Karlsruhe Institute of Technology

Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite

Andreas Geiger[†] Philip Lenz[†] Raquel Urtasun[‡]

[†] Karlsruhe Institute of Technology [‡] Toyota Technological Institute at Chicago

Andreas Geiger (KIT)

Philip Lenz (KIT)

Raquel Urtasun (TTI-C)

The Challenge of Autonomous Cars

DARPA GRAND CHALLENGE

DARPA URBAN CHALLENGE

State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars

State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars

State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars

3D Laser-scanner

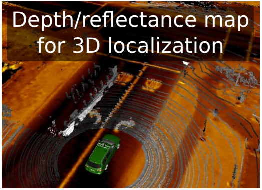
State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars



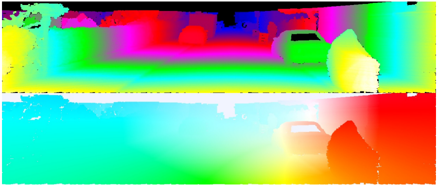
State of the art

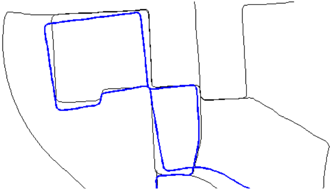
- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars





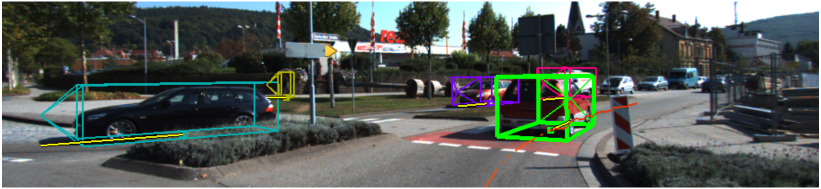
State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars



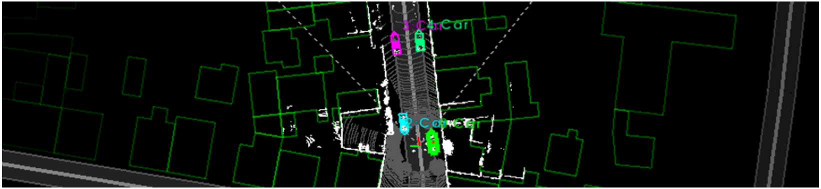
State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars



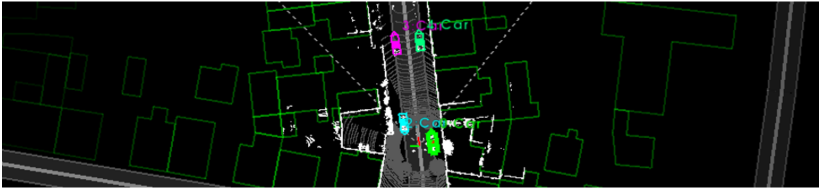
State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The Challenge of Autonomous Cars



State of the art

- Localization, path planning, obstacle avoidance
- Heavy use of 3D laser scanner and detailed maps

Problems for computer vision

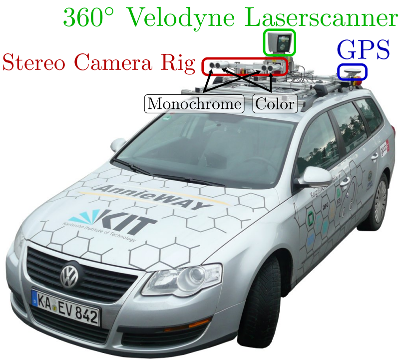
- Stereo, optical flow, localization
- Object detection, recognition and tracking
- Semantic segmentation, 3D scene understanding

The KITTI Vision Benchmark Suite

- **Two stereo rigs** (1392 $\times$ 512 px, 54 cm base, 90° opening)
- **Velodyne** laser scanner, **GPS+IMU** localization
- **6 hours** of recordings, 10 frames per second

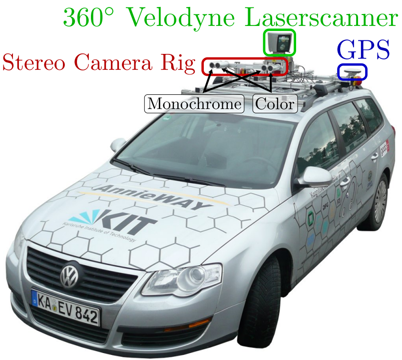
The KITTI Vision Benchmark Suite

- **Two stereo rigs** (1392 × 512 px, 54 cm base, 90° opening)
- **Velodyne** laser scanner, **GPS+IMU** localization
- **6 hours** of recordings, 10 frames per second

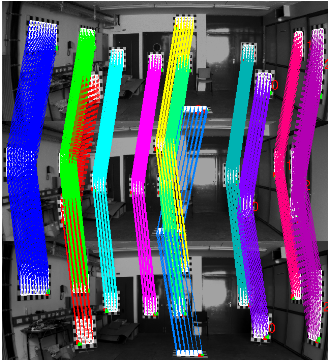


The KITTI Vision Benchmark Suite

- **Two stereo rigs** (1392 $\times$ 512 px, 54 cm base, 90° opening)
- **Velodyne** laser scanner, **GPS+IMU** localization
- **6 hours** of recordings, 10 frames per second



Sensor Calibration Challenges

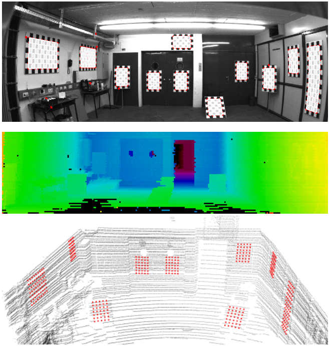


- Camera $\leftrightarrow$ camera calibration
- Velodyne $\leftrightarrow$ camera registration
- GPS $\leftrightarrow$ Velodyne registration

