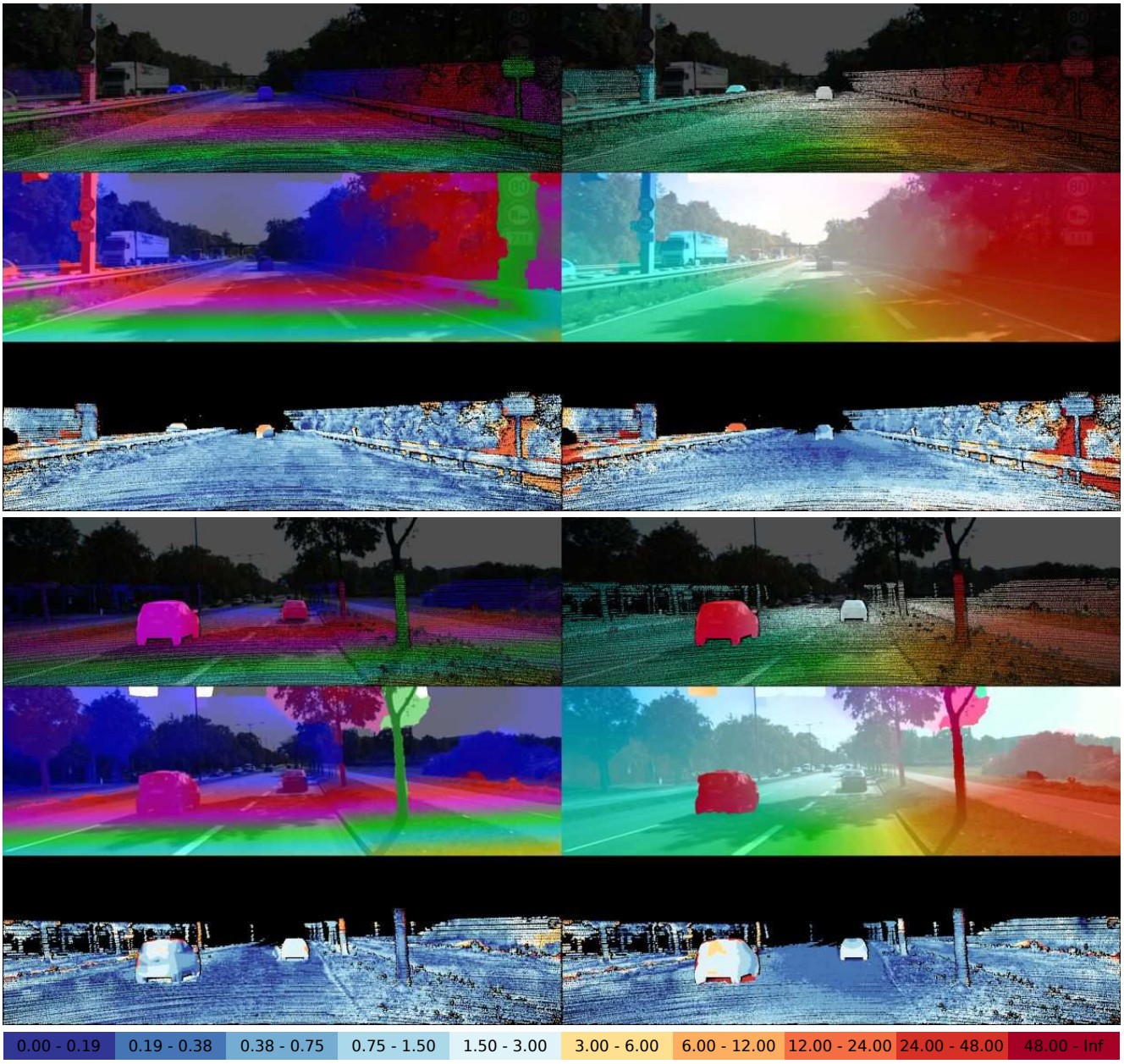


Figure 12: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

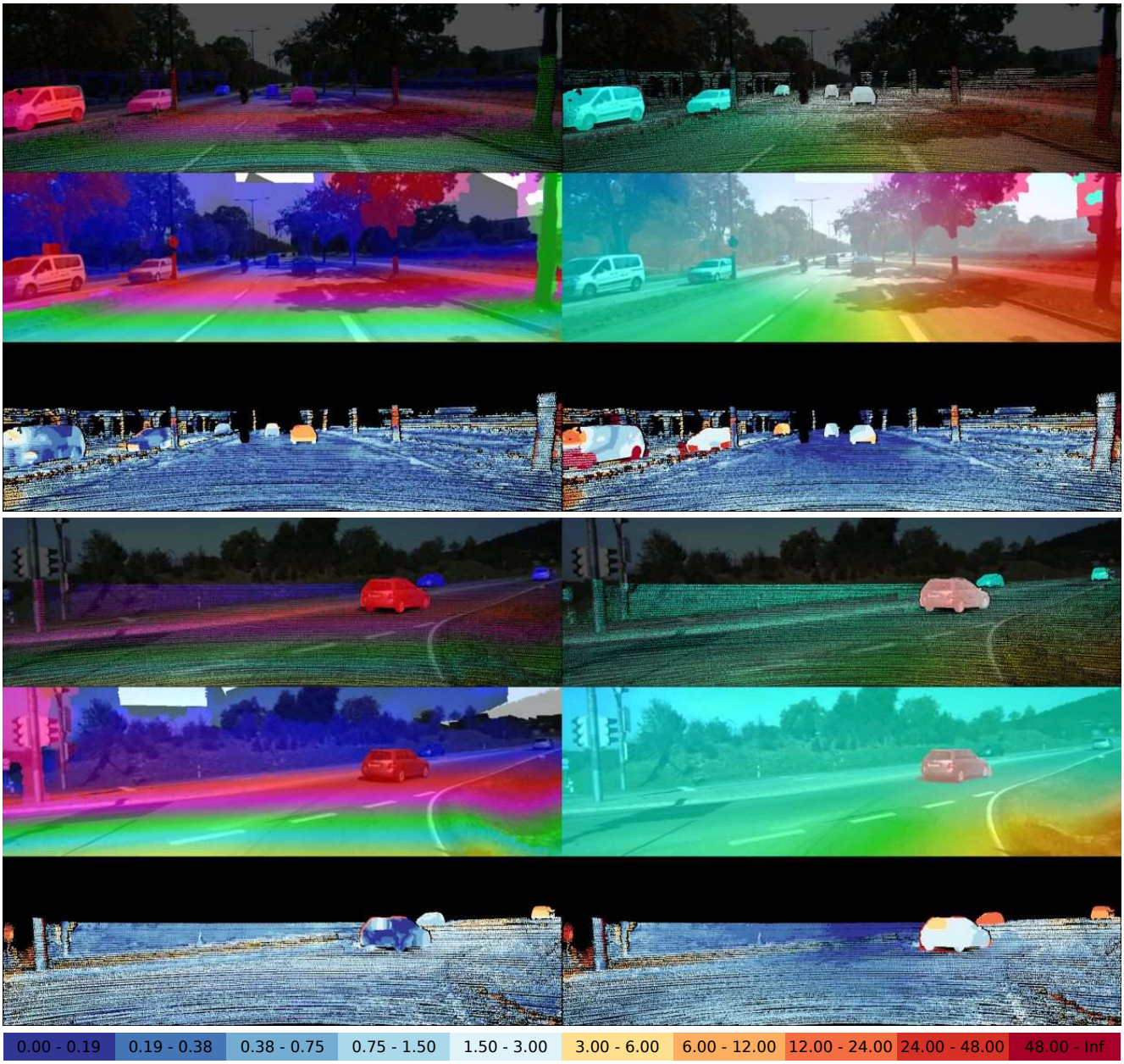


Figure 13: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

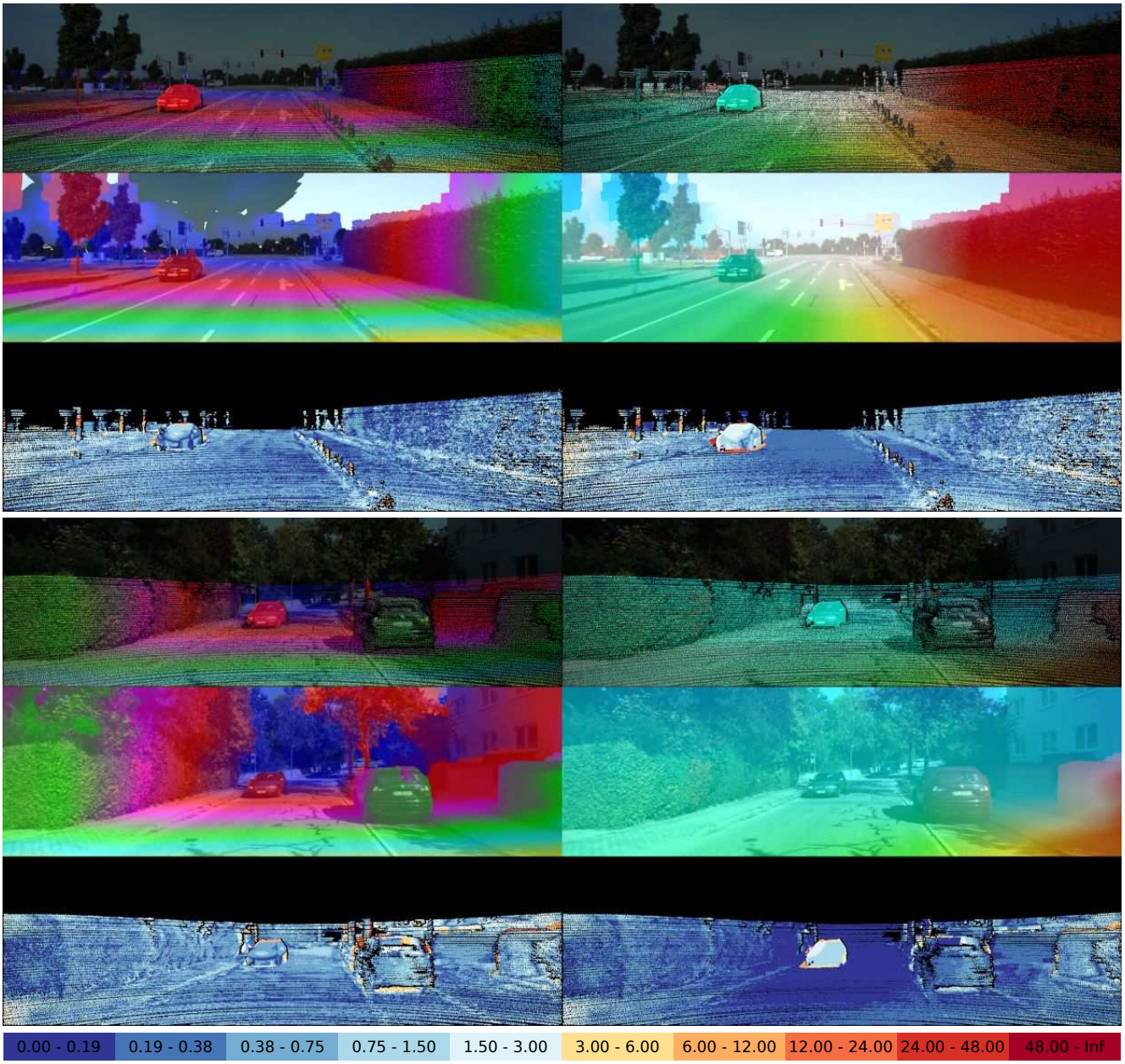


Figure 14: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

