Computer Vision for Autonomous Vehicles

Problems, Datasets and State of the Art

Other titles in Foundations and Trends® in Computer Graphics and Vision

Discrete Graphical Models - An Optimization Perspective
Bogdan Savchynskyy
ISBN: 978-1-68083-638-7

Line Drawings from 3D Models: A Tutorial
Pierre Bénard and Aaron Hertzmann
ISBN: 978-1-68083-590-8

Publishing and Consuming 3D Content on the Web: A Survey
Marco Potenziani, Marco Callieri, Matteo Dellepiane and Roberto Scopigno
ISBN: 978-1-68083-536-6

Crowdsourcing in Computer Vision
Adriana Kovashka, Olga Russakovsky, Li Fei-Fei and Kristen Grauman
ISBN: 978-1-68083-212-9

The Path to Path-Traced Movies
Per H. Christensen and Wojciech Jarosz
ISBN: 978-1-68083-210-5

(Hyper)-Graphs Inference through Convex Relaxations and Move Making Algorithms
Nikos Komodakis, M. Pawan Kumar and Nikos Paragios
ISBN: 978-1-68083-138-2

Computer Vision for Autonomous Vehicles

Problems, Datasets and State of the Art

Joel Janai
Max-Planck-Institute for Intelligent Systems Tübingen, Germany
University of Tübingen, Germany
joel.janai@tuebingen.mpg.de

Fatma Güney
College of Engineering
Koç University, Turkey
fguney@ku.edu.tr

Aseem Behl
Max-Planck-Institute for Intelligent Systems Tübingen, Germany
University of Tübingen, Germany
aseem.behl@tuebingen.mpg.de

Andreas Geiger
Max-Planck-Institute for Intelligent Systems Tübingen, Germany
University of Tübingen, Germany
andreas.geiger@tuebingen.mpg.de

the essence of knowledge
Boston — Delft

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

J. Janai, F. Güney, A. Behl and A. Geiger. *Computer Vision for Autonomous Vehicles*. Foundations and Trends® in Computer Graphics and Vision, vol. 12, no. 1–3, pp. 1–308, 2020.

ISBN: 978-1-68083-689-9

Foundations and Trends® in Computer Graphics and Vision

Volume 12, Issue 1–3, 2020

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2020, Volume 12, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Contents

1 Introduction **3**

2 History of Autonomous Driving **7**

3 Sensors **15**

3.1 Camera Models 16

3.1.1 Omnidirectional Cameras 17

3.1.2 Event Cameras 18

3.2 Calibration 19

4 Datasets & Benchmarks **21**

4.1 Computer Vision Datasets 23

4.1.1 Object Recognition 23

4.1.2 Object Tracking 25

4.1.3 Stereo and 3D Reconstruction 26

4.1.4 Optical Flow 27

4.2 Autonomous Driving Datasets 29

4.2.1 Object Detection and Semantic Segmentation 31

4.2.2 Tracking 33

4.2.3 Traffic Sign Detection 33

4.2.4 Road and Lane Detection 34

4.2.5 Flow and Stereo 34
4.2.6 Long-Term Autonomy 35
4.3 Synthetic Data Generation Using Game Engines 36

5 Object Detection **39**
5.1 Problem Definition 39
5.2 Methods 41
5.2.1 Classical Pipeline 41
5.2.2 Part-Based Approaches 43
5.2.3 Deep Learning for Detection 45
5.2.4 Real-Time Pedestrian Detection 47
5.2.5 Human Pose Estimation 48
5.2.6 Traffic Sign Detection 48
5.2.7 3D Object Detection from 2D Images 50
5.2.8 3D Object Detection from 3D Point Clouds 51
5.3 Datasets 52
5.4 Metrics 53
5.5 State of the Art on KITTI 54
5.6 Discussion 61

6 Object Tracking **65**
6.1 Problem Definition 65
6.2 Methods 66
6.2.1 Tracking by Detection 66
6.2.2 Pedestrian Tracking 69
6.2.3 Joint Detection and Tracking 70
6.2.4 Deep Learning for Multi-Object Tracking 71
6.2.5 3D Object Tracking 72
6.3 Datasets 74
6.4 Metrics 76
6.5 State of the Art on MOT & KITTI 78
6.6 Discussion 81

7 Semantic Segmentation **83**
7.1 Problem Definition 83
7.2 Methods 83

7.2.1 Deep Learning for Semantic Segmentation 86
7.2.2 Videos 90
7.2.3 Street Side Views 92
7.2.4 3D Data 94
7.2.5 Road Segmentation 98
7.2.6 Free Space Estimation 99
7.2.7 Stixels 102
7.2.8 Aerial Images 103
7.3 Datasets 106
7.4 Metrics 106
7.5 State of the Art on Cityscapes 107
7.6 Discussion 108

8 Semantic Instance Segmentation **110**
8.1 Problem Definition 110
8.2 Methods 110
8.2.1 Proposal-Based Approaches 110
8.2.2 Proposal-Free Approaches 113
8.2.3 Panoptic Segmentation 114
8.3 Datasets 116
8.4 Metrics 116
8.5 State of the Art on Cityscapes 116
8.6 Discussion 118