Geiger et al., ICRA 2012

ICP + Hand-eye calibration

Sensor Calibration Challenges

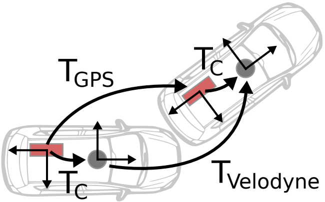


- Camera $\leftrightarrow$ camera calibration
- Velodyne $\leftrightarrow$ camera registration
- GPS $\leftrightarrow$ Velodyne registration

} Geiger et al., ICRA 2012

} ICP + Hand-eye calibration

Sensor Calibration Challenges

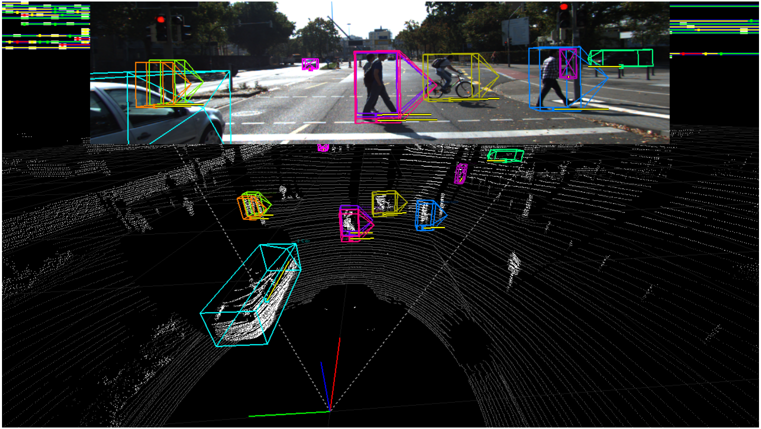


- Camera ↔ camera calibration
- Velodyne ↔ camera registration
- GPS ↔ Velodyne registration

} Geiger et al., ICRA 2012 (covers the first two points)

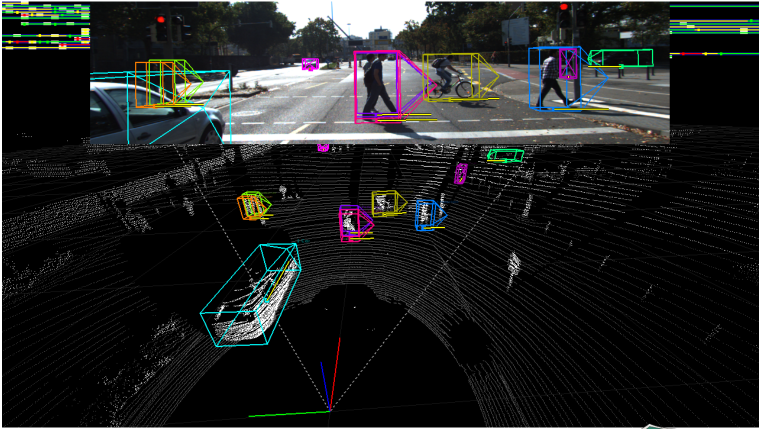
} ICP + Hand-eye calibration (covers the third point)

Data Annotation Challenges

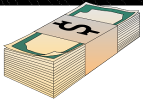


- **3D object labels:** 22 Annotators
- **Occlusion labels:** Mechanical Turk

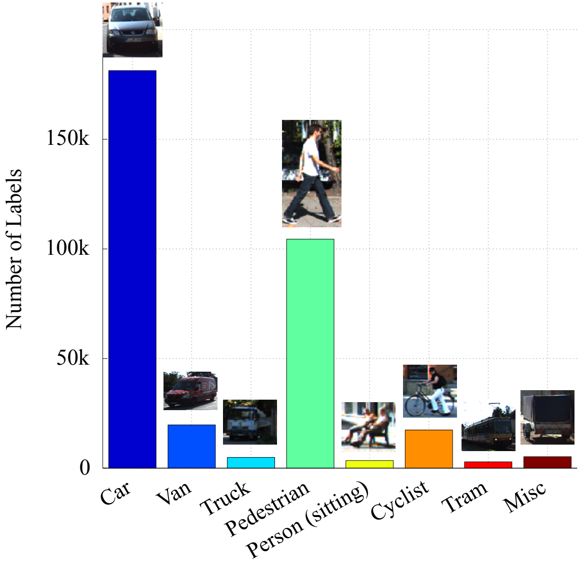
Data Annotation Challenges



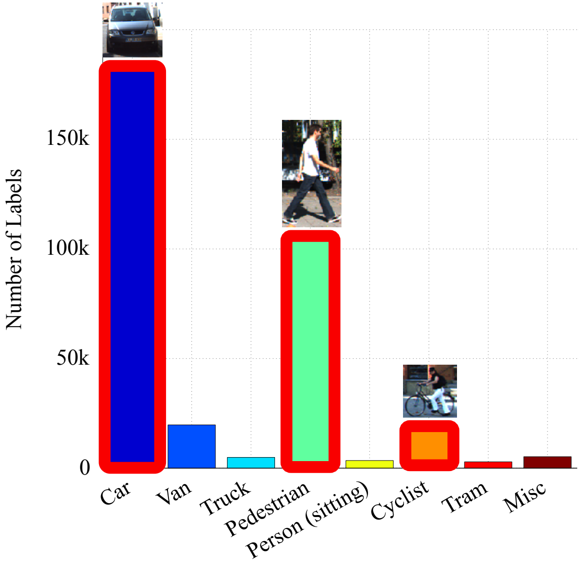
- **3D object labels:** 22 Annotators
- **Occlusion labels:** Mechanical Turk