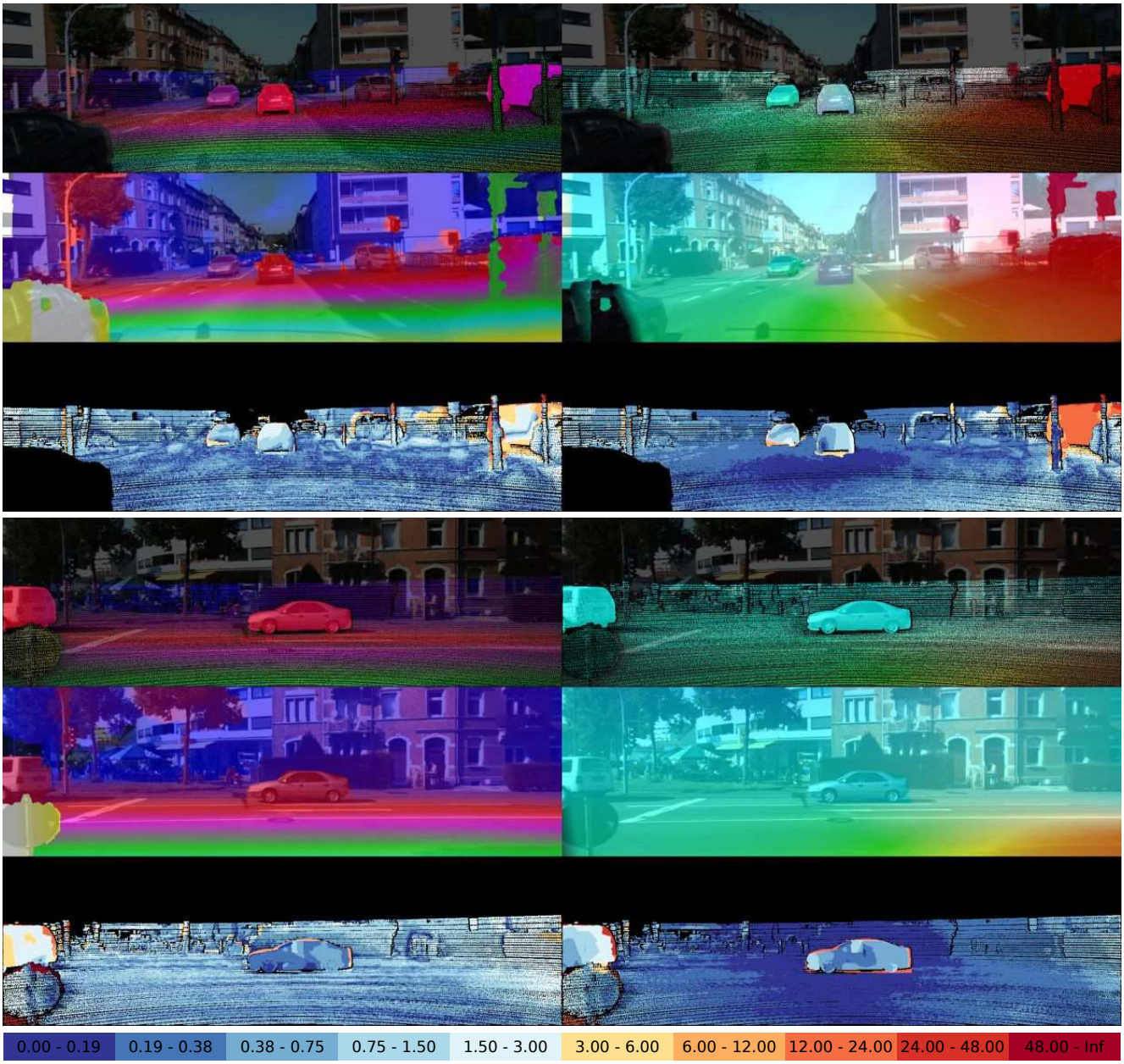


Figure 15: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

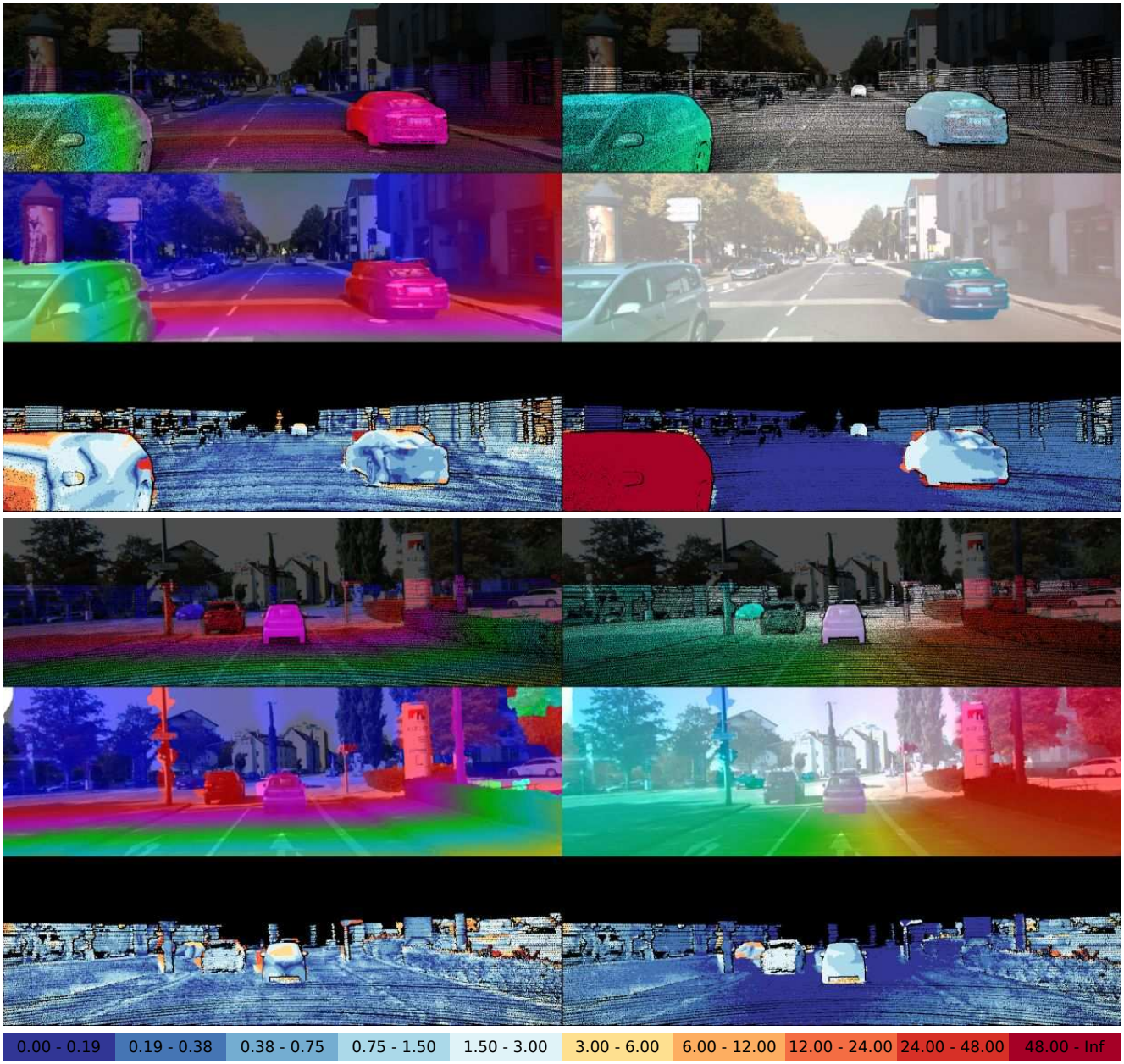


Figure 16: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

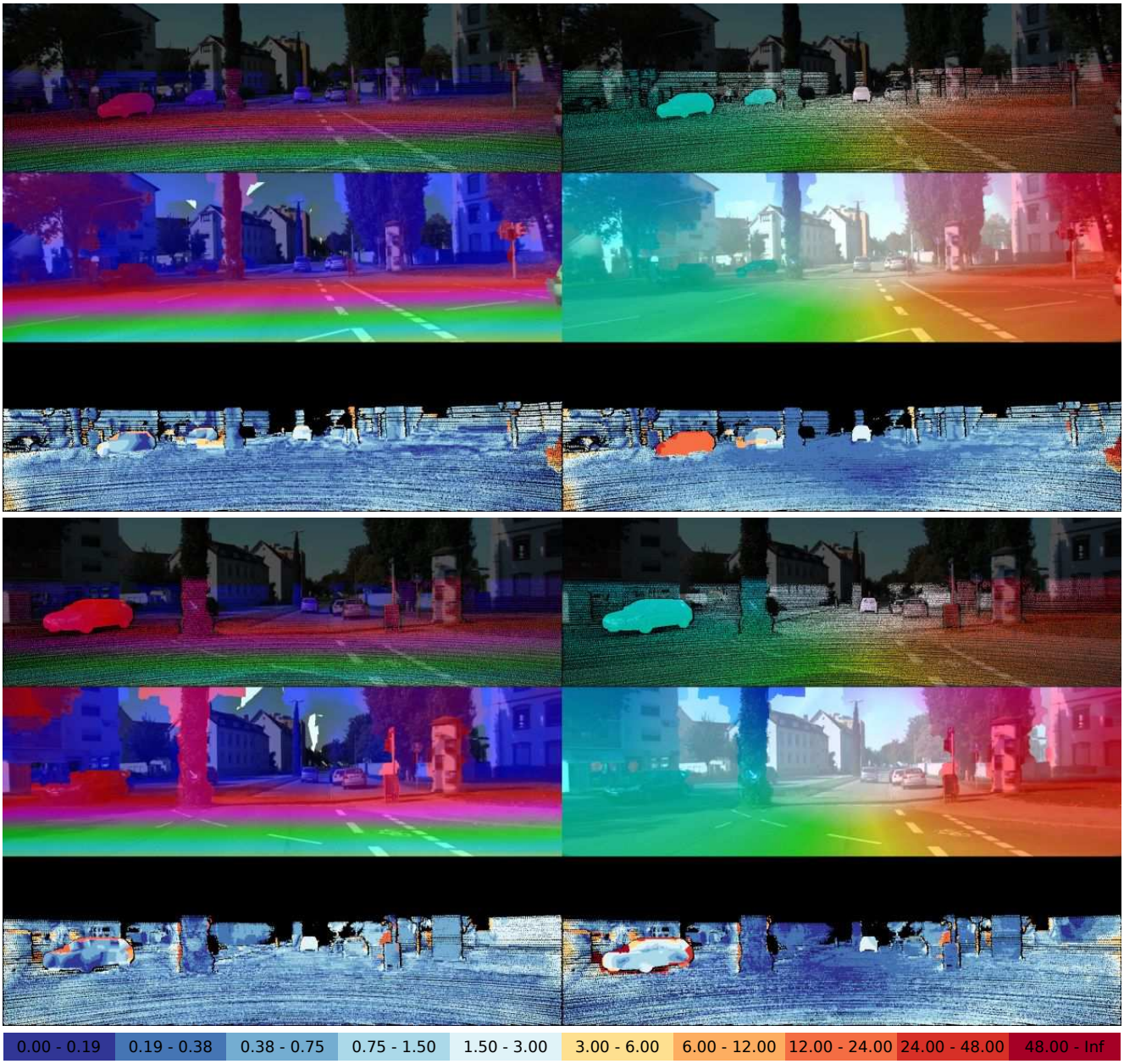


Figure 17: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

