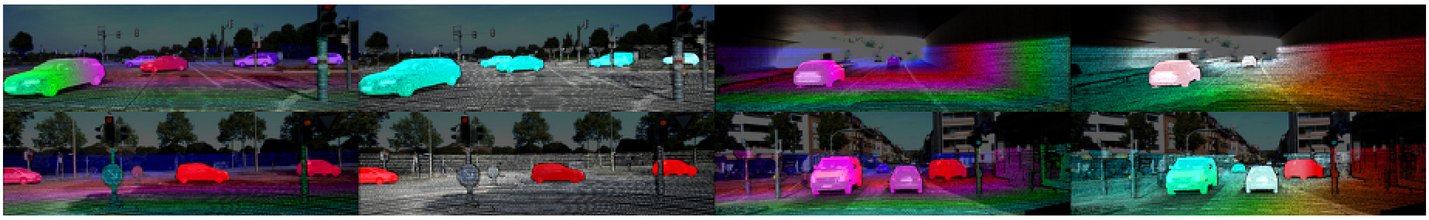
The KITTI Vision Benchmark Suite

A project of Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago

- home
- setup
- stereo
  - Stereo 2012
  - Stereo 2015
- flow
  - Flow 2012
  - Flow 2015
- scene flow
- odometry
- object
  - 2d object
  - 3d object
  - bird's eye view
- tracking
- road
- semantics
- raw data
- submit results

Andreas Geiger (MPI Tübingen) | Philip Lenz (KIT) | Christoph Stiller (KIT) | Raquel Urtasun (University of Toronto)

Stereo Evaluation 2015



The **stereo 2015** / **flow 2015** / **scene flow 2015** benchmark consists of 200 training scenes and 200 test scenes (4 color images per scene, saved in loss less png format). Compared to the stereo 2012 and flow 2012 benchmarks, it comprises dynamic scenes for which the ground truth has been established in a semi-automatic process. Our evaluation server computes the percentage of bad pixels averaged over all ground truth pixels of all 200 test images. For this benchmark, we consider a pixel to be correctly estimated if the disparity or flow end-point error is **<3px or <5%** (for scene flow this criterion needs to be fulfilled for both disparity maps and the flow map). We require that all methods use the same parameter set for all test pairs. Our development kit provides details about the data format as well as MATLAB / C++ utility functions for reading and writing disparity maps and flow fields. More details can be found in Object Scene Flow for Autonomous Vehicles (CVPR 2015).

- Download stereo 2015/flow 2015/scene flow 2015 data set (2 GB)
- Download calibration files (1 MB)
- Download multi-view extension (20 frames per scene) (14 GB)
- Download development kit (3 MB)

Our evaluation table ranks all methods according to the number of erroneous pixels. All methods providing less than 100 % density have been interpolated using simple background interpolation as explained in the corresponding header file in the development kit. Legend:

- **D1:** Percentage of stereo disparity outliers in first frame
- **D2:** Percentage of stereo disparity outliers in second frame
- **Fl:** Percentage of optical flow outliers
- **SF:** Percentage of scene flow outliers (=outliers in either D0, D1 or Fl)
- **bg:** Percentage of outliers averaged only over background regions
- **fg:** Percentage of outliers averaged only over foreground regions
- **all:** Percentage of outliers averaged over all ground truth pixels

Note: On 13.03.2017 we have fixed several small errors in the flow (noc+occ) ground truth of the dynamic foreground objects and manually verified all images for correctness by warping them according to the ground truth. As a consequence, all error numbers have decreased slightly. Please download the devkit and the annotations with the improved ground truth for the training set again if you have downloaded the files prior to 13.03.2017 and consider reporting these new number in all future publications. The last leaderboards before these corrections can be found here (optical flow 2015) and here (scene flow 2015). The leaderboards for the KITTI 2015 stereo benchmarks did not change.