9 Stereo **119**
9.1 Problem Definition 119
9.2 Methods 119
9.2.1 Matching Cost 120
9.2.2 Energy Optimization 121
9.2.3 Higher-Order Models 123
9.2.4 Piecewise Planar Priors 123
9.2.5 Segmentation-Based Models 124
9.2.6 Deep Learning for Stereo Matching 125
9.2.7 Variable Baseline 127
9.2.8 Omnidirectional Cameras 128
9.3 Datasets 128

9.4 Metrics . . . 128
9.5 State of the Art on KITTI . . . 129
9.6 Discussion . . . 130

10 Multi-View 3D Reconstruction 133
10.1 Problem Definition . . . 133
10.2 Structure from Motion . . . 134
10.3 Multi-View Stereo . . . 136
10.3.1 Planarity and Primitives . . . 138
10.3.2 Shape Priors . . . 138
10.3.3 Semantics . . . 140
10.3.4 Efficient Reconstruction . . . 142
10.3.5 Deep Learning for Multi-View Stereo . . . 142
10.3.6 Omnidirectional Cameras . . . 144
10.4 Datasets . . . 145
10.5 Metrics . . . 145
10.6 State of the Art on ETH3D & Tanks and Temples . . . 146
10.7 Discussion . . . 148

11 Optical Flow 149
11.1 Problem Definition . . . 149
11.2 Methods . . . 150
11.2.1 Sparse Matches . . . 152
11.2.2 Epipolar Flow . . . 153
11.2.3 Semantic Segmentation . . . 154
11.2.4 Deep Learning for Optical Flow . . . 155
11.2.5 High-Speed Flow . . . 158
11.2.6 Confidences . . . 159
11.3 Datasets . . . 161
11.4 Metrics . . . 161
11.5 State of the Art on KITTI . . . 161
11.6 Discussion . . . 164

12 3D Scene Flow 165
12.1 Problem Definition . . . 165
12.2 Methods . . . 166

12.2.1 Piecewise Rigidity 167
12.2.2 Semantic Segmentation 168
12.2.3 Scene Flow from 3D Point Clouds 169
12.3 Datasets 170
12.4 Metrics 170
12.5 State of the Art on KITTI 170
12.6 Discussion 172

13 Mapping, Localization & Ego-Motion Estimation 174
13.1 Problem Definition 174
13.2 Mapping 175
13.2.1 Metric Maps 175
13.2.2 Semantic Maps 176
13.3 Localization 177
13.3.1 Structure-Based Localization 180
13.3.2 Cross-View Localization 183
13.3.3 Semantic Alignment from LiDAR 185
13.4 Ego-Motion Estimation 186
13.4.1 Drift 190
13.4.2 Loop Closure Detection 191
13.4.3 Simultaneous Localization and Mapping (SLAM) . 193
13.5 Datasets 196
13.6 Metrics 197
13.7 State of the Art on KITTI 198
13.8 Discussion 202

14 Scene Understanding 205
14.1 Problem Definition 205
14.2 Methods 206
14.2.1 Road Topology and Traffic Participants 206
14.2.2 Physical and Temporal Relationships 207
14.3 Discussion 208

15 End-to-End Learning for Autonomous Driving 210
15.1 Problem Definition 210
15.2 Methods 211

15.2.1 Behavior Cloning 211
15.2.2 Reinforcement Learning 214
15.2.3 Combined Methods 216
15.2.4 Intermediate Representations 217
15.2.5 Transferring from Simulation to the Real World . . 219
15.3 Datasets 221
15.4 Metrics 222
15.5 Discussion 223

16 Conclusion **224**

Acknowledgements **226**

References **227**

Computer Vision for Autonomous Vehicles

Joel Janai[1], Fatma Güney[2], Aseem Behl[3] and Andreas Geiger[4]

[1] *Max-Planck-Institute for Intelligent Systems Tübingen, Germany University of Tübingen, Germany; joel.janai@tuebingen.mpg.de*
[2] *College of Engineering, Koç University, Turkey; fguney@ku.edu.tr*
[3] *Max-Planck-Institute for Intelligent Systems Tübingen, Germany University of Tübingen, Germany; aseem.behl@tuebingen.mpg.de*
[4] *Max-Planck-Institute for Intelligent Systems Tübingen, Germany University of Tübingen, Germany; andreas.geiger@tuebingen.mpg.de*

ABSTRACT

Recent years have witnessed enormous progress in AI-related fields such as computer vision, machine learning, and autonomous vehicles. As with any rapidly growing field, it becomes increasingly difficult to stay up-to-date or enter the field as a beginner. While several survey papers on particular sub-problems have appeared, no comprehensive survey on problems, datasets, and methods in computer vision for autonomous vehicles has been published. This monograph attempts to narrow this gap by providing a survey on the state-of-the-art datasets and techniques. Our survey includes both the historically most relevant literature as well as the current state of the art on several specific topics, including recognition, reconstruction, motion estimation, tracking, scene understanding, and end-to-end learning for autonomous driving. Towards this goal, we analyze the performance of the state of the art on several challenging benchmarking datasets, including KITTI, MOT, and

Joel Janai, Fatma Güney, Aseem Behl and Andreas Geiger (2020), "Computer Vision for Autonomous Vehicles", Foundations and Trends® in Computer Graphics and Vision: Vol. 12, No. 1–3, pp 1–308. DOI: 10.1561/0600000079.