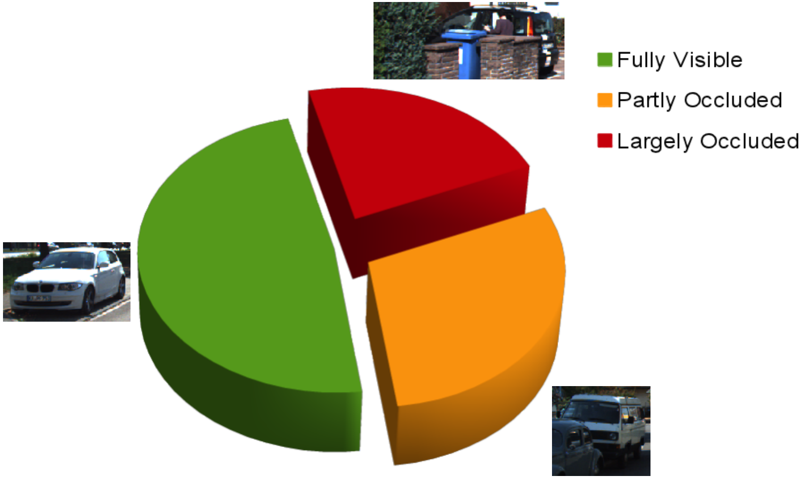
Data Statistics



Data Statistics

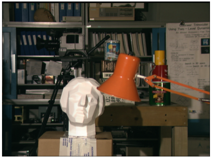


Data Statistics



Novel Challenges

Middlebury Stereo Evaluation – Version 2

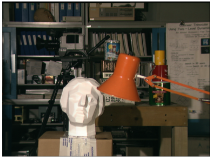


Error Threshold = 1 Algorithm	Avg.	Tsukuba ground truth			Venus ground truth			Teddy ground truth			Cones ground truth		
CoopRegion [41]	8.8	0.87 4	1.16 1	4.61 3	0.11 4	0.21 3	1.54 7	5.16 16	8.31 11	13.0 13	2.79 17	7.18 4	8.01 23
AdaptingBP [17]	9.0	1.11 19	1.37 7	5.79 19	0.10 3	0.21 4	1.44 5	4.22 8	7.06 6	11.8 9	2.48 7	7.92 11	7.32 10
ADCensus [94]	7.3	1.07 15	1.48 13	5.73 17	0.09 2	0.25 7	1.15 3	4.10 6	6.22 3	10.9 6	2.42 5	7.25 5	6.95 6
SurfaceStereo [79]	18.2	1.28 32	1.65 21	6.78 37	0.19 18	0.28 10	2.61 32	3.12 2	5.10 1	8.65 1	2.89 21	7.95 13	8.26 30
GC+SegmBorder [57]	27.1	1.47 45	1.82 32	7.86 58	0.19 19	0.31 12	2.44 26	4.25 9	5.55 2	10.9 7	4.99 77	5.78 1	8.66 37
WarpMat [55]	20.8	1.16 20	1.35 6	6.04 24	0.18 17	0.24 6	2.44 26	5.02 13	9.30 17	13.0 15	3.49 39	8.47 22	9.01 44
RDP [102]	12.5	0.97 10	1.39 9	5.00 9	0.21 23	0.38 19	1.89 13	4.84 10	9.94 19	12.6 11	2.53 8	7.69 8	7.38 11
RVbased [116]	11.6	0.95 9	1.42 11	4.98 8	0.11 6	0.29 11	1.07 1	5.98 21	11.6 31	15.4 27	2.35 3	7.61 6	6.81 5
OutlierConf [42]	12.9	0.88 5	1.43 12	4.74 5	0.18 16	0.26 9	2.40 22	5.01 12	9.12 16	12.8 12	2.78 16	8.57 23	6.99 7

- Average errors: 2 – 3% (non-occluded regions)

Novel Challenges

Middlebury Stereo Evaluation – Version 2



Error Threshold = 1 Algorithm	Avg.	Tsukuba ground truth			Venus ground truth			Teddy ground truth			Cones ground truth		
CoopRegion [41]	8.8	0.87 4	**1.16** 1	4.61 3	0.11 4	0.21 3	1.54 7	5.16 16	8.31 11	13.0 13	2.79 17	7.18 4	8.01 23
AdaptingBP [17]	9.0	1.11 19	1.37 7	5.79 19	0.10 3	0.21 4	1.44 5	4.22 8	7.06 6	11.8 9	2.48 7	7.92 11	7.32 10
ADCensus [94]	7.3	1.07 15	1.48 13	5.73 17	0.09 2	0.25 7	1.15 3	4.10 6	6.22 3	10.9 6	2.42 5	7.25 5	6.95 6
SurfaceStereo [79]	18.2	1.28 32	1.65 21	6.78 37	0.19 18	0.28 10	2.61 32	3.12 2	**5.10** 1	**8.65** 1	2.89 21	7.95 13	8.26 30
GC+SegmBorder [57]	27.1	1.47 45	1.82 32	7.86 58	0.19 19	0.31 12	2.44 26	4.25 9	5.55 2	10.9 7	4.99 77	**5.78** 1	8.66 37
WarpMat [55]	20.8	1.16 20	1.35 6	6.04 24	0.18 17	0.24 6	2.44 26	5.02 13	9.30 17	13.0 15	3.49 39	8.47 22	9.01 44
RDP [102]	12.5	0.97 10	1.39 9	5.00 9	0.21 23	0.38 19	1.89 13	4.84 10	9.94 19	12.6 11	2.53 8	7.69 8	7.38 11
RVbased [116]	11.6	0.95 9	1.42 11	4.98 8	0.11 6	0.29 11	**1.07** 1	5.98 21	11.6 31	15.4 27	2.35 3	7.61 6	6.81 5
OutlierConf [42]	12.9	0.88 5	1.43 12	4.74 5	0.18 16	0.26 9	2.40 22	5.01 12	9.12 16	12.8 12	2.78 16	8.57 23	6.99 7