Additional information used by the methods

- Flow: Method uses optical flow (2 temporally adjacent images)
- Multiview: Method uses more than 2 temporally adjacent images
- Motion stereo: Method uses epipolar geometry for computing optical flow
- Additional training data: Use of additional data sources for training (see details)

Evaluation ground truth All pixels **Evaluation area** All pixels

	Method	Setting	Code	D1-bg	D1-fg	D1-all	Density	Runtime	Environment	Compare
1	CRL			2.48 %	**3.59 %**	**2.67 %**	100.00 %	0.47 s	Nvidia GTX 1080	

J. Pang, W. Sun, J. Ren, C. Yang and Y. Qiong: Cascade Residual Learning: A Two-stage Convolutional Neural Network for Stereo Matching. arXiv preprint arXiv:1708.09204 2017.

2	GC-NET			**2.21 %**	6.16 %	2.87 %	100.00 %	0.9 s	Nvidia GTX Titan X	

A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach and A. Bry: End-to-End Learning of Geometry and Context for Deep Stereo Regression. arXiv preprint arxiv:1703.04309 2017.

3	DRR			2.58 %	6.04 %	3.16 %	100.00 %	0.4 s	Nvidia GTX Titan X	

S. Gidaris and N. Komodakis: Detect, Replace, Refine: Deep Structured Prediction For Pixel Wise Labeling. arXiv preprint arXiv:1612.04770 2016.

4	L-ResMatch		code	2.72 %	6.95 %	3.42 %	100.00 %	48 s	1 core @ 2.5 Ghz (C/C++)	

A. Shaked and L. Wolf: Improved Stereo Matching with Constant Highway Networks and Reflective Loss. arXiv preprint arxiv:1701.00165 2016.

5	Displets v2		code	3.00 %	5.56 %	3.43 %	100.00 %	265 s	>8 cores @ 3.0 Ghz (Matlab + C/C++)	

F. Guney and A. Geiger: Displets: Resolving Stereo Ambiguities using Object Knowledge. Conference on Computer Vision and Pattern Recognition (CVPR) 2015.

6	D3DNet			2.88 %	6.60 %	3.50 %	100.00 %	0.35 s	Nvidia GTX Titan X	
7	SsSMnet			2.86 %	7.12 %	3.57 %	100.00 %	0.8 s	Titan Xp	

Y. Zhong, Y. Dai and H. Li: Self-Supervised Learning for Stereo Matching with Self-Improving Ability. arXiv:1709.00930 2017.

8	CNNF+SGM			2.78 %	7.69 %	3.60 %	100.00 %	71 s	TESLA K40C	
9	PBCP			2.58 %	8.74 %	3.61 %	100.00 %	68 s	Nvidia GTX Titan X	

A. Seki and M. Pollefeys: Patch Based Confidence Prediction for Dense Disparity Map. British Machine Vision Conference (BMVC) 2016.

10	SGM-Net			2.66 %	8.64 %	3.66 %	100.00 %	67 s	Titan X	

A. Seki and M. Pollefeys: SGM-Nets: Semi-Global Matching With Neural Networks. CVPR 2017.

11	MRDF-CNN			3.55 %	4.81 %	3.76 %	100.00 %	0.62 s	Nvidia GTX Titan X	
12	MC-CNN-acrt		code	2.89 %	8.88 %	3.89 %	100.00 %	67 s	Nvidia GTX Titan X (CUDA, Lua/Torch7)	

J. Zbontar and Y. LeCun: Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. Submitted to JMLR .

13	ASTCC			2.94 %	8.95 %	3.94 %	100.00 %	130 s	GPU @ 2.5 Ghz (Python + C/C++)	
14	CNN-SPS			3.30 %	7.92 %	4.07 %	100.00 %	80 s	GPU @ 2.5 Ghz (C/C++)	

L. Chen, J. Chen and L. Fan: A Convolutional Neural Networks based Full Density Stereo Matching Framework. .

15	RGL			4.22 %	4.02 %	4.19 %	100.00 %	0.1 s	1 core @ 2.5 Ghz (C/C++)	
16	GPNET			4.23 %	4.05 %	4.20 %	100.00 %	0.1 s	1 core @ 2.5 Ghz (C/C++)	
17	PRSM	Flow, Multiview	code	3.02 %	10.52 %	4.27 %	99.99 %	300 s	1 core @ 2.5 Ghz (C/C++)	

C. Vogel, K. Schindler and S. Roth: 3D Scene Flow Estimation with a Piecewise Rigid Scene Model. ijcv 2015.

18	CAL			4.33 %	4.13 %	4.30 %	100.00 %	0.1 s	Nvidia GTX Titan X	
19	DispNetC		code	4.32 %	4.41 %	4.34 %	100.00 %	0.06 s	Nvidia GTX Titan X (Caffe)	

N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy and T. Brox: A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. CVPR 2016.