Cityscapes. Besides, we discuss open problems and current research challenges. To ease accessibility and accommodate missing references, we also provide a website that allows navigating topics as well as methods and provides additional information.

1

Introduction

Since the first successful demonstrations in the 1980s [169, 171, 661], great progress has been made in the field of autonomous vehicles. However, despite these advances and ambitious commercial goals, fully autonomous navigation in general environments has not been realized to date. The reason for this is two-fold: First, autonomous systems which operate in complex dynamic environments require models which generalize to unpredictable situations and reason in a timely manner. Second, informed decisions require accurate perception, yet most of the existing computer vision models are still inferior to human perception and reasoning.

Existing approaches to self-driving can be roughly categorized into modular pipelines and monolithic end-to-end learning approaches. Both approaches are contrasted at a conceptual level in Figure 1.1. The modular pipeline is the standard approach to autonomous driving, mostly followed in the industry. The key idea is to break down the complex mapping function from high-dimensional inputs to low-dimensional control variables into modules which can be independently developed, trained, and tested. In Figure 1.1 (top), these modules comprise low-level perception, scene parsing, path planning, and vehicle control. However,

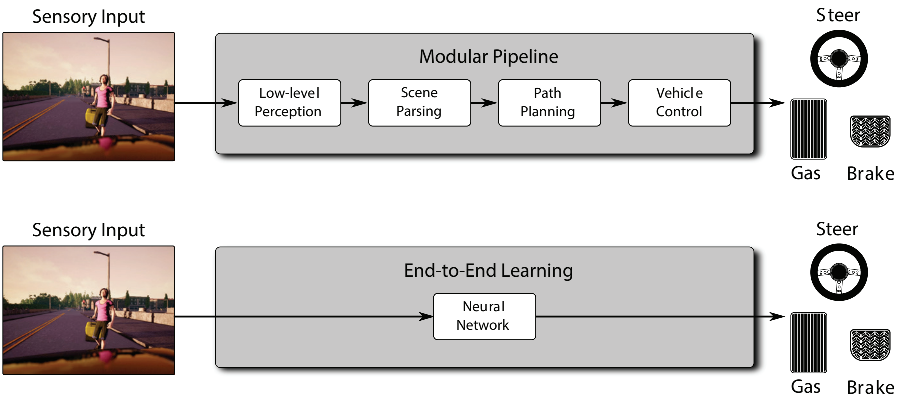

Figure 1.1: Approaches to self-driving. Classical modular pipeline (top) vs. monolithic end-to-end learning approach (bottom). See text for details.

this is just one particular example of modularizing a self-driving stack and other or more fine-grained modularizations are also possible. Existing approaches typically leverage machine learning (e.g., deep neural networks) to extract low-level features or to parse the scene into individual components. In contrast, path planning and vehicle control are dominated by classical state machines, search algorithms, and control models.

The major advantage of modular pipelines is that they deploy human interpretable intermediate representations such as detected objects or free space information which allow gaining insights into failure modes of the system. Furthermore, the development of modular pipelines can be easily parallelized within companies where typically different teams work on different aspects of the driving problem simultaneously. Furthermore, it is comparably easy to integrate first principles and prior knowledge about the problem into the system. Examples include traffic laws that can be explicitly enforced in the planner or knowledge about the vehicle dynamics, which lead to improved vehicle control. Other aspects that are more difficult to specify by hand, such as the appearance of pedestrians, are learned from large annotated datasets.

A major drawback of modular approaches is the fact that human-designed intermediate representations are not necessarily optimal for the driving task, which typically includes aspects like safety, comfort, and time for reaching the goal. Moreover, most modules are trained and validated independently from each other, making use of auxiliary loss functions. Consider the problem of object detection as an example. Most objects in the scene are not directly relevant for the driving task, yet the learning algorithm is not informed about the relevance of each object and therefore tasks a neural network to detect all objects with equal importance. Thus, the network is wasting capacity on irrelevant objects while not being able to detect the driving relevant objects with the necessary accuracy. This demonstrates the difficulty of defining appropriate intermediate representations and auxiliary loss functions.

An alternative to modular pipelines is end-to-end learning-based models which try to learn a policy, i.e., a function from observations to actions using a generic model such as a deep neural network. This approach is illustrated in Figure 1.1 (bottom) and discussed in detail in Section 15. The network parameters can be learned either via imitation learning by replicating the behavior of a teacher or using reinforcement learning by exploring the world and taking actions that are likely to yield a high user-specified reward. However, reinforcement learning approaches suffer from the credit assignment and reward shaping problems, are typically slow and can only be applied in non-safety-critical simulation environments. Imitation learning, on the other hand, suffers from overfitting and does not easily generalize to novel scenarios. Furthermore, holistic neural network-based approaches are often hard to interpret as they present themselves as "black boxes" to the user which do not reveal *why* a certain error has occurred.