- Average errors: 2 – 3% (non-occluded regions)

Novel Challenges

Fast guided cost-volume filtering (Rhemann et al., CVPR 2011)

Middlebury, Errors: **2.7%**

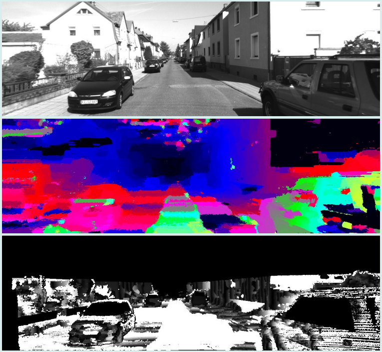
- Error threshold: 1 px (Middlebury) / 3 px (KITTI)

Novel Challenges

Fast guided cost-volume filtering (Rhemann et al., CVPR 2011)

Middlebury, Errors: 2.7%

KITTI, Errors: 46.3%



- Error threshold: 1 px (Middlebury) / 3 px (KITTI)

Novel Challenges

So what is the difference?

Middlebury

- Laboratory
- Lambertian
- Rich in texture
- Medium-size label set
- Largely fronto-parallel

KITTI

- Moving vehicle
- Specularities
- Sensor saturation
- Large label set
- Strong slants

Novel Challenges

So what is the difference?

Middlebury

- Laboratory
- Lambertian
- Rich in texture
- Medium-size label set
- Largely fronto-parallel

KITTI

- Moving vehicle
- Specularities
- Sensor saturation
- Large label set
- Strong slants

Novel Challenges

So what is the difference?

Middlebury

- Laboratory
- Lambertian
- Rich in texture
- Medium-size label set
- Largely fronto-parallel

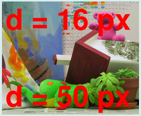
KITTI

- Moving vehicle
- Specularities
- Sensor saturation
- Large label set
- Strong slants

Novel Challenges

So what is the difference?

Middlebury



- Laboratory
- Lambertian
- Rich in texture
- Medium-size label set
- Largely fronto-parallel

KITTI

- Moving vehicle
- Specularities
- Sensor saturation
- Large label set
- Strong slants

Novel Challenges

So what is the difference?

Middlebury

- Laboratory
- Lambertian
- Rich in texture
- Medium-size label set
- Largely fronto-parallel

KITTI

- Moving vehicle
- Specularities
- Sensor saturation
- Large label set
- Strong slants

Stereo Evaluation

200 training images / 200 test images

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare	
1	PCBP	4.72 %	6.16 %	1.0 px	1.4 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)	☐	sub.
Anonymous submission										
2	ITGV	6.31 %	7.40 %	1.3 px	1.5 px	100.00 %	7 s	1 core @ 3.0 Ghz (Matlab + C/C++)	☐	sub.
Rene Ranftl, Stefan Gehrig, Thomas Pock and Horst Bischof. Pushing the Limits of Stereo Using Variational Stereo Estimation. IEEE Intelligent Vehicles Symposium 2012.										
3	OCV-SGBM	7.64 %	9.13 %	1.8 px	2.0 px	86.50 %	1.1 s	1 core @ 2.5 Ghz (C/C++)	☐	
Heiko Hirschmueller. Stereo processing by semiglobal matching and mutual information. PAMI 2008.										
4	ELAS	8.24 %	9.95 %	1.4 px	1.6 px	94.55 %	0.3 s	1 core @ 2.5 Ghz (C/C++)	☐	
Andreas Geiger, Martin Roser and Raquel Urtasun. Efficient Large-Scale Stereo Matching. ACCV 2010.										
5	SDM	10.98 %	12.19 %	2.0 px	2.3 px	63.58 %	1 min	1 core @ 2.5 Ghz (C/C++)	☐	
Jana Kostkova. Stratified dense matching for stereopsis in complex scenes. BMVC 2003.										
6	GCSF	12.06 %	13.26 %	1.9 px	2.1 px	60.77 %	2.4 s	1 core @ 2.5 Ghz (C/C++)	☐	sub.
Jan Cech, Jordi Sanchez-Riera and Radu P. Horaud. Scene Flow Estimation by Growing Correspondence Seeds. CVPR 2011.										
7	GCS	13.37 %	14.54 %	2.1 px	2.3 px	51.06 %	2.2 s	1 core @ 2.5 Ghz (C/C++)	☐	
Jan Cech and Radim Sara. Efficient Sampling of Disparity Space for Fast And Accurate Matching. BenCOS 2007.										
8	CostFilter	19.96 %	21.05 %	5.0 px	5.4 px	100.00 %	4 min	1 core @ 2.5 Ghz (Matlab)	☐	
Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother and Margrit Gelautz. Fast Cost-Volume Filtering for Visual Correspondence and Beyond. CVPR 2011.										
9	OCV-BM	25.39 %	26.72 %	7.6 px	7.9 px	55.84 %	0.1 s	1 core @ 2.5 Ghz (C/C++)	☐	
G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools 2000.										
10	GC+occ	33.50 %	34.74 %	8.6 px	9.2 px	87.57 %	6 min	1 core @ 2.5 Ghz (C/C++)	☐	
Vladimir Kolmogorov and Ramin Zabih. Computing Visual Correspondence with Occlusions using Graph Cuts. ICCV 2001.										