20	SSF	Flow		3.55 %	8.75 %	4.42 %	100.00 %	5 min	1 core @ 2.5 Ghz (Matlab + C/C++)	
21	CGNet			4.39 %	4.59 %	4.43 %	100.00 %	2.3 s	1 core @ 2.5 Ghz (Matlab + C/C++)	
22	ISF	Flow		4.12 %	6.17 %	4.46 %	100.00 %	10 min	1 core @ 3 Ghz (C/C++)	

A. Behl, O. Jafari, S. Mustikovela, H. Alhaija, C. Rother and A. Geiger: Bounding Boxes, Segmentations and Object Coordinates: How Important is Recognition for 3D Scene Flow Estimation in Autonomous Driving Scenarios?. International Conference on Computer Vision (ICCV) 2017.

23	Content-CNN			3.73 %	8.58 %	4.54 %	100.00 %	1 s	Nvidia GTX Titan X (Torch)	

W. Luo, A. Schwing and R. Urtasun: Efficient Deep Learning for Stereo Matching. CVPR 2016.

24	FCVF-net			3.96 %	7.74 %	4.59 %	100.00 %	0.5 s	GPU @ 1.5 Ghz (Python)
25	MCSC			3.61 %	10.13 %	4.69 %	100.00 %	1 s	Nvidia GTX 1080 (Caffe)
26	MC-CNN-WS			3.78 %	10.93 %	4.97 %	100.00 %	1.35 s	1 core 2.5 Ghz + K40 NVIDIA, Lua-Torch
	S. Tulyakov, A. Ivanov and F. Fleuret: Weakly supervised learning of deep metrics for stereo reconstruction. ICCV 2017.								
27	3DMST			3.36 %	13.03 %	4.97 %	100.00 %	93 s	1 core @ >3.5 Ghz (C/C++)
	X. Lincheng Li and L. Zhang: 3D Cost Aggregation with Multiple Minimum Spanning Trees for Stereo Matching. submitted to Applied Optics .								
28	LPU			3.55 %	12.30 %	5.01 %	100.00 %	1650 s	1 core @ 2.5 Ghz (Matlab + C/C++)
29	OSF+TC			4.11 %	9.64 %	5.03 %	100.00 %	50 min	1 core @ 2.5 Ghz (C/C++)
	M. Neoral and J. Šochman: Object Scene Flow with Temporal Consistency. 22nd Computer Vision Winter Workshop (CVWW) 2017.								
30	SOSF			4.30 %	8.72 %	5.03 %	100.00 %	55 min	1 core @ 2.5 Ghz (Matlab + C/C++)
31	SGM+CNN			3.93 %	10.56 %	5.04 %	100.00 %	2 s	Nvidia GTX 970
32	SPS-St		code	3.84 %	12.67 %	5.31 %	100.00 %	2 s	1 core @ 3.5 Ghz (C/C++)
	K. Yamaguchi, D. McAllester and R. Urtasun: Efficient Joint Segmentation, Occlusion Labeling, Stereo and Flow Estimation. ECCV 2014.								
33	MNP			3.92 %	12.37 %	5.33 %	100.00 %	3 min	>8 cores @ 2.5 Ghz (C/C++)
34	MDP			4.19 %	11.25 %	5.36 %	100.00 %	11.4 s	4 cores @ 3.5 Ghz (Matlab + C/C++)
	A. Li, D. Chen, Y. Liu and Z. Yuan: Coordinating Multiple Disparity Proposals for Stereo Computation. IEEE Conference on Computer Vision and Pattern Recognition 2016.								
35	CPM2		code	4.13 %	12.03 %	5.44 %	99.95 %	3 s	1 core @ 3.5 Ghz (C/C++)
36	CNN-MS			3.89 %	13.28 %	5.45 %	100.00 %	3 min	GPU @ TITAN X (Lua/Torch)
37	UCNN			4.15 %	12.08 %	5.47 %	99.98 %	3 s	Nvidia GTX Titan X GPU (cuDNN)
38	JMR			4.35 %	11.25 %	5.50 %	99.81 %	1.3 sec	GTX TitanX (C/C++)
39	OSF		code	4.54 %	12.03 %	5.79 %	100.00 %	50 min	1 core @ 2.5 Ghz (C/C++)