In this survey, we focus on perception for autonomous vehicles. In particular, we discuss the perception-related modules of the modular pipeline as well as end-to-end learning-based approaches. Other aspects of the self-driving problem are discussed in related surveys: For example, Winner *et al.* [716] put emphasis on driver assistance systems, considering both their structure and their function. Similarly, Klette [357] provides an overview of vision-based driver assistance systems. They describe most aspects of the perception problem at a high level but do not provide an in-depth review of the state of the art in each task as we pursue in this survey. Complementary to our work, Zhu *et al.* [798] provide an overview of environment perception for intelligent vehicles, focusing on lane detection, traffic sign/light recognition as well as vehicle tracking. In contrast, our goal is to bridge the gap between the robotics, intelligent vehicles, and computer vision communities by providing an extensive overview and comparison, including works from all three fields.

This survey is structured as follows: first, we provide a brief history of autonomous driving, followed by an introduction to camera models and calibration techniques. We then provide an overview of autonomous driving-related datasets with a particular focus on perception before surveying the relevant perception tasks and the state-of-the-art algorithms for solving them. More specifically, we review object detection, tracking, semantic (instance) segmentation, reconstruction, motion estimation, and scene understanding techniques. Each section starts with the problem definition, an overview over the most important methods and main design choices, a qualitative and quantitative analysis of the top-performing techniques on the most popular datasets, as well as a discussion of the state of the art in this area. Finally, we provide an overview of state-of-the-art end-to-end models for autonomous driving before concluding this survey. To ease navigation, we also provide an interactive online tool[1] which visualizes the surveyed papers with an interactive graph and additional information in an easily accessible manner. We hope that our survey will become a useful tool for researchers in the field of autonomous vision and lowers the entry barrier for beginners by providing a thorough overview of the field.

[1]http://www.cvlibs.net/projects/autonomous_vision_survey.

References

[1] Acuna, D., H. Ling, A. Kar, and S. Fidler (2018). "Efficient interactive annotation of segmentation datasets with polygon-RNN++". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[2] Agarwal, S., N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski (2009). "Building Rome in a day". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[3] Aghdam, H. H., E. J. Heravi, and D. Puig (2016). "A practical approach for detection and classification of traffic signs using Convolutional Neural Networks". *Robotics and Autonomous Systems (RAS)*. 84: 97–112.

[4] Alvarez, J. M., T. Gevers, and A. M. Lopez (2010). "3D scene priors for road detection". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[5] Álvarez, J. M., T. Gevers, Y. LeCun, and A. M. López (2012). "Road scene segmentation from a single image". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[6] Álvarez, J. M. and A. M. López (2011). "Road detection based on illuminant invariance". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 12(1): 184–193.

[7] Aly, M. (2008). "Real time detection of lane markers in urban streets". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[8] Anandan, P. (1989). "A computational framework and an algorithm for the measurement of visual motion". *International Journal of Computer Vision (IJCV)*. 2(3): 283–310.

[9] Andreasson, H. and A. J. Lilienthal (2010). "6D scan registration using depth-interpolated local image features". In: *Robotics and Autonomous Systems (RAS)*.

[10] Andriluka, M., S. Roth, and B. Schiele (2008). "People-tracking-by-detection and people-detection-by-tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[11] Andriluka, M., S. Roth, and B. Schiele (2010). "Monocular 3D pose estimation and tracking by detection". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[12] Andriyenko, A. and K. Schindler (2011). "Multi-target tracking by continuous energy minimization". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[13] Andriyenko, A., K. Schindler, and S. Roth (2012). "Discrete-continuous optimization for multi-target tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[14] Anguelov, D., C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. S. Ogale, L. Vincent, and J. Weaver (2010). "Google street view: Capturing the world at street level". *IEEE Computer*. 43(6): 32–38.

[15] Arandjelović, R., P. Gronat, A. Torii, T. Pajdla, and J. Sivic (2016). "NetVLAD: CNN architecture for weakly supervised place recognition". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[16] Arbeláez, P. A., J. Pont-Tuset, J. T. Barron, F. Marqués, and J. Malik (2014). "Multiscale combinatorial grouping". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[17] Arnab, A. and P. H. S. Torr (2017). "Pixelwise instance segmentation with a dynamically instantiated network". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[18] Babaee, M., A. Athar, and G. Rigoll (2018). "Multiple people tracking using hierarchical deep tracklet re-identification". arXiv: 1811.04091 [cs.CV].

[19] Badino, H., U. Franke, and R. Mester (2007). "Free space computation using stochastic occupancy grids and dynamic programming". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV) Workshops.*

[20] Badino, H., U. Franke, and D. Pfeiffer (2009). "The stixel world - A compact medium level representation of the 3D-world". In: *Proc. of the DAGM Symposium on Pattern Recognition (DAGM).*

[21] Badino, H., D. Huber, and T. Kanade (2012). "Real-time topometric localization". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).*

[22] Badrinarayanan, V., I. Budvytis, and R. Cipolla (2014). "Mixture of trees probabilistic graphical model for video segmentation". *International Journal of Computer Vision (IJCV).* 110(1): 14–29.