Stereo Evaluation

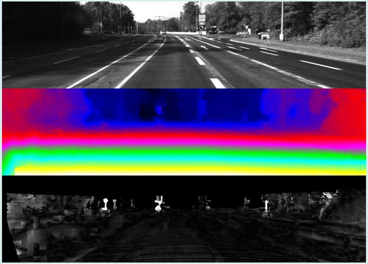
200 training images / 200 test images

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare	
1	PCBP	4.72 %	6.16 %	1.0 px	1.4 px	100.00 %	5 min	4 cores @ 2.5 Ghz (Matlab + C/C++)	☐	**sub.**
Anonymous submission										
2	ITGV	6.31 %	7.40 %	1.3 px	1.5 px	100.00 %	7 s	1 core @ 3.0 Ghz (Matlab + C/C++)	☐	**sub.**
Rene Ranftl, Stefan Gehrig, Thomas Pock and Horst Bischof. Pushing the Limits of Stereo Using Variational Stereo Estimation. IEEE Intelligent Vehicles Symposium 2012.										
3	OCV-SGBM	7.64 %	9.13 %	1.8 px	2.0 px	86.50 %	1.1 s	1 core @ 2.5 Ghz (C/C++)	☐	
Heiko Hirschmueller. Stereo processing by semiglobal matching and mutual information. PAMI 2008.										
4	ELAS	8.24 %	9.95 %	1.4 px	1.6 px	94.55 %	0.3 s	1 core @ 2.5 Ghz (C/C++)	☐	
Andreas Geiger, Martin Roser and Raquel Urtasun. Efficient Large-Scale Stereo Matching. ACCV 2010.										
5	SDM	10.98 %	12.19 %	2.0 px	2.3 px	63.58 %	1 min	1 core @ 2.5 Ghz (C/C++)	☐	
Jana Kostkova. Stratified dense matching for stereopsis in complex scenes. BMVC 2003.										
6	GCSF	12.06 %	13.26 %	1.9 px	2.1 px	60.77 %	2.4 s	1 core @ 2.5 Ghz (C/C++)	☐	**sub.**
Jan Cech, Jordi Sanchez-Riera and Radu P. Horaud. Scene Flow Estimation by Growing Correspondence Seeds. CVPR 2011.										
7	GCS	13.37 %	14.54 %	2.1 px	2.3 px	51.06 %	2.2 s	1 core @ 2.5 Ghz (C/C++)	☐	
Jan Cech and Radim Sara. Efficient Sampling of Disparity Space for Fast And Accurate Matching. BenCOS 2007.										
8	CostFilter	19.96 %	21.05 %	5.0 px	5.4 px	100.00 %	4 min	1 core @ 2.5 Ghz (Matlab)	☐	
Christoph Rhemann, Asmaa Hosni, Michael Bleyer, Carsten Rother and Margrit Gelautz. Fast Cost-Volume Filtering for Visual Correspondence and Beyond. CVPR 2011.										
9	OCV-BM	25.39 %	26.72 %	7.6 px	7.9 px	55.84 %	0.1 s	1 core @ 2.5 Ghz (C/C++)	☐	
G. Bradski. The OpenCV Library. Dr. Dobb's Journal of Software Tools 2000.										
10	GC+occ	33.50 %	34.74 %	8.6 px	9.2 px	87.57 %	6 min	1 core @ 2.5 Ghz (C/C++)	☐	
Vladimir Kolmogorov and Ramin Zabih. Computing Visual Correspondence with Occlusions using Graph Cuts. ICCV 2001.										

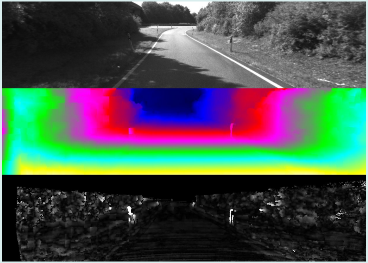
Stereo Evaluation

Particle Convex Belief Propagation (PCBP): **Best Results**

Errors: 0.5%



Errors: 0.5%

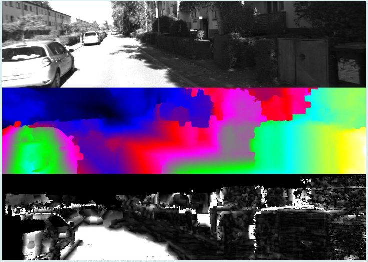


- Natural scenes, lots of texture, no objects
- A couple of wrong pixels at poles

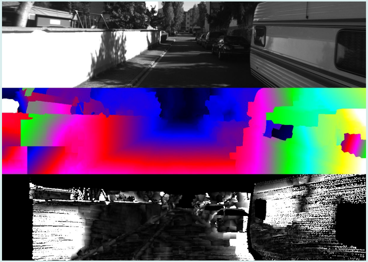
Stereo Evaluation

Particle Convex Belief Propagation (PCBP): **Worst Results**

Errors: 19.5%



Errors: 21.1%



- Inner city scenes, lots of objects
- Textureless surfaces, sensor saturation, reflections