	M. Menze and A. Geiger: Object Scene Flow for Autonomous Vehicles. Conference on Computer Vision and Pattern Recognition (CVPR) 2015.								
40	sqrtSGM			4.84 %	11.64 %	5.97 %	100.00 %	7.77 s	4 cores @ 3.5 Ghz (C/C++)
41	CSF			4.57 %	13.04 %	5.98 %	99.99 %	80 s	1 core @ 2.5 Ghz (C/C++)
	Z. Lv, C. Beall, P. Alcantarilla, F. Li, Z. Kira and F. Dellaert: A Continuous Optimization Approach for Efficient and Accurate Scene Flow. European Conf. on Computer Vision (ECCV) 2016.								
42	MBM			4.69 %	13.05 %	6.08 %	100.00 %	0.13 s	1 core @ 3.0 Ghz (C/C++)
	N. Einecke and J. Eggert: A Multi-Block-Matching Approach for Stereo. IV 2015.								
43	PR-Sceneflow		code	4.74 %	13.74 %	6.24 %	100.00 %	150 s	4 core @ 3.0 Ghz (Matlab + C/C++)
	C. Vogel, K. Schindler and S. Roth: Piecewise Rigid Scene Flow. ICCV 2013.								
44	SGM+DAISY		code	4.86 %	13.42 %	6.29 %	95.26 %	5 s	1 core @ 2.5 Ghz (C/C++)
45	DeepCostAggr			5.34 %	11.35 %	6.34 %	99.98 %	0.03 s	GPU @ 2.5 Ghz (C/C++)
46	SceneFFields			5.12 %	13.83 %	6.57 %	100.00 %	65 s	4 cores @ >3.5 Ghz (C/C++)
47	FSF+MS			5.72 %	11.84 %	6.74 %	100.00 %	2.7 s	4 cores @ 3.5 Ghz (C/C++)
	T. Taniai, S. Sinha and Y. Sato: Fast Multi-frame Stereo Scene Flow with Motion Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017) 2017.								
48	AABM			4.88 %	16.07 %	6.74 %	100.00 %	0.08 s	1 core @ 3.0 Ghz (C/C++)
	N. Einecke and J. Eggert: Stereo Image Warping for Improved Depth Estimation of Road Surfaces. IV 2013.								
49	SGM+C+NL		code	5.15 %	15.29 %	6.84 %	100.00 %	4.5 min	1 core @ 2.5 Ghz (C/C++)
	H. Hirschmüller: Stereo Processing by Semiglobal Matching and Mutual Information. PAMI 2008. D. Sun, S. Roth and M. Black: A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them. IJCV 2013.								
50	SGM+LDOF		code	5.15 %	15.29 %	6.84 %	100.00 %	86 s	1 core @ 2.5 Ghz (C/C++)
	H. Hirschmüller: Stereo Processing by Semiglobal Matching and Mutual Information. PAMI 2008. T. Brox and J. Malik: Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation. PAMI 2011.								
51	SGM+SF			5.15 %	15.29 %	6.84 %	100.00 %	45 min	16 core @ 3.2 Ghz (C/C++)
	H. Hirschmüller: Stereo Processing by Semiglobal Matching and Mutual Information. PAMI 2008. M. Hornacek, A. Fitzgibbon and C. Rother: SphereFlow: 6 DoF Scene Flow from RGB-D Pairs. CVPR 2014.								
52	SNCC			5.36 %	16.05 %	7.14 %	100.00 %	0.08 s	1 core @ 3.0 Ghz (C/C++)
	N. Einecke and J. Eggert: A Two-Stage Correlation Method for Stereoscopic Depth Estimation. DICTA 2010.								
53	rcam			6.17 %	14.01 %	7.47 %	100.00 %	12 s	8 cores @ 2.5 Ghz (Python + C/C++)
54	DMDE			6.89 %	12.92 %	7.90 %	100.00 %	7 s	4 cores @ 3.0 Ghz (C/C++)