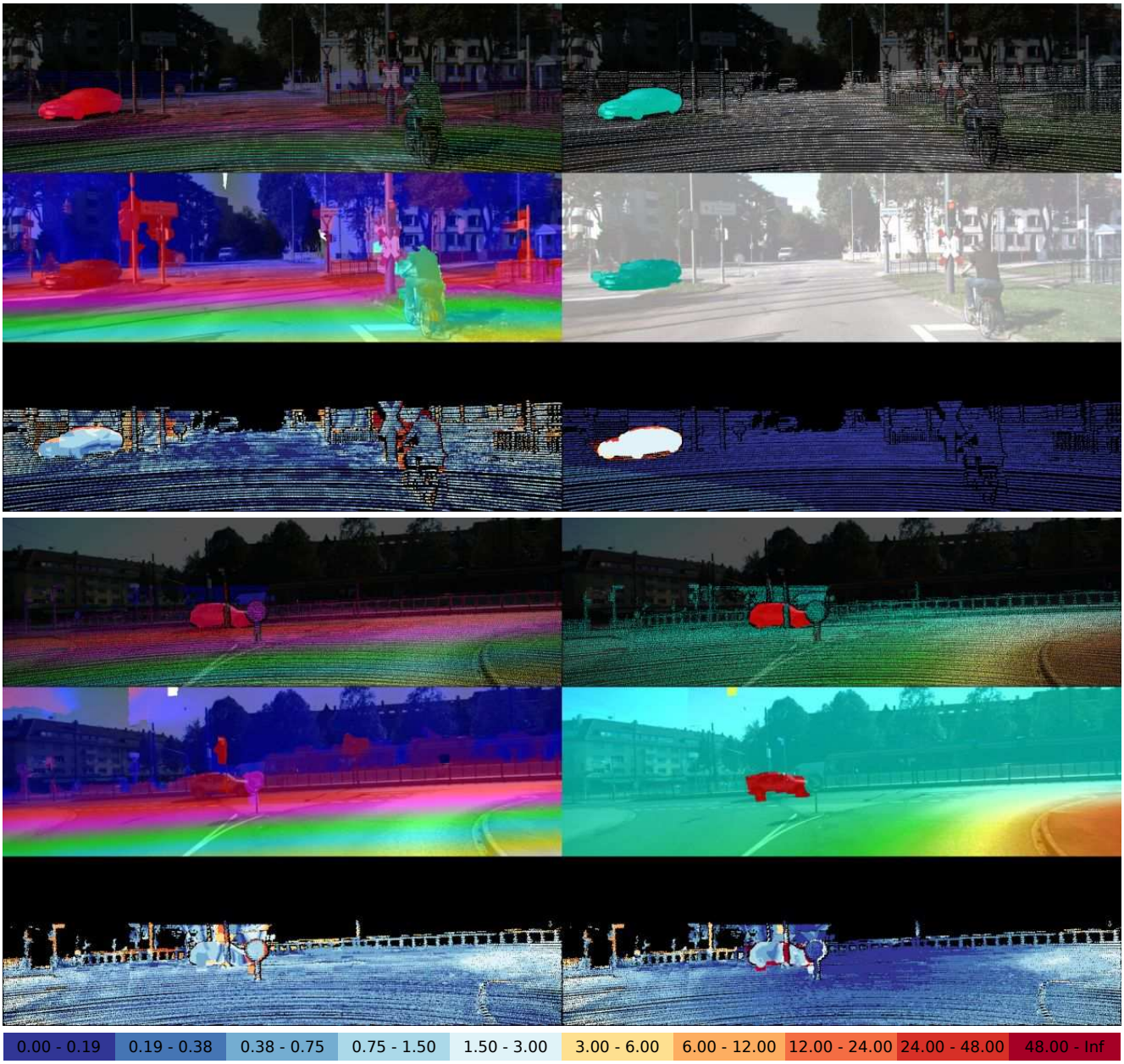


Figure 18: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

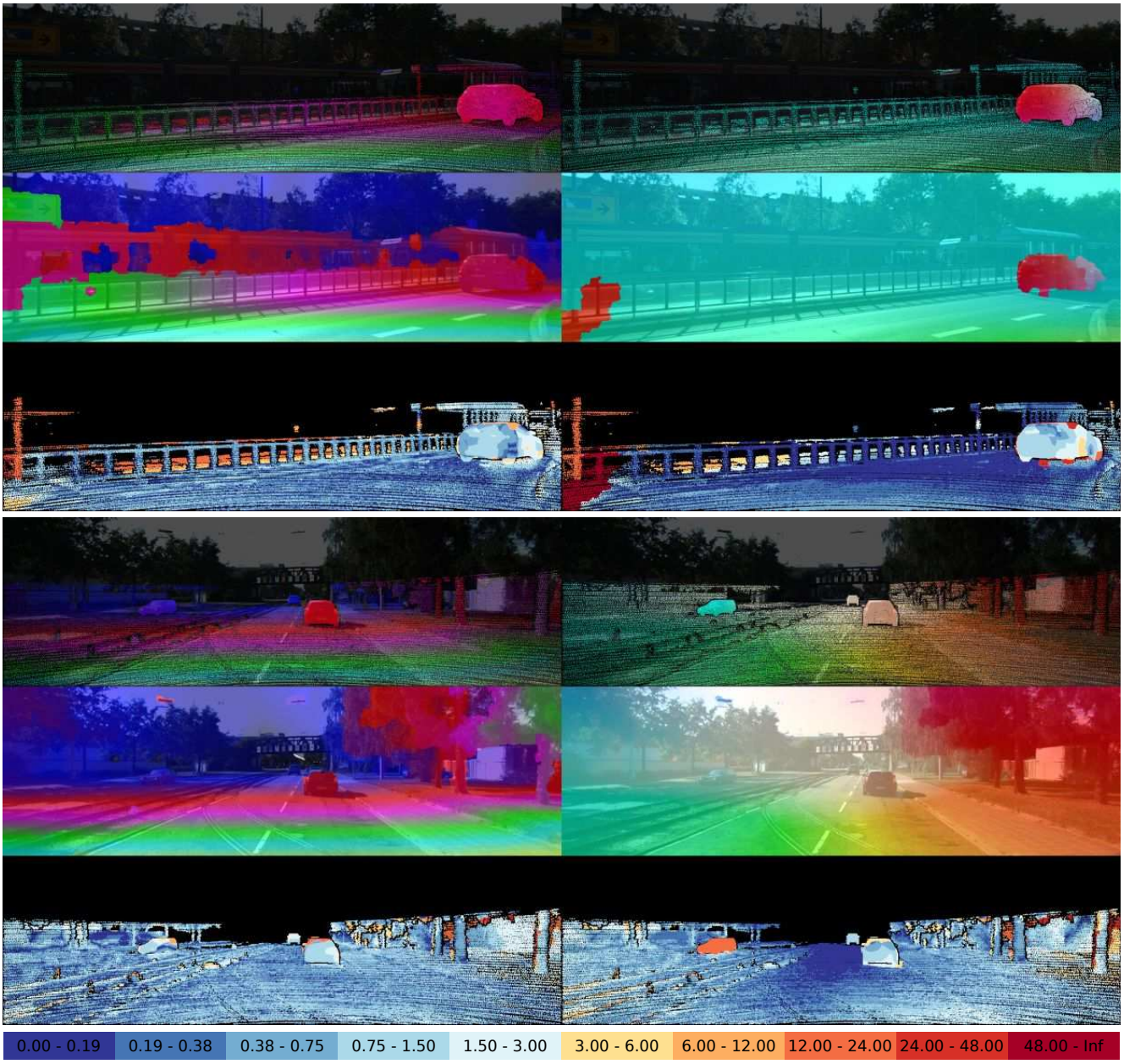


Figure 19: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

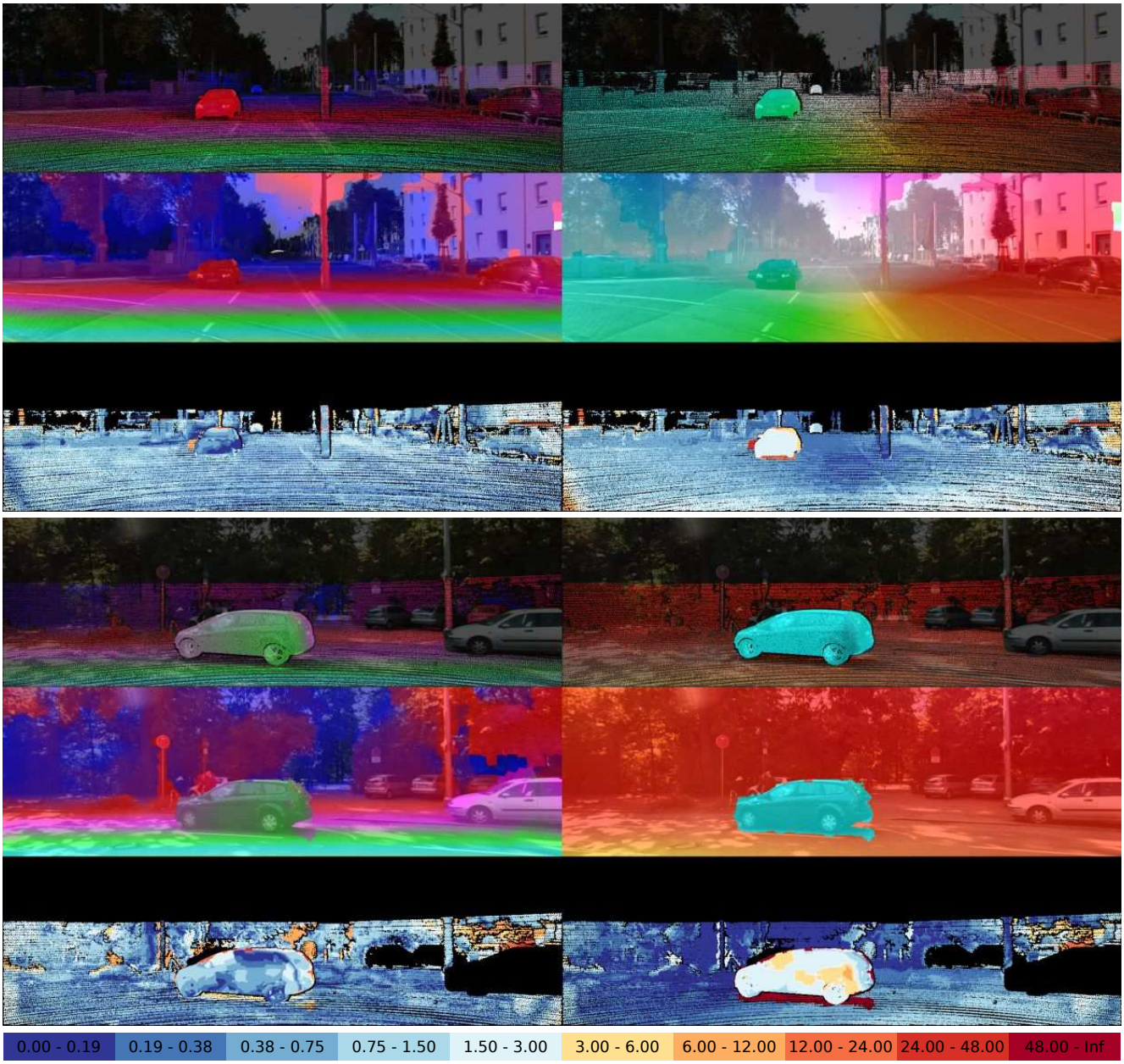


Figure 20: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