Optical Flow Evaluation

200 training images / 200 test images

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare	
1	TGV2CENSUS	11.14 %	18.42 %	2.9 px	6.6 px	100.00 %	4 s	1 core @ 3.0 Ghz (Matlab + C/C++)	☐	**sub.**
	Manuel Werlberger. Convex Approaches for High Performance Video Processing. 2012.									
2	HS	19.92 %	28.86 %	5.8 px	11.7 px	100.00 %	3 min	1 core @ 2.5 Ghz (Matlab + C/C++)	☐	
	Berthold K. P. Horn and Brian G. Schunck. Determining optical flow: A Retrospective. AI 1993.									
3	RSRS-Flow	20.74 %	29.68 %	6.2 px	12.1 px	100.00 %	4 min	1 core @ 2.5 Ghz (Matlab)	☐	**sub.**
	Pratim Ghosh and B.S. Manjunath. Robust Simultaneous Registration and Segmentation with Sparse Error Reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence 2012.									
4	LDOF	21.86 %	31.31 %	5.5 px	12.4 px	100.00 %	1 min	1 core @ 2.5 Ghz (C/C++)	☐	
	T. Brox and J. Malik. Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation. PAMI 2011.									
5	C+NL	24.64 %	33.35 %	9.0 px	16.4 px	100.00 %	3 min	1 core @ 2.5 Ghz (C/C++)	☐	
	Deqing Sun, Stefan Roth and Michael J. Black. Secrets of optical flow estimation and their principles. CVPR 2010.									
6	HMM	29.35 %	38.13 %	6.8 px	14.7 px	100.00 %	10 min	1 core @ 2.5 Ghz (C/C++)	☐	**sub.**
	Anonymous submission									
7	DB-TV-L1	30.75 %	39.13 %	7.8 px	14.6 px	100.00 %	16 s	1 core @ 2.5 Ghz (Matlab)	☐	
	C. Zach, T. Pock and H. Bischof. A Duality Based Approach for Realtime TV- L1 Optical Flow. DAGM 2007.									
8	GCSF	33.23 %	41.74 %	7.0 px	15.3 px	48.27 %	2.4 s	1 core @ 2.5 Ghz (C/C++)	☐	**sub.**
	Jan Cech, Jordi Sanchez-Riera and Radu P. Horaud. Scene Flow Estimation by Growing Correspondence Seeds. CVPR 2011.									
9	HAOF	35.76 %	43.36 %	11.1 px	18.2 px	100.00 %	16.2 s	1 core @ 2.5 Ghz (C/C++)	☐	
	Thomas Brox, Andrés Bruhn, Nils Papenberg and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. ECCV 2004.									
10	PolyExpand	47.54 %	53.95 %	17.2 px	25.2 px	100.00 %	1 s	1 core @ 2.5 Ghz (C/C++)	☐	
	Gunnar Farneback. Two-Frame Motion Estimation Based on Polynomial Expansion. Proceedings of the 13th Scandinavian Conference on Image Analysis 2003.									

Optical Flow Evaluation

Karlsruhe Institute of Technology

200 training images / 200 test images

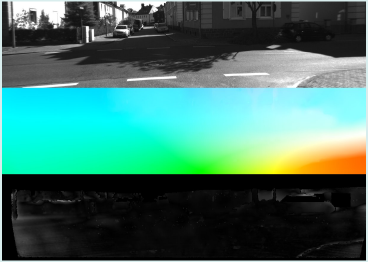
Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All	Density	Runtime	Environment	Compare	
1	TGV2CENSUS	11.14 %	18.42 %	2.9 px	6.6 px	100.00 %	4 s	1 core @ 3.0 Ghz (Matlab + C/C++)	☐	sub.
Manuel Werlberger. Convex Approaches for High Performance Video Processing. 2012.										
2	HS	19.92 %	28.86 %	5.8 px	11.7 px	100.00 %	3 min	1 core @ 2.5 Ghz (Matlab + C/C++)	☐	
Berthold K. P. Horn and Brian G. Schunck. Determining optical flow: A Retrospective. AI 1993.										
3	RSRS-Flow	20.74 %	29.68 %	6.2 px	12.1 px	100.00 %	4 min	1 core @ 2.5 Ghz (Matlab)	☐	sub.
Pratim Ghosh and B.S. Manjunath. Robust Simultaneous Registration and Segmentation with Sparse Error Reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence 2012.										
4	LDOF	21.86 %	31.31 %	5.5 px	12.4 px	100.00 %	1 min	1 core @ 2.5 Ghz (C/C++)	☐	
T. Brox and J. Malik. Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation. PAMI 2011.										
5	C+NL	24.64 %	33.35 %	9.0 px	16.4 px	100.00 %	3 min	1 core @ 2.5 Ghz (C/C++)	☐	
Deqing Sun, Stefan Roth and Michael J. Black. Secrets of optical flow estimation and their principles. CVPR 2010.										
6	HMM	29.35 %	38.13 %	6.8 px	14.7 px	100.00 %	10 min	1 core @ 2.5 Ghz (C/C++)	☐	sub.
Anonymous submission										
7	DB-TV-L1	30.75 %	39.13 %	7.8 px	14.6 px	100.00 %	16 s	1 core @ 2.5 Ghz (Matlab)	☐	
C. Zach, T. Pock and H. Bischof. A Duality Based Approach for Realtime TV- L1 Optical Flow. DAGM 2007.										
8	GCSF	33.23 %	41.74 %	7.0 px	15.3 px	48.27 %	2.4 s	1 core @ 2.5 Ghz (C/C++)	☐	sub.
Jan Cech, Jordi Sanchez-Riera and Radu P. Horaud. Scene Flow Estimation by Growing Correspondence Seeds. CVPR 2011.										
9	HAOF	35.76 %	43.36 %	11.1 px	18.2 px	100.00 %	16.2 s	1 core @ 2.5 Ghz (C/C++)	☐	
Thomas Brox, Andrés Bruhn, Nils Papenberg and Joachim Weickert. High accuracy optical flow estimation based on a theory for warping. ECCV 2004.										
10	PolyExpand	47.54 %	53.95 %	17.2 px	25.2 px	100.00 %	1 s	1 core @ 2.5 Ghz (C/C++)	☐	
Gunnar Farneback. Two-Frame Motion Estimation Based on Polynomial Expansion. Proceedings of the 13th Scandinavian Conference on Image Analysis 2003.										