[23] Badrinarayanan, V., F. Galasso, and R. Cipolla (2010). "Label propagation in video sequences". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[24] Badrinarayanan, V., A. Kendall, and R. Cipolla (2017). "SegNet: A deep convolutional encoder-decoder architecture for image segmentation". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 39(12): 2481–2495.

[25] Bai, M., W. Luo, K. Kundu, and R. Urtasun (2016). "Exploiting semantic information and deep matching for optical flow". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[26] Bai, M. and R. Urtasun (2017). "Deep Watershed transform for instance segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 2858–2866.

[27] Bailer, C., B. Taetz, and D. Stricker (2015). "Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[28] Baker, S., D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski (2011). “A database and evaluation methodology for optical flow”. *International Journal of Computer Vision (IJCV).* 92: 1–31.

[29] Balntas, V. (2019). “SILDa: A multi-task dataset for evaluating visual localization”. https://medium.com/scape-technologies/silda-a-multi-task-dataset-for-evaluating-visual-localization-7fc6c2c56c74. Online: accessed 17-June-2019.

[30] Balntas, V., L. Hammarstrand, H. Heijnen, F. Kahl, W. Maddern, K. Mikolajczyk, T. Pajdla, M. Pollefeys, T. Sattler, J. L. Schönberger, P. Speciale, J. Sivic, C. Toft, and A. Torii (2019). “Workshop on long-term visual localization under changing conditions”. https://sites.google.com/view/ltvl2019/. Online: accessed 17-June-2019.

[31] Balntas, V., S. Li, and V. Prisacariu (2018). “RelocNet: Continuous metric learning relocalisation using neural nets”. In: *Proc. of the European Conf. on Computer Vision (ECCV).* 782–799.

[32] Bansal, M., A. Krizhevsky, and A. Ogaler (2018). “ChauffeurNet: Learning to drive by imitating the best and synthesizing the worst”. arXiv: 1812.03079.

[33] Bansal, M., H. S. Sawhney, H. Cheng, and K. Daniilidis (2011). “Geo-localization of street views with aerial image databases”. In: *Proc. of the International Conf. on Multimedia (ICM).*

[34] Bao, S., M. Chandraker, Y. Lin, and S. Savarese (2013). “Dense object reconstruction with semantic priors”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[35] Barnes, N., A. Zelinsky, and L. S. Fletcher (2008). “Real-time speed sign detection using the radial symmetry detector”. *IEEE Trans. on Intelligent Transportation Systems (T-ITS).* 9(2): 322–332.

[36] Behl, A., O. H. Jafari, S. K. Mustikovela, H. A. Alhaija, C. Rother, and A. Geiger (2017). “Bounding boxes, segmentations and object coordinates: How important is recognition for 3D scene flow estimation in autonomous driving scenarios?” In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[37] Behl, A., D. Paschalidou, S. Donne, and A. Geiger (2019). "PointFlowNet: Learning representations for rigid motion estimation from point clouds". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[38] Behley, J., M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall (2019). "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences". arXiv: 1904.01416 [cs.CV].

[39] Behley, J., V. Steinhage, and A. B. Cremers (2012). "Performance of histogram descriptors for the classification of 3D laser range data in urban environments". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[40] Behley, J., V. Steinhage, and A. B. Cremers (2013). "Laser-based segment classification using a mixture of bag-of-words". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[41] Beltrán, J., C. Guindel, F. M. Moreno, D. Cruzado, F. García, and A. de la Escalera (2018). "BirdNet: A 3D object detection framework from LiDAR information". In: *Proc. IEEE Conf. on Intelligent Transportation Systems (ITSC)*. 3517–3523.

[42] Benenson, R., M. Mathias, R. Timofte, and L. J. Van Gool (2012). "Pedestrian detection at 100 frames per second". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[43] Benenson, R., M. Omran, J. H. Hosang, and B. Schiele (2014). "Ten years of pedestrian detection, what have we learned?" In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[44] Berclaz, J., F. Fleuret, and P. Fua (2009). "Multiple object tracking using flow linear programming". *Performance Evaluation of Tracking and Surveillance*.

[45] Berclaz, J., F. Fleuret, E. Turetken, and P. Fua (2011). "Multiple object tracking using K-shortest paths optimization". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 33(9): 1806–1819.

[46] Bertasius, G., L. Torresani, and J. Shi (2018). "Object detection in video with spatiotemporal sampling networks". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[47] Bertozzi, M., L. Bombini, A. Broggi, M. Buzzoni, E. Cardarelli, S. Cattani, P. Cerri, A. Coati, S. Debattisti, A. Falzoni, R. I. Fedriga, M. Felisa, L. Gatti, A. Giacomazzo, P. Grisleri, M. C. Laghi, L. Mazzei, P. Medici, M. Panciroli, P. P. Porta, P. Zani, and P. Versari (2011). "VIAC: An out of ordinary experiment". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. 175–180.

[48] Bertozzi, M., A. Broggi, and A. Fascioli (2000). "Vision-based intelligent vehicles: State of the art and perspectives". *Robotics and Autonomous Systems (RAS)*. 32(1): 1–16.