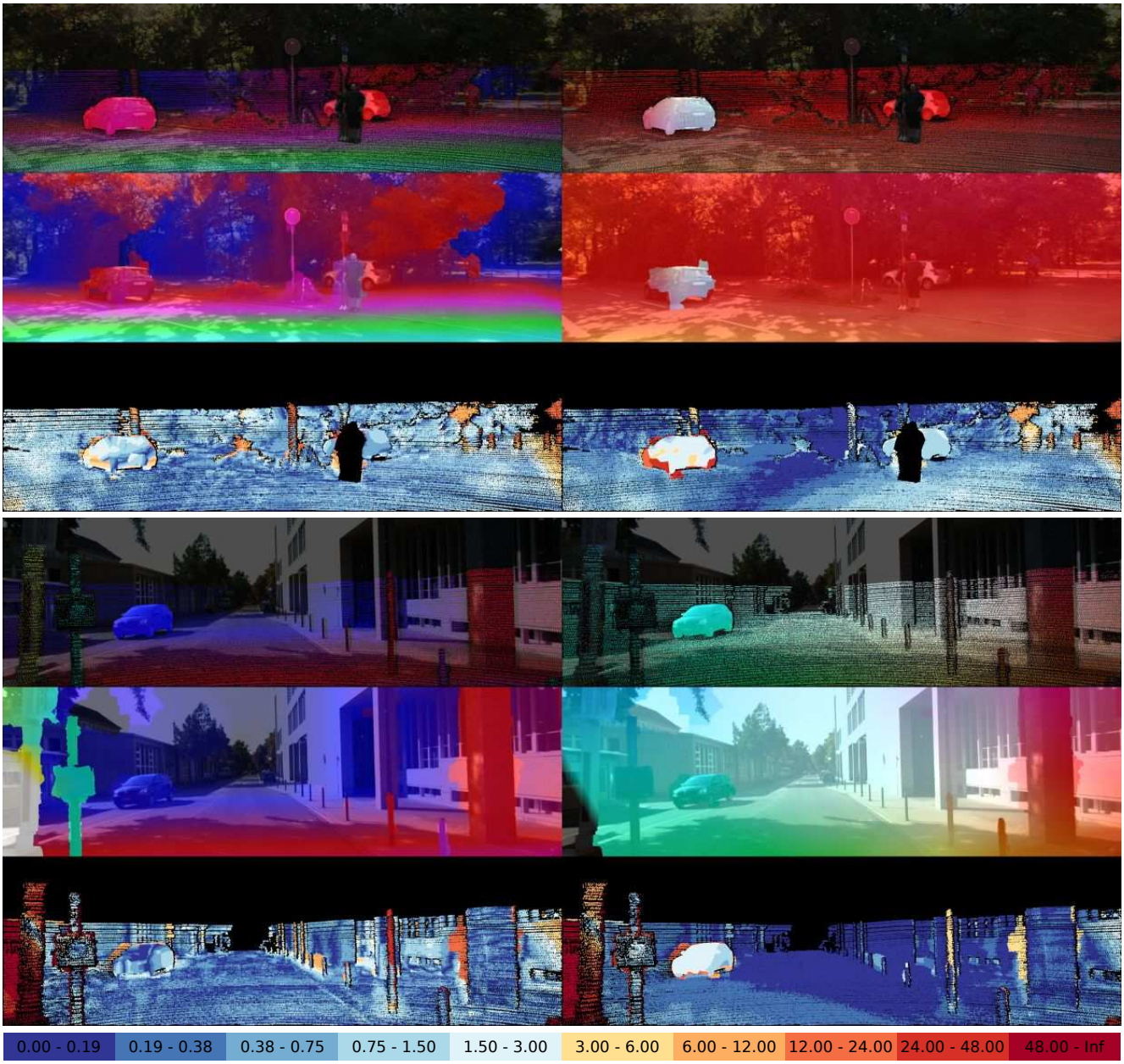


Figure 21: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

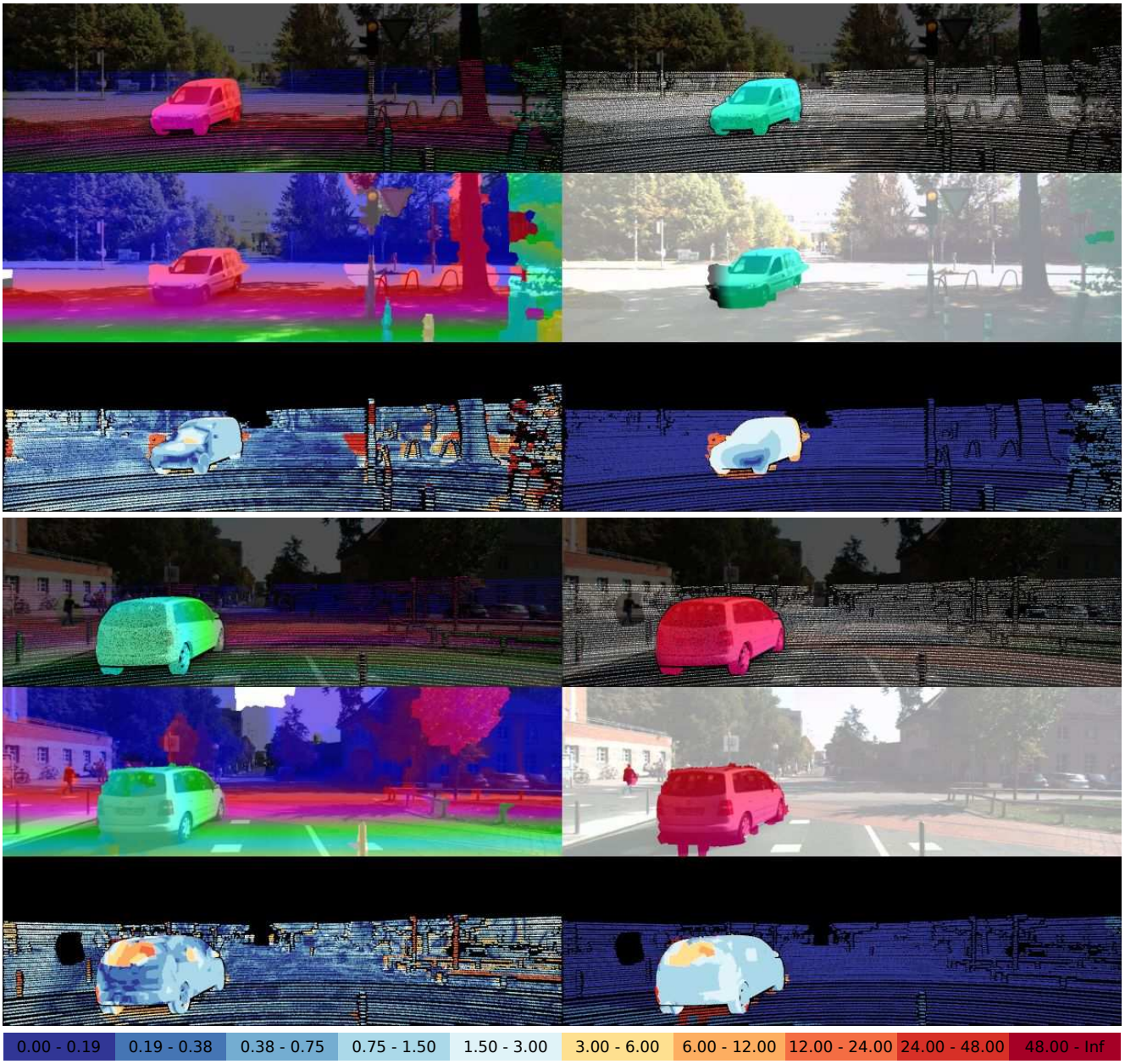


Figure 22: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

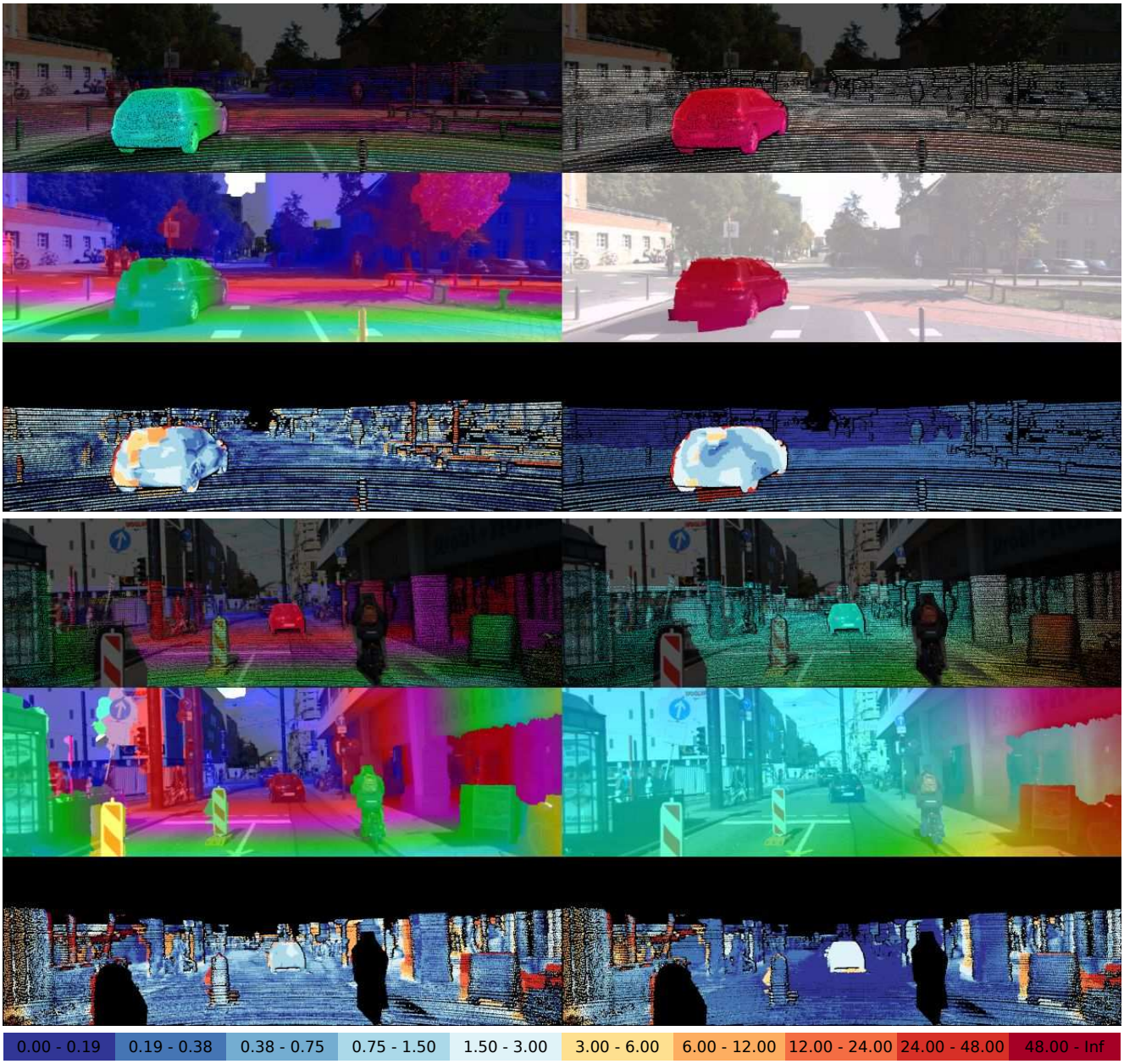


Figure 23: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