Optical Flow Evaluation

Second Order Total Generalized Variation: **Best Results**

Errors: 0.5%

Errors: 0.5%

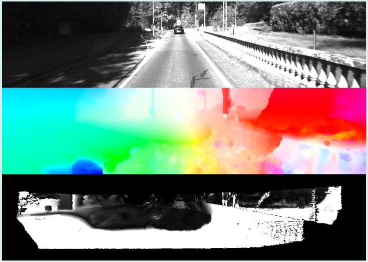


- City scenes with slow motion (intersections)
- Small flow vectors ($<$ 30 px)

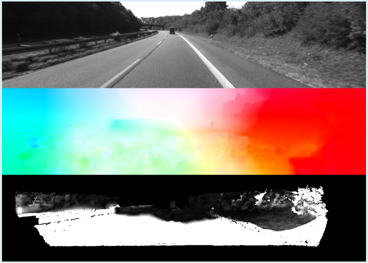
Optical Flow Evaluation

Second Order Total Generalized Variation: **Worst Results**

Errors: 56.5%



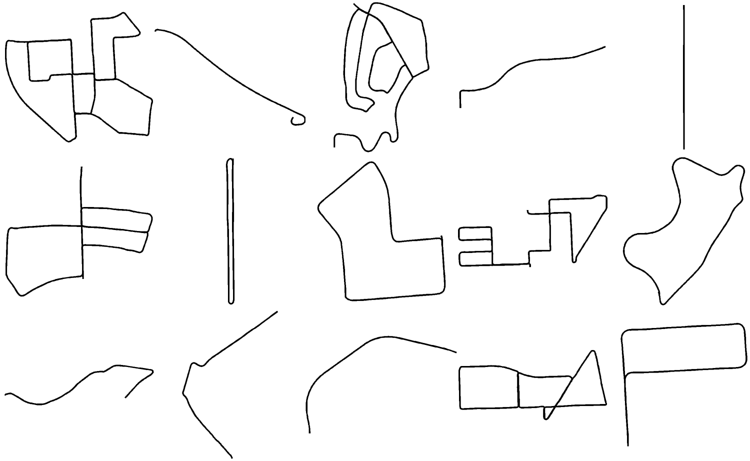
Errors: 58.8%



- Difficult lighting conditions, highway driving
- Large flow vectors ($>$ 150 px)

SLAM Evaluation

22 sequences – 40 kilometers



SLAM Evaluation

Karlsruhe Institute of Technology

TOYOTA TECHNOLOGICAL INSTITUTE CHICAGO

11 training sequences / 11 test sequences

Rank	Method	Submitted	Translation	Rotation	Runtime	Environment	Compare	
1	GT_VO3pt	8 Jun. 2012	**2.21 %**	0.0117 [deg/m]	1.26 s	1 core @ 2.5 Ghz (C/C++)	☐	**sub.**
2	VISO2-S	12 Mar. 2012	2.28 %	0.0154 [deg/m]	0.05 s	1 core @ 2.5 Ghz (C/C++)	☐	
Andreas Geiger, Julius Ziegler and Christoph Stiller. StereoScan: Dense 3d Reconstruction in Real-time. IEEE Intelligent Vehicles Symposium 2011.								
3	VO3pt	15 Mar. 2012	2.93 %	**0.0116 [deg/m]**	0.56 s	1 core @ 2.0 Ghz (C/C++)	☐	**sub.**
P. F. Alcantarilla. Vision Based Localization: From Humanoid Robots to Visually Impaired People. 2011. P.F. Alcantarilla, J.J. Yebes, J. Almazán and L.M. Bergasa. On Combining Visual SLAM and Dense Scene Flow to Increase the Robustness of Localization and Mapping in Dynamic Environments. IEEE Intl. Conf. on Robotics and Automation (ICRA) 2012.								
4	VO3ptLBA	14 Mar. 2012	3.17 %	0.0180 [deg/m]	0.57 s	1 core @ 2.0 Ghz (C/C++)	☐	**sub.**
P. F. Alcantarilla. Vision Based Localization: From Humanoid Robots to Visually Impaired People. 2011. P.F. Alcantarilla, J.J. Yebes, J. Almazán and L.M. Bergasa. On Combining Visual SLAM and Dense Scene Flow to Increase the Robustness of Localization and Mapping in Dynamic Environments. IEEE Intl. Conf. on Robotics and Automation (ICRA) 2012.								
5	VOFS	15 Mar. 2012	4.21 %	0.0158 [deg/m]	0.51 s	1 core @ 2.0 Ghz (C/C++)	☐	**sub.**
Michael Kaess, Kai Ni and Frank Dellaert. Flow separation for fast and robust stereo odometry. ICRA 2009. P.F. Alcantarilla, L.M. Bergasa and F. Dellaert. Visual Odometry priors for robust {EKF-SLAM}. ICRA 2010.								
6	VOFSLBA	15 Mar. 2012	4.35 %	0.0189 [deg/m]	0.52 s	1 core @ 2.0 Ghz (C/C++)	☐	**sub.**
Michael Kaess, Kai Ni and Frank Dellaert. Flow separation for fast and robust stereo odometry. ICRA 2009. P.F. Alcantarilla, L.M. Bergasa and F. Dellaert. Visual Odometry priors for robust {EKF-SLAM}. ICRA 2010.								
7	VISO2-M	4 Apr. 2012	13.79 %	0.0372 [deg/m]	0.1 s	1 core @ 2.5 Ghz (C/C++)	☐	
Andreas Geiger, Julius Ziegler and Christoph Stiller. StereoScan: Dense 3d Reconstruction in Real-time. IEEE Intelligent Vehicles Symposium 2011.								