[49] Besl, P. and H. McKay (1992). "A method for registration of 3D shapes". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 14: 239–256.

[50] Bewley, A., Z. Ge, L. Ott, F. Ramos, and B. Upcroft (2016). "Simple online and realtime tracking". In: *Proc. IEEE International Conf. on Image Processing (ICIP)*.

[51] Bewley, A., J. Rigley, Y. Liu, J. Hawke, R. Shen, V. D. Lam, and A. Kendall (2018). "Learning to drive from simulation without real world labels". arXiv: 1812.03823 [cs.CV].

[52] Bhotika, R., D. J. Fleet, and K. N. Kutulakos (2002). "A probabilistic theory of occupancy and emptiness". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[53] Binas, J., D. Neil, S. Liu, and T. Delbrück (2017). "DDD17: End-To-end DAVIS driving dataset". *Proc. of the International Conf. on Machine Learning (ICML) Workshops*.

[54] Black, M. J. and P. Anandan (1993). "A framework for the robust estimation of optical flow". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[55] Blaha, M., C. Vogel, A. Richard, J. D. Wegner, T. Pock, and K. Schindler (2016). "Large-scale semantic 3D reconstruction: An adaptive multi-resolution model for multi-class volumetric labeling". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[56] Blanco-Claraco, J., F. Moreno-Dueñas, and J. G. Jiménez (2014). "The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario". *International Journal of Robotics Research (IJRR)*. 33(2): 207–214.

[57] Bochinski, E., V. Eiselein, and T. Sikora (2017). "High-speed tracking-by-detection without using image information". In: *Proc. of International Conf. on Advanced Video and Signal Based Surveillance (AVSS)*. 1–6.

[58] Bódis-Szomorú, A., H. Riemenschneider, and L. Van Gool (2016). "Efficient volumetric fusion of airborne and street-side data for urban reconstruction". In: *Proc. of the International Conf. on Pattern Recognition (ICPR)*.

[59] Bogo, F., A. Kanazawa, C. Lassner, P. V. Gehler, J. Romero, and M. J. Black (2016). "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[60] Bojarski, M., D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba (2016). "End to end learning for self-driving cars". arXiv: 1604.07316 [cs.CV].

[61] Borkar, A., M. Hayes, and M. T. Smith (2012). "A novel lane detection system with efficient ground truth generation". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 13(1): 365–374.

[62] Bouguet, J.-Y. (2010). "Camera calibration toolbox for Matlab". URL: http://www.vision.caltech.edu/bouguetj/calib_doc.

[63] Boukhers, Z., K. Shirahama, and M. Grzegorzek (2018). "Less restrictive camera odometry estimation from monocular camera". *Multimedia Tools Appl.* 77(13): 16199–16222.

[64] Boykov, Y., O. Veksler, and R. Zabih (1999). "Fast approximate energy minimization via graph cuts". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 23: 2001.

[65] Brabandere, B. D., D. Neven, and L. Van Gool (2017). "Semantic instance segmentation with a discriminative loss function". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops.*

[66] Brahmbhatt, S., J. Gu, K. Kim, J. Hays, and J. Kautz (2018). "Geometry-aware learning of maps for camera localization". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2616–2625.

[67] Braid, D., A. Broggi, and G. Schmiedel (2006). "The TerraMax autonomous vehicle". *Journal of Field Robotics (JFR)*. 23(9): 693–708.

[68] Braun, M., S. Krebs, F. Flohr, and D. M. Gavrila (2019). "The EuroCity persons dataset: A novel benchmark for object detection". arXiv: 1805.07193 [cs.CV].

[69] Braun, M., Q. Rao, Y. Wang, and F. Flohr (2016). "Pose-RCNN: Joint object detection and pose estimation using 3D object proposals". In: *Proc. IEEE Conf. on Intelligent Transportation Systems (ITSC)*. IEEE. 1546–1551.

[70] Braunschweig, T. U. (2010). "Project Stadtpilot". https://www.tu-braunschweig.de/stadtpilot. Online: accessed 18-October-2019.

[71] Bredies, K., K. Kunisch, and T. Pock (2010). "Total generalized variation". *Journal of Imaging Sciences (SIAM)*. 3(3): 492–526.

[72] Breitenstein, M. D., F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool (2009). "Robust tracking-by-detection using a detector confidence particle filter". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[73] Breitenstein, M. D., F. Reichlin, B. Leibe, E. Koller-Meier, and L. J. Van Gool (2011). "Online multiperson tracking-by-detection from a single, uncalibrated camera". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 33(9): 1820–1833.

[74] Brendel, W., M. R. Amer, and S. Todorovic (2011). "Multi-object tracking as maximum weight independent set". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[75] Broggi, A., M. Bertozzi, A. Fascioli, and G. Conte (1999). *Automatic Vehicle Guidance: The Experience of the Argo Vehicle*. Singapore: World Scientific.