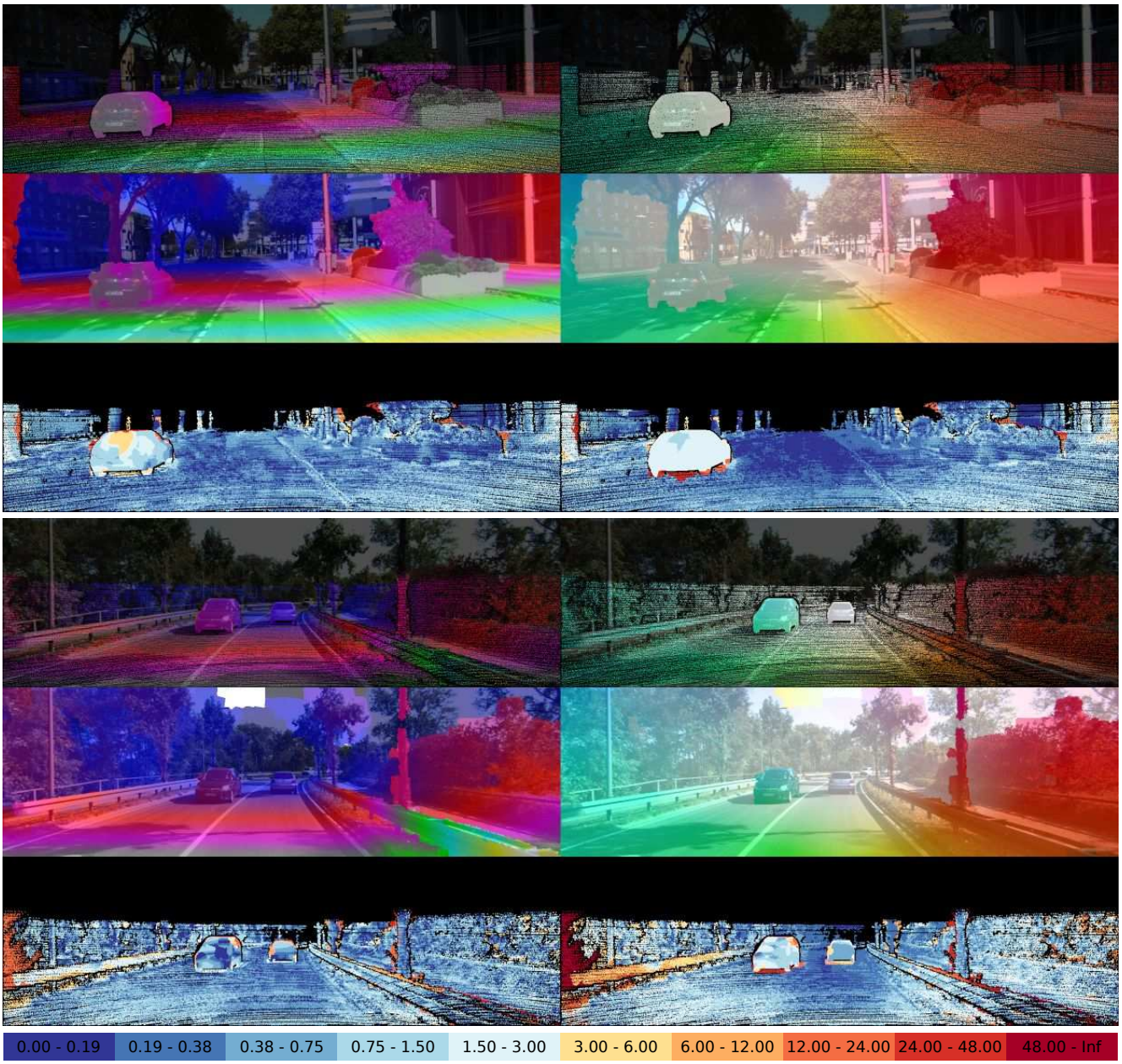


Figure 24: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

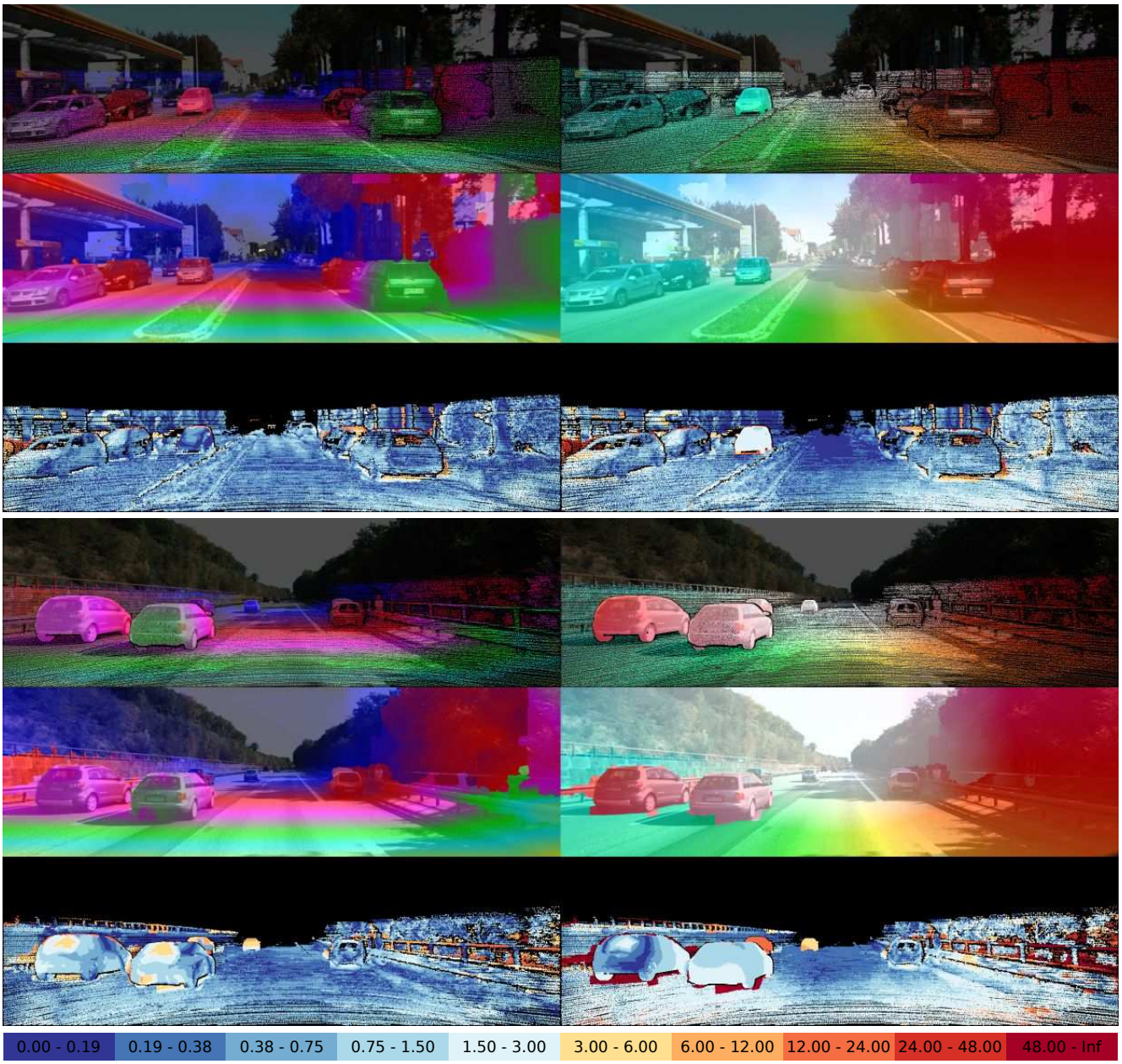


Figure 25: **Qualitiative Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by our scene flow algorithm, and the respective error images using the color scheme depicted in the legend.