Conclusion

Where are we now?

- Realistic dataset with 3D ground truth
  - Stereo
  - Optical flow
  - SLAM
  - Object detection / orientation estimation
- Complement existing benchmarks / reduce overfitting
- Submit your results: **www.cvlibs.net**

Conclusion

Where are we now?

- Realistic dataset with 3D ground truth
  - Stereo
  - Optical flow
  - SLAM
  - Object detection / orientation estimation
- Complement existing benchmarks / reduce overfitting
- Submit your results: **www.cvlibs.net**

Conclusion

Where are we now?

- Realistic dataset with 3D ground truth
  - Stereo
  - Optical flow
  - SLAM
  - Object detection / orientation estimation
- Complement existing benchmarks / reduce overfitting
- Submit your results: **www.cvlibs.net**

Conclusion

Where are we now?

- Realistic dataset with 3D ground truth
  - Stereo
  - Optical flow
  - SLAM
  - Object detection / orientation estimation
- Complement existing benchmarks / reduce overfitting
- Submit your results: **www.cvlibs.net**

Conclusion

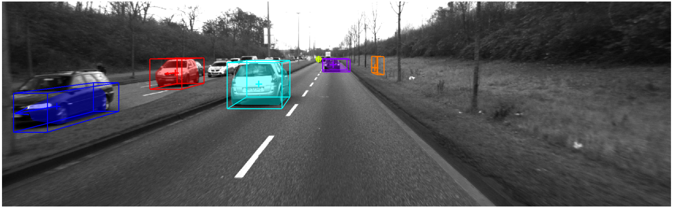
Where are we now?

- Realistic dataset with 3D ground truth
  - Stereo
  - Optical flow
  - SLAM
  - Object detection / orientation estimation
- Complement existing benchmarks / reduce overfitting
- Submit your results: **www.cvlibs.net**

Conclusion

But KITTI is much more ...

- 3D object tracking
- Loop closure (SLAM)
- Structure-from-Motion
- Semantic segmentation (class labels)
- 3D scene understanding (layout and objects)
- Use of maps

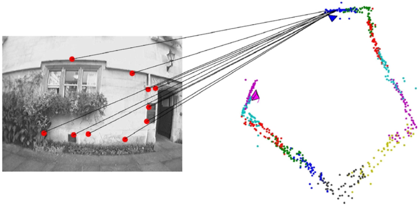


Lenz et al., IV 2011

Conclusion

But KITTI is much more ...

- 3D object tracking
- Loop closure (SLAM)
- Structure-from-Motion
- Semantic segmentation (class labels)
- 3D scene understanding (layout and objects)
- Use of maps

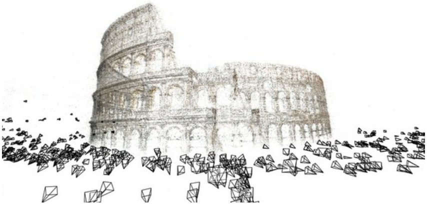


Williams et al., RAS 2009

Conclusion

But KITTI is much more ...

- 3D object tracking
- Loop closure (SLAM)
- Structure-from-Motion
- Semantic segmentation (class labels)
- 3D scene understanding (layout and objects)
- Use of maps



Agarwal et al., ICCV 2009

Conclusion

But KITTI is much more ...

- 3D object tracking
- Loop closure (SLAM)
- Structure-from-Motion
- Semantic segmentation (class labels)
- 3D scene understanding (layout and objects)
- Use of maps

Wojek et al., ECCV 2008

Conclusion

But KITTI is much more ...

- 3D object tracking
- Loop closure (SLAM)
- Structure-from-Motion
- Semantic segmentation (class labels)
- 3D scene understanding (layout and objects)
- Use of maps

Geiger et al., CVPR 2011 and NIPS 2011

Conclusion

But KITTI is much more ...

- 3D object tracking
- Loop closure (SLAM)
- Structure-from-Motion
- Semantic segmentation (class labels)
- 3D scene understanding (layout and objects)
- Use of maps

OpenStreetMap – The free Wiki World Map

Related Datasets and Benchmarks

Stereo / Optical Flow	setting	#images	ground truth
EISATS	synthetic	500	dense
Middlebury	laboratory	40	dense
Make3D Stereo	real	260	0.5 %
Ladicky	real	70	manual
KITTI	**real**	**400**	**50 %**

SLAM	setting	length	metric
TUM RGB-D	indoor	0.4 km	✓
New College	outdoor	2.2 km	
Malaga 2009	outdoor	6.4 km	
Ford Campus	outdoor	5.1 km	
KITTI	**outdoor**	**39.2 km**	✓

Related Datasets and Benchmarks

Object Detection	#cat.	#labels/cat.	occlusion	3D	orientation
Caltech 101	101	40-800			
MIT StreetScenes	9	3k			
LabelMe	4000	60			
ETHZ Pedestrian	1	12k			
PASCAL 2011	20	1k	✓		
Daimler	1	56k	✓		
Caltech Pedestrian	1	350k	✓		
COIL-100	100	72		✓	discrete
EPFL Multi-View	20	90		✓	discrete
Caltech 3D	100	144		✓	discrete
KITTI	**3**	**1k - 40k**	✓	✓	**continuous**