[76] Broggi, A., M. Bertozzi, A. Fascioli, and M. Sechi (2000). "Shape-based pedestrian detection". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[77] Broggi, A., P. Cerri, S. Debattisti, M. C. Laghi, P. Medici, D. Molinari, M. Panciroli, and A. Prioletti (2015). "PROUD – Public Road Urban Driverless-Car Test". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 16(6): 3508–3519.

[78] Broggi, A., P. Cerri, P. Medici, P. P. Porta, and G. Ghisio (2007). "Real time road signs recognition". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 981–986.

[79] Broggi, A., P. Medici, E. Cardarelli, P. Cerri, A. Giacomazzo, and N. Finardi (2010). "Development of the control system for the Vislab Intercontinental Autonomous Challenge". In: *Proc. IEEE Conf. on Intelligent Transportation Systems (ITSC)*. 635–640.

[80] Brooks, R. A. (1981). "Model-based three dimensional interpretations of two dimensional images". In: *Proc. of the International Joint Conf. on Artificial Intelligence (IJCAI)*. 619–624.

[81] Brox, T. and J. Malik (2011). "Large displacement optical flow: Descriptor matching in variational motion estimation". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 33(Mar.): 500–513.

[82] Brubaker, M. A., A. Geiger, and R. Urtasun (2016). "Map-based probabilistic visual self-localization". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 38(4): 652–665.

[83] Bruhn, A. and J. Weickert (2006). "A confidence measure for variational optic flow methods". In: *Geometric Properties for Incomplete Data*. Ed. by R. Klette, R. Kozera, L. Noakes, and J. Weickert. Dordrecht: Springer Netherlands. 283–298. DOI: 10.1007/1-4020-3858-8_15.

[84] Buczko, M. and V. Willert (2016a). "Flow-decoupled normalized reprojection error for visual odometry". In: *Proc. IEEE Conf. on Intelligent Transportation Systems (ITSC)*.

[85] Buczko, M. and V. Willert (2016b). "How to distinguish inliers from outliers in visual odometry for high-speed automotive applications". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[86] Buczko, M. and V. Willert (2017). "Monocular outlier detection for visual odometry". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. 739–745.

[87] Buczko, M., V. Willert, J. Schwehr, and J. Adamy (2018). "Self-validation for automotive visual odometry". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. 1–6.

[88] Budvytis, I., V. Badrinarayanan, and R. Cipolla (2010). "Label propagation in complex video sequences using semi-supervised learning". In: *Proc. of the British Machine Vision Conf. (BMVC)*.

[89] Buehler, M., K. Iagnemma, and S. Singh (2007). *The 2005 Darpa Grand Challenge: The Great Robot Race*. Vol. 36. Springer.

[90] Buehler, M., K. Iagnemma, and S. Singh (2009). "The DARPA urban challenge". *DARPA Challenge*. Advanced Robotics 56.

[91] Bunschoten, R., B. J. A. Kröse, and N. A. Vlassis (2003). "Robust scene reconstruction from an omnidirectional vision system". *IEEE Trans. on Robotics and Automation (TRA)*. 19(2): 351–357.

[92] Butler, D. J., J. Wulff, G. B. Stanley, and M. J. Black (2012). "A naturalistic open source movie for optical flow evaluation". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[93] Caesar, H., V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom (2019). "nuScenes: A multimodal dataset for autonomous driving". arXiv: 1903.11027.

[94] Cai, Z., Q. Fan, R. S. Feris, and N. Vasconcelos (2016). "A unified multi-scale deep convolutional neural network for fast object detection". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[95] Camposeco, F., A. Cohen, M. Pollefeys, and T. Sattler (2019). "Hybrid scene compression for visual localization". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[96] Carlevaris-Bianco, N., A. K. Ushani, and R. M. Eustice (2016). "University of Michigan North Campus long-term vision and lidar dataset". *International Journal of Robotics Research (IJRR)*. 35(9): 1023–1035.

[97] Carreira, J., R. Caseiro, J. P. Batista, and C. Sminchisescu (2012). "Semantic segmentation with second-order pooling". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 430–443.

[98] Carreira, J. and C. Sminchisescu (2012). "CPMC: Automatic object segmentation using constrained parametric min-cuts". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 34(7): 1312–1328.

[99] Cech, J., J. Sanchez-Riera, and R. P. Horaud (2011). "Scene flow estimation by growing correspondence seeds". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[100] Center, G. H. (2017). "Self-driving cars, in 1956?" https://www.gmheritagecenter.com/featured/Autonomous_Vehicles.html. Online: accessed 18-October-2019.

[101] Cernea, D. (2015). "OpenMVS: open multiple view stereovision". http://cdcseacave.github.io/openMVS. Online: accessed 23-April-2019.

[102] Chabot, F., M. Chaouch, J. Rabarisoa, C. Teulière, and T. Chateau (2017). "Deep MANTA: A coarse-to-fine many-task network for joint 2D and 3D vehicle analysis from monocular image". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2040–2049.

[103] Chang, J. and Y. Chen (2018). "Pyramid stereo matching network". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 5410–5418.

[104] Chang, M., J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays (2019). "Argoverse: 3D tracking and forecasting with rich maps". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 8748–8757.