9. Qualitative Comparison to State-of-the-Art Methods

Finally, we show a qualitative comparison of our results to those of the state-of-the-art methods [6], [2], [1], [5], and [9]. For each scene, we show the ground truth disparity and optical flow map overlayed with the input images, the estimated disparity and optical flow maps and the corresponding error images using a logarithmic color coding where red shades represent errors above 3 pixels and blue shades denote errors below 3 pixels. The percentage error, which is also included in our evaluation metric, is mapped so that inlier pixels with an error below 3 pixels or 5% are shown in blue shades.

(a) Huguet [6]

(b) GCSF [2]

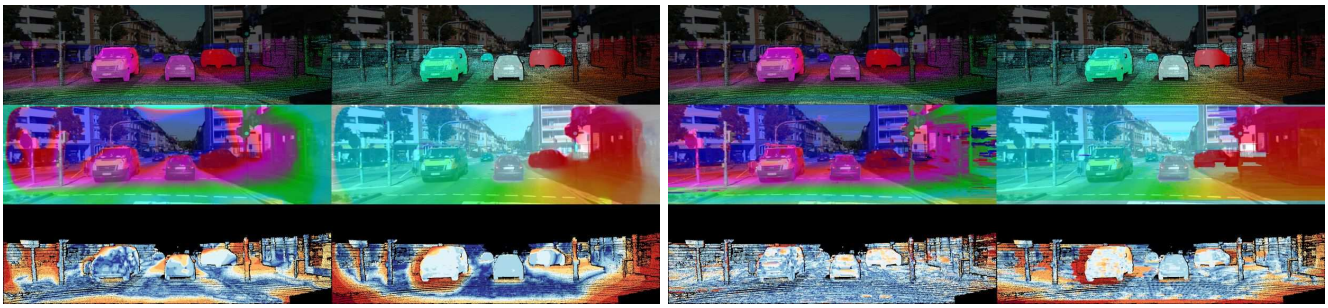
(c) SGM [4] + LDOF [1]

(d) SGM [4] + Sphere Flow [5]

(e) PRSF [9]

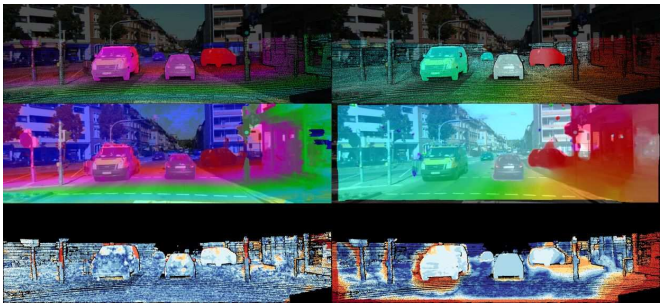
(f) Our full method

Figure 26: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

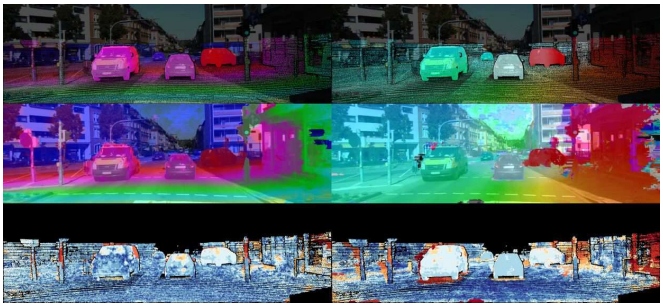


(a) Huguet [6]

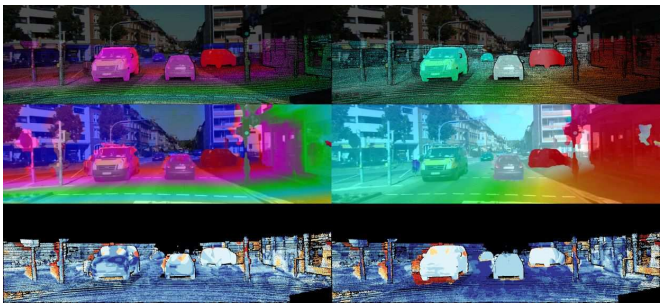
(b) GCSF [2]



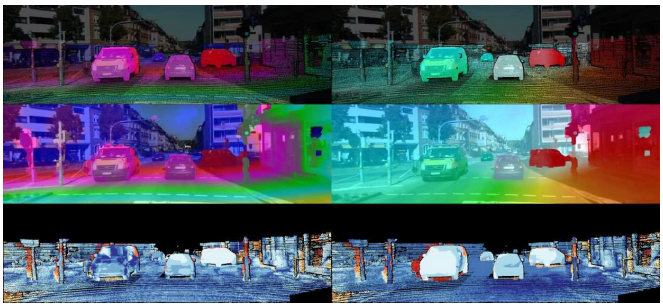
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

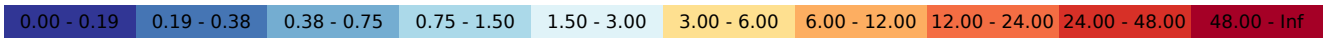
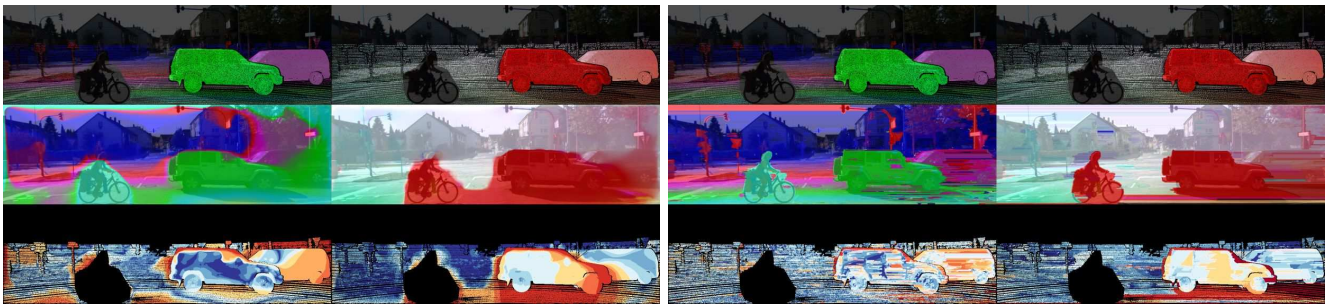
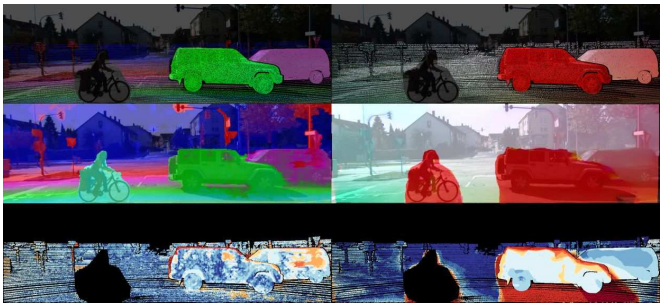


Figure 27: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

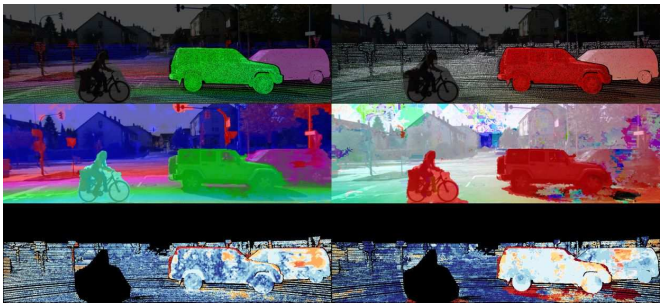


(a) Huguet [6]

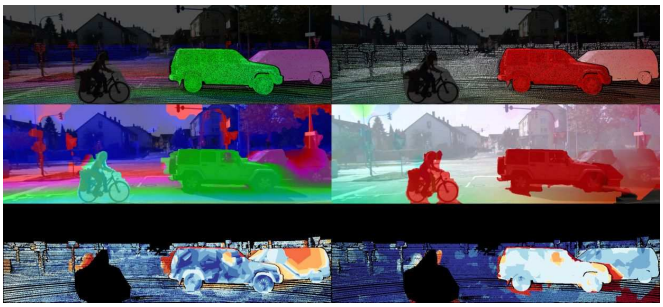
(b) GCSF [2]



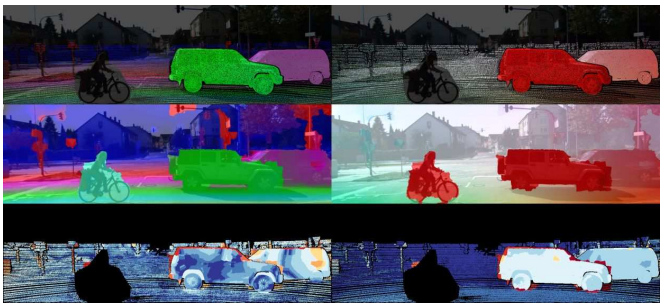
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

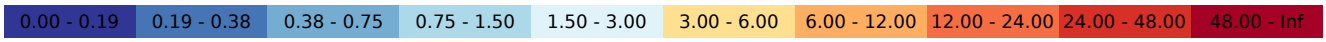
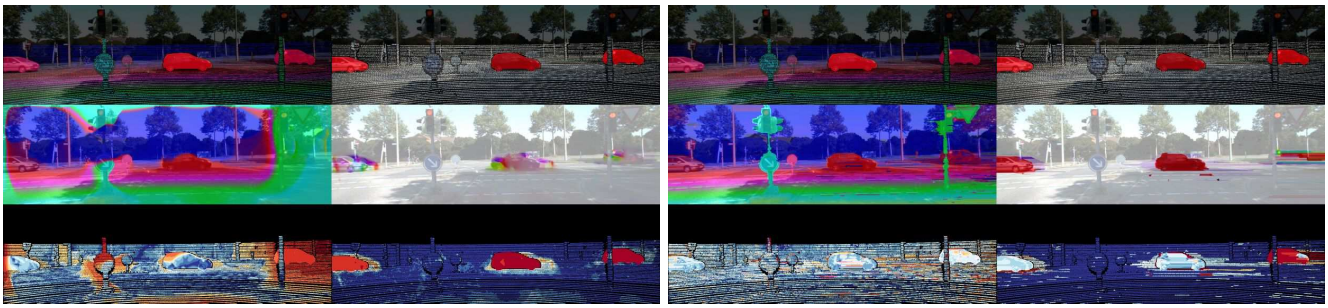
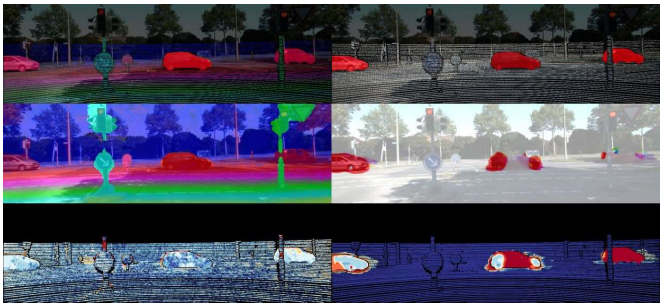