55 CSCT+SGM+MF 6.91 % 14.87 % 8.24 % 100.00 % 0.0064 s Nvidia GTX Titan X @ 1.0 Ghz (CUDA)

D. Hernandez-Juarez, A. Chacon, A. Espinosa, D. Vazquez, J. Moure and A. Lopez: Embedded real-time stereo estimation via Semi-Global Matching on the GPU. Procedia Computer Science 2016.

56 MeshStereo code 5.82 % 21.21 % 8.38 % 100.00 % 87 s 1 core @ 2.5 Ghz (C/C++)

C. Zhang, Z. Li, Y. Cheng, R. Cai, H. Chao and Y. Rui: MeshStereo: A Global Stereo Model With Mesh Alignment Regularization for View Interpolation. The IEEE International Conference on Computer Vision (ICCV) 2015.

57 PCOF + ACTF 6.31 % 19.24 % 8.46 % 100.00 % 0.08 s GPU @ 2.0 Ghz (C/C++)

M. Derome, A. Plyer, M. Sanfourche and G. Le Besnerais: A Prediction-Correction Approach for Real-Time Optical Flow Computation Using Stereo. German Conference on Pattern Recognition 2016.

58 PCOF-LDOF 6.31 % 19.24 % 8.46 % 100.00 % 50 s 1 core @ 3.0 Ghz (C/C++)

M. Derome, A. Plyer, M. Sanfourche and G. Le Besnerais: A Prediction-Correction Approach for Real-Time Optical Flow Computation Using Stereo. German Conference on Pattern Recognition 2016.

59 BRIEF 7.04 % 18.72 % 8.99 % 100.00 % 3.72 s 4 cores @ >3.5 Ghz (C/C++)

60 CPL+SP 7.09 % 19.89 % 9.22 % 99.78 % 5 min 1 core @ 2.0 Ghz (C/C++)

61 ELAS code 7.86 % 19.04 % 9.72 % 92.35 % 0.3 s 1 core @ 2.5 Ghz (C/C++)

A. Geiger, M. Roser and R. Urtasun: Efficient Large-Scale Stereo Matching. ACCV 2010.

62 REAF code 8.43 % 18.51 % 10.11 % 100.00 % 1.1 s 1 core @ 2.5 Ghz (C/C++)

C. Cigla: Recursive Edge-Aware Filters for Stereo Matching. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops 2015.

63 iGF 8.64 % 21.85 % 10.84 % 100.00 % 220 s 1 core @ 3.0 Ghz (C/C++)

R. Hamzah, H. Ibrahim and A. Hassan: Stereo matching algorithm based on per pixel difference adjustment, iterative guided filter and graph segmentation. Journal of Visual Communication and Image Representation 2016.