[105] Chen, C., A. Seff, A. L. Kornhauser, and J. Xiao (2015a). "DeepDriving: Learning affordance for direct perception in autonomous driving". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 2722–2730.

[106] Chen, D., B. Zhou, V. Koltun, and P. Krähenbühl (2019). "Learning by cheating". In: *Proc. Conf. on Robot Learning (CoRL)*.

[107] Chen, L., M. D. Collins, Y. Zhu, G. Papandreou, B. Zoph, F. Schroff, H. Adam, and J. Shlens (2018a). "Searching for efficient multi-scale architectures for dense image prediction". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 8713–8724.

[108] Chen, L.-C., S. Fidler, A. L. Yuille, and R. Urtasun (2014). "Beat the MTurkers: Automatic image labeling from weak 3D supervision". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[109] Chen, L., A. Hermans, G. Papandreou, F. Schroff, P. Wang, and H. Adam (2018b). "MaskLab: Instance segmentation by refining object detection with semantic and direction features". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 4013–4022.

[110] Chen, L.-C., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille (2015b). "Semantic image segmentation with deep convolutional nets and fully connected CRFs". In: *Proc. of the International Conf. on Learning Representations (ICLR)*.

[111] Chen, L., G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille (2018c). "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 40(4): 834–848.

[112] Chen, L., G. Papandreou, F. Schroff, and H. Adam (2017a). "Rethinking atrous convolution for semantic image segmentation". arXiv: 1706.05587 [cs.CV].

[113] Chen, L., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam (2018d). "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 833–851.

[114] Chen, Q. and V. Koltun (2016). "Full flow: Optical flow estimation by global optimization over regular grids". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[115] Chen, X., K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun (2015c). "3D Object proposals for accurate object class detection". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[116] Chen, X., K. Kundu, Y. Zhu, H. Ma, S. Fidler, and R. Urtasun (2018e). "3D object proposals using stereo imagery for accurate object class detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 40(5): 1259–1272.

[117] Chen, X., H. Ma, J. Wan, B. Li, and T. Xia (2017b). "Multi-view 3D object detection network for autonomous driving". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[118] Cherabier, I., C. Häne, M. R. Oswald, and M. Pollefeys (2016). "Multi-label semantic 3D reconstruction using voxel blocks". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[119] Cherabier, I., J. Schönberger, M. Oswald, M. Pollefeys, and A. Geiger (2018). "Learning priors for semantic 3D reconstruction". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[120] Choi, W., C. Pantofaru, and S. Savarese (2013). "A general framework for tracking multiple people from a moving camera". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 35(7): 1577–1591.

[121] Choi, W. (2015). "Near-online multi-target tracking with aggregated local flow descriptor". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[122] Choi, W. and S. Savarese (2012). "A unified framework for multi-target tracking and collective activity recognition". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[123] Chollet, F. (2017). "Xception: Deep learning with depthwise separable convolutions". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 1800–1807.

[124] Chu, P., H. Fan, C. C. Tan, and H. Ling (2019). "Online multi-object tracking with instance-aware tracker and dynamic model refreshment". In: *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*. 161–170.

[125] Chu, P. and H. Ling (2019). "FAMNet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking". arXiv: 1904.04989 [cs.CV].

[126] Chu, Q., W. Ouyang, H. Li, X. Wang, B. Liu, and N. Yu (2017). "Online multi-object tracking using cnn-based single object tracker with spatial-temporal attention mechanism". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 4846–4855.

[127] Ciresan, D. C., U. Meier, J. Masci, and J. Schmidhuber (2011). "A committee of neural networks for traffic sign classification". In: *International Joint Conference on Neural Networks (IJCNN)*. 1918–1921.

[128] Ciresan, D. C., U. Meier, J. Masci, and J. Schmidhuber (2012). "Multi-column deep neural network for traffic sign classification". *Neural Networks*. 32: 333–338.

[129] Codevilla, F., A. M. Lopez, V. Koltun, and A. Dosovitskiy (2018a). "On offline evaluation of vision-based driving models". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[130] Codevilla, F., M. Miiller, A. López, V. Koltun, and A. Dosovitskiy (2018b). "End-to-end driving via conditional imitation learning". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 1–9.

[131] Codevilla, F., E. Santana, A. M. López, and A. Gaidon (2019). "Exploring the limitations of behavior cloning for autonomous driving". arXiv: 1904.08980 [cs.CV].

[132] Collins, R. T. (1996). "A space-sweep approach to true multi-image matching". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 358–363.

[133] Cootes, T. F., C. J. Taylor, D. H. Cooper, and J. Graham (1995). "Active shape models-their training and application". *Computer Vision and Image Understanding (CVIU)*. 61(1): 38–59.

[134] Cordts, M., M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele (2016). "The cityscapes dataset for semantic urban scene understanding". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[135] Cordts, M., L. Schneider, M. Enzweiler, U. Franke, and S. Roth (2014). "Object-level priors for stixel generation". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[136] Cornelis, N., B. Leibe, K. Cornelis, and L. J. Van Gool (2008). "3D urban scene modeling integrating recognition and reconstruction". *International Journal of Computer Vision (