Figure 28: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

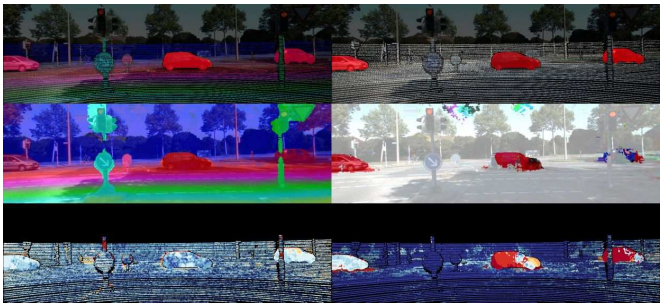


(a) Huguet [6]

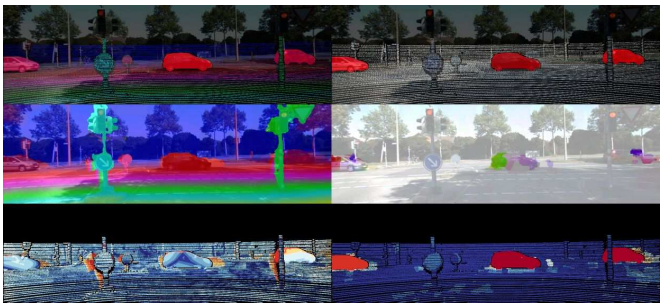
(b) GCSF [2]



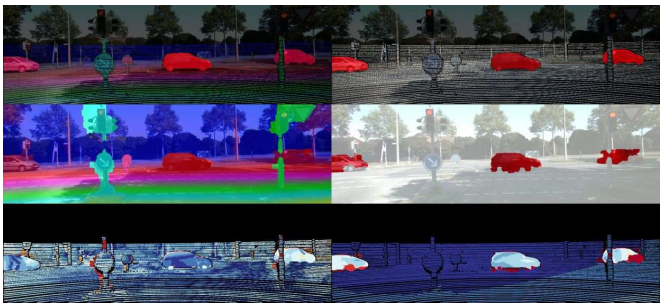
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

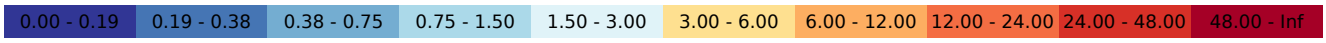
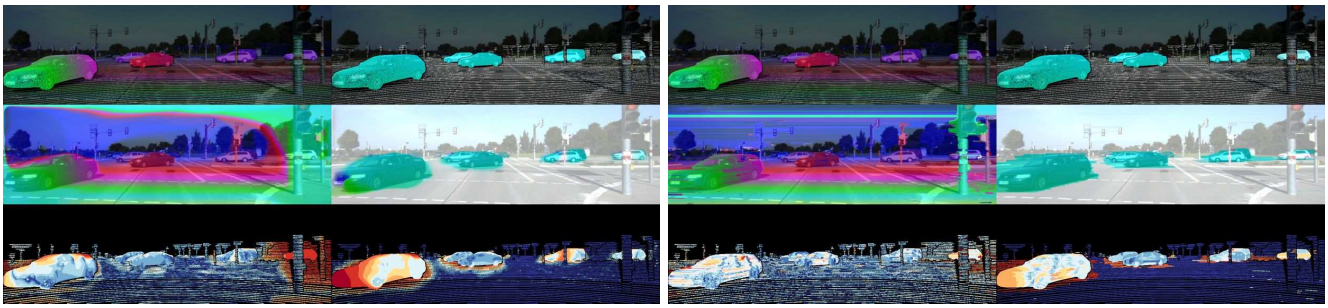
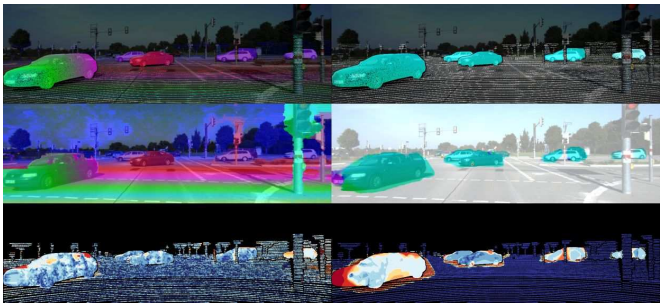


Figure 29: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

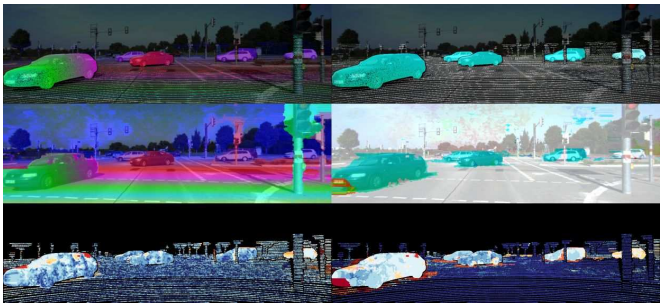


(a) Huguet [6]

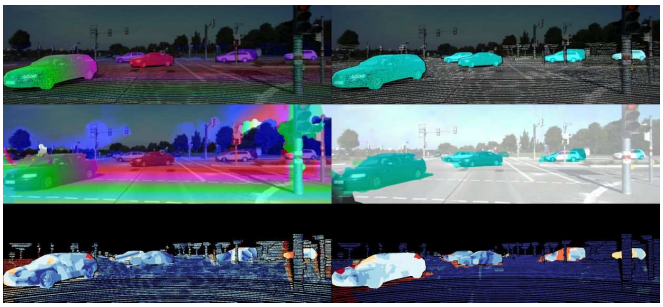
(b) GCSF [2]



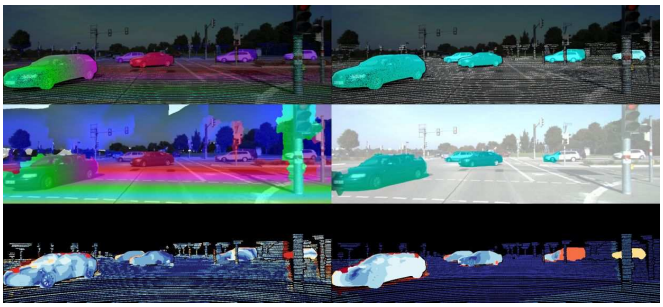
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

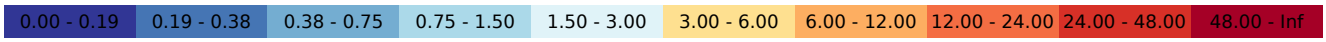
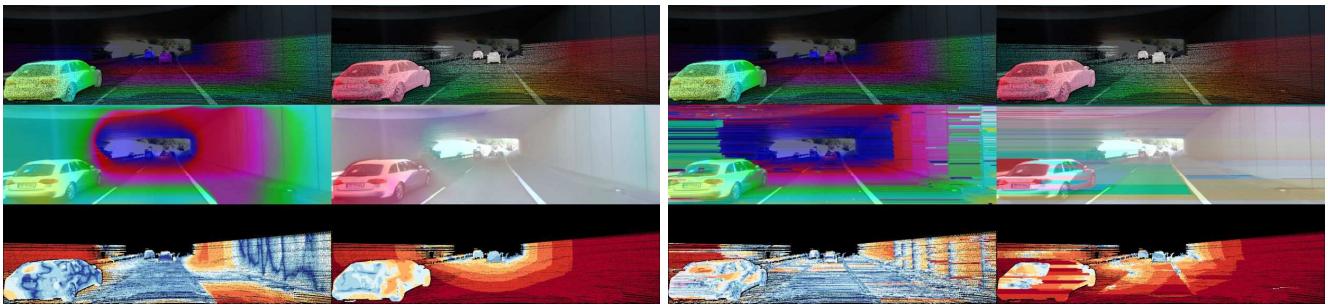
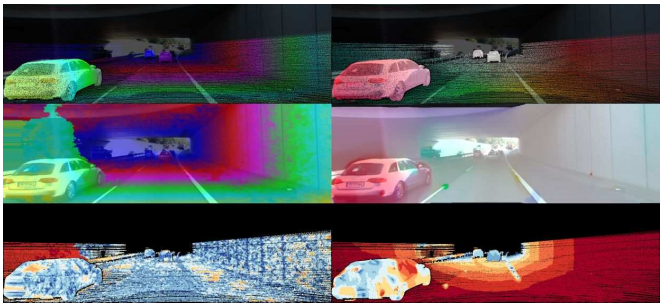


Figure 30: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

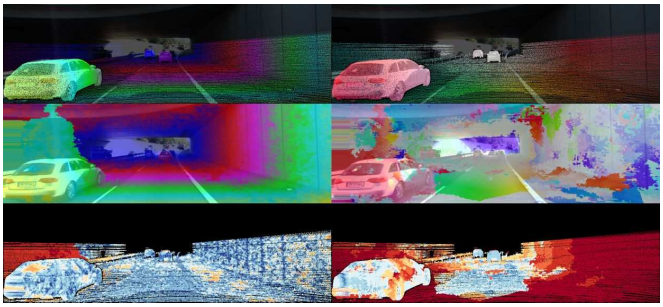


(a) Huguet [6]

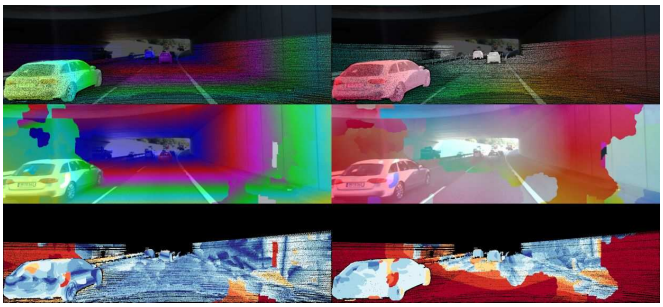
(b) GCSF [2]



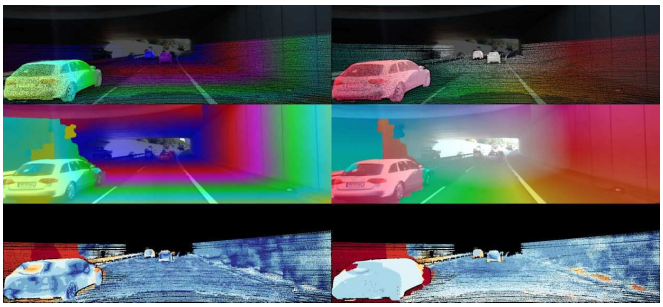
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