64 OCV-SGBM code 8.92 % 20.59 % 10.86 % 90.41 % 1.1 s 1 core @ 2.5 Ghz (C/C++)

H. Hirschmueller: Stereo processing by semiglobal matching and mutual information. PAMI 2008.

65 SDM 9.41 % 24.75 % 11.96 % 62.56 % 1 min 1 core @ 2.5 Ghz (C/C++)

J. Kostkova: Stratified dense matching for stereopsis in complex scenes. BMVC 2003.

66 DSGCA 10.54 % 20.79 % 12.25 % 100.00 % 144 s >8 cores @ 3.5 Ghz (C/C++)

67 SGM&FlowFie+ 11.93 % 20.57 % 13.37 % 81.24 % 29 s 1 core @ 3.5 Ghz (C/C++)

68 GCSF code 11.64 % 27.11 % 14.21 % 100.00 % 2.4 s 1 core @ 2.5 Ghz (C/C++)

J. Cech, J. Sanchez-Riera and R. Horaud: Scene Flow Estimation by growing Correspondence Seeds. CVPR 2011.

69 CostFilter code 17.53 % 22.88 % 18.42 % 100.00 % 4 min 1 core @ 2.5 Ghz (Matlab)

C. Rhemann, A. Hosni, M. Bleyer, C. Rother and M. Gelautz: Fast Cost-Volume Filtering for Visual Correspondence and Beyond. CVPR 2011.

70 DWBSF 19.61 % 22.69 % 20.12 % 100.00 % 7 min 4 cores @ 3.5 Ghz (C/C++)

C. Richardt, H. Kim, L. Valgaerts and C. Theobalt: Dense Wide-Baseline Scene Flow From Two Handheld Video Cameras. 3DV 2016.

71 OCV-BM code 24.29 % 30.13 % 25.27 % 58.54 % 0.1 s 1 core @ 2.5 Ghz (C/C++)

G. Bradski: The OpenCV Library. Dr. Dobb's Journal of Software Tools 2000.

72 VSF code 27.31 % 21.72 % 26.38 % 100.00 % 125 min 1 core @ 2.5 Ghz (C/C++)

F. Huguet and F. Devernay: A Variational Method for Scene Flow Estimation from Stereo Sequences. ICCV 2007.

73 SED code 25.01 % 40.43 % 27.58 % 4.02 % 0.68 s 1 core @ 2.0 Ghz (C/C++)

D. Pe\~{n}a and A. Sutherland: Disparity Estimation by Simultaneous Edge Drawing. Computer Vision -- ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part II 2017.

74 MST code 45.83 % 38.22 % 44.57 % 100.00 % 7 s 1 core @ 2.5 Ghz (Matlab + C/C++)

Q. Yang: A Non-Local Cost Aggregation Method for Stereo Matching. CVPR 2012.

75 Test AD 58.86 % 57.65 % 58.66 % 100.00 % 181 s 2 cores @ 3.0 Ghz (C/C++)

Table as LaTeX | Only published Methods

Related Datasets

- HCI/Bosch Robust Vision Challenge: Optical flow and stereo vision challenge on high resolution imagery recorded at a high frame rate under diverse weather conditions (e.g., sunny, cloudy, rainy). The Robert Bosch AG provides a prize for the best performing method.
- Image Sequence Analysis Test Site (EISATS): Synthetic image sequences with ground truth information provided by UoA and Daimler AG. Some of the images come with 3D range sensor information.
- Middlebury Stereo Evaluation: The classic stereo evaluation benchmark, featuring four test images in version 2 of the benchmark, with very accurate ground truth from a structured light system. 38 image pairs are provided in total.
- Daimler Stereo Dataset: Stereo bad weather highway scenes with partial ground truth for freespace
- Make3D Range Image Data: Images with small-resolution ground truth used to learn and evaluate depth from single monocular images.
- Lubor Ladicky's Stereo Dataset: Stereo Images with manually labeled ground truth based on polygonal areas.

- Middlebury Optical Flow Evaluation: The classic optical flow evaluation benchmark, featuring eight test images, with very accurate ground truth from a shape from UV light pattern system. 24 image pairs are provided in total.

Citation

When using this dataset in your research, we will be happy if you cite us:

```
@INPROCEEDINGS{Menze2015CVPR,
  author = {Moritz Menze and Andreas Geiger},
  title = {Object Scene Flow for Autonomous Vehicles},
  booktitle = {Conference on Computer Vision and Pattern Recognition (CVPR)},
  year = {2015}
}
@INPROCEEDINGS{Menze2015ISA,
  author = {Moritz Menze and Christian Heipke and Andreas Geiger},
  title = {Joint 3D Estimation of Vehicles and Scene Flow},
  booktitle = {ISPRS Workshop on Image Sequence Analysis (ISA)},
  year = {2015}
}
```