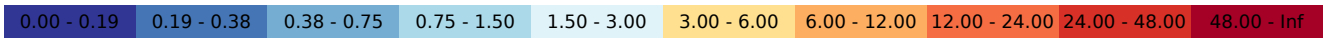
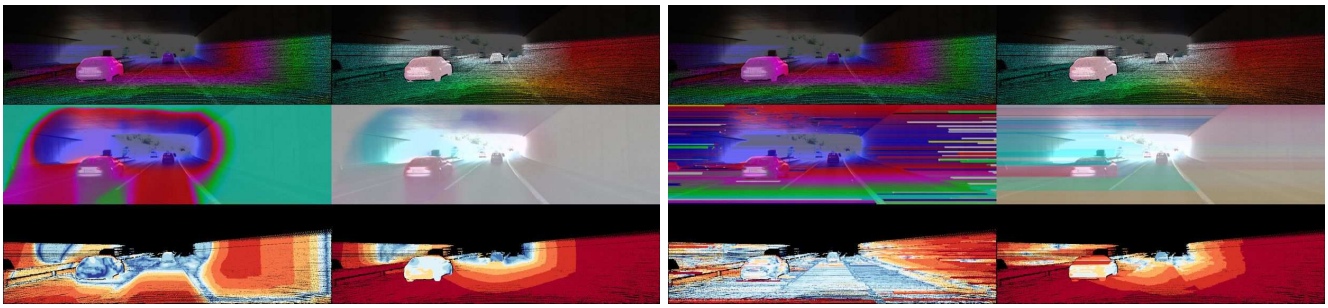
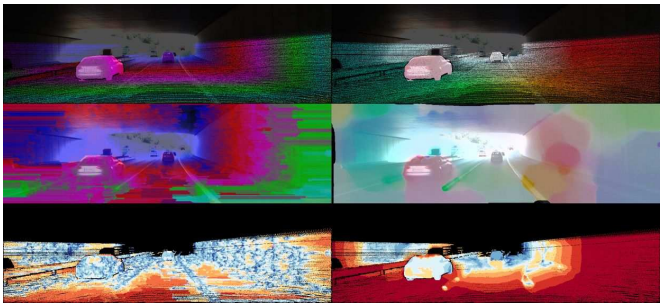


Figure 31: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

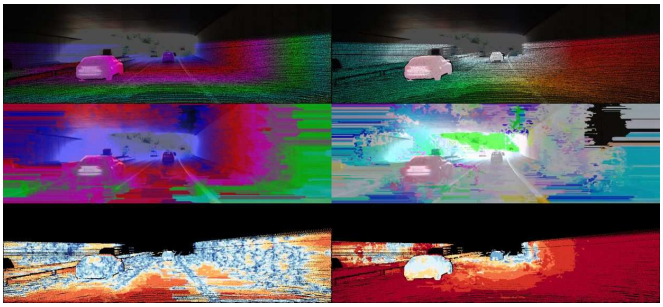


(a) Huguet [6]

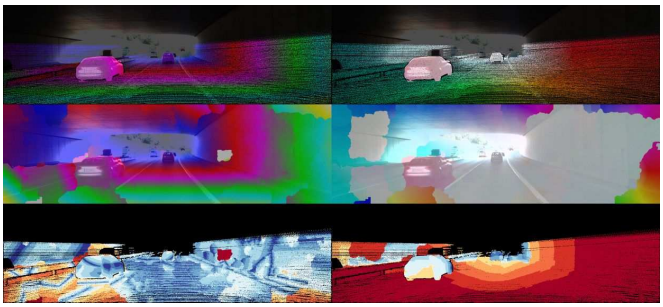
(b) GCSF [2]



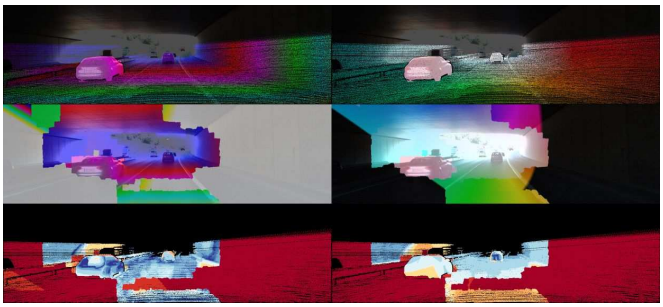
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

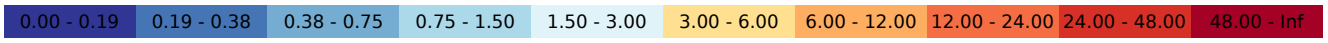
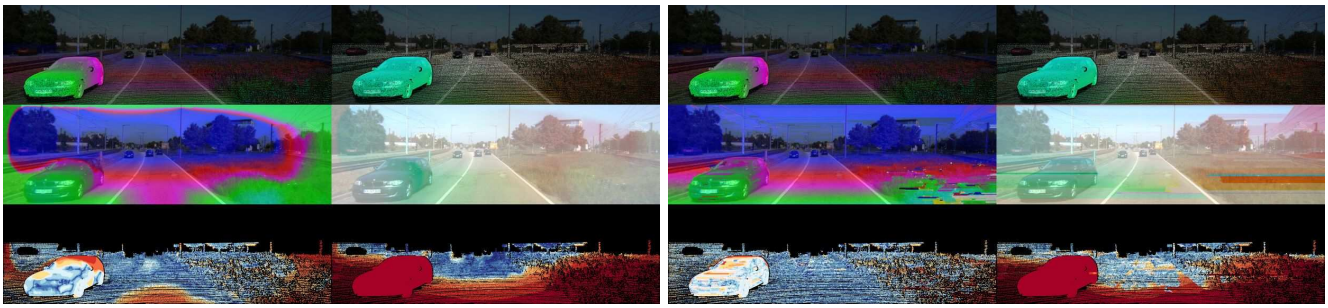
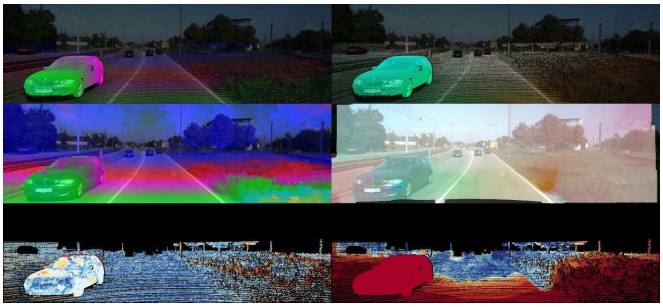


Figure 32: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

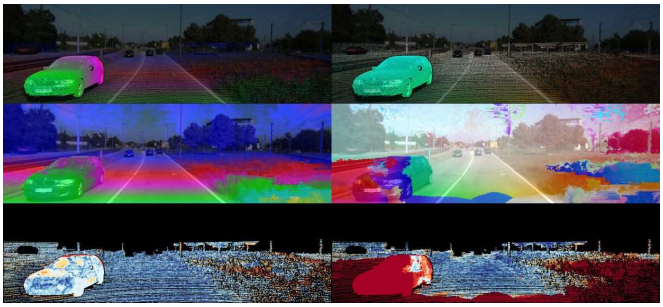


(a) Huguet [6]

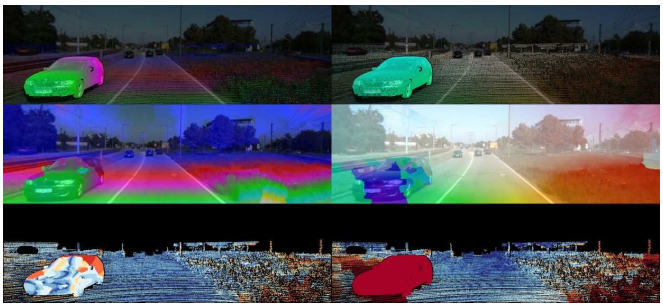
(b) GCSF [2]



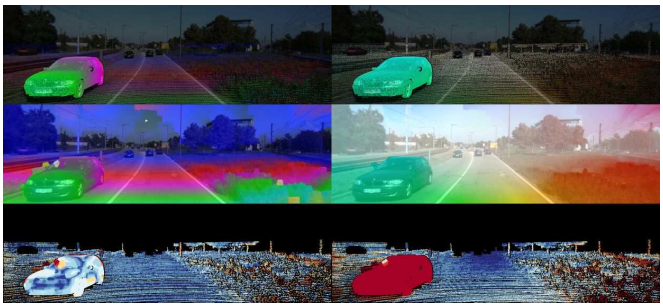
(c) SGM [4] + LDOF [1]



(d) SGM [4] + Sphere Flow [5]



(e) PRSF [9]



(f) Our full method

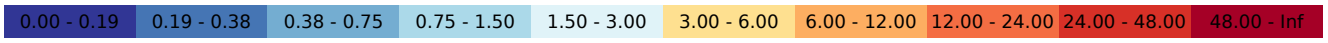


Figure 33: **Comparison of Results.** Each subfigure shows from top-to-bottom: The disparity and optical flow ground truth in the reference view, the disparity map (D1) and optical flow map (Fl) estimated by the specified algorithm, and the respective error images using the color scheme depicted in the legend.

References

[1] T. Brox and J. Malik. Large displacement optical flow: Descriptor matching in variational motion estimation. *PAMI*, 33:500–513, March 2011. 4, 28, 29, 30, 31, 32, 33, 34, 35

[2] J. Cech, J. Sanchez-Riera, and R. P. Horaud. Scene flow estimation by growing correspondence seeds. In *CVPR*, 2011. 4, 28, 29, 30, 31, 32, 33, 34, 35

[3] A. Geiger, J. Ziegler, and C. Stiller. StereoScan: Dense 3D reconstruction in real-time. In *IV*, 2011. 2

[4] H. Hirschmueller. Stereo processing by semiglobal matching and mutual information. *PAMI*, 30(2):328–341, 2008. 4, 28, 29, 30, 31, 32, 33, 34, 35

[5] M. Hornacek, A. Fitzgibbon, and C. Rother. SphereFlow: 6 DoF scene flow from RGB-D pairs. In *CVPR*, 2014. 4, 28, 29, 30, 31, 32, 33, 34, 35

[6] F. Huguet and F. Devernay. A variational method for scene flow estimation from stereo sequences. In *ICCV*, 2007. 4, 5, 28, 29, 30, 31, 32, 33, 34, 35

[7] D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *IJCV*, 106(2):115–137, 2013. 4, 5

[8] L. Valgaerts, A. Bruhn, H. Zimmer, J. Weickert, C. Stoll, and C. Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In *ECCV*, 2010. 5

[9] C. Vogel, K. Schindler, and S. Roth. Piecewise rigid scene flow. In *ICCV*, 2013. 4, 5, 28, 29, 30, 31, 32, 33, 34, 35

[10] A. Wedel, C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers. Efficient dense scene flow from sparse or dense stereo data. In *ECCV*, 2008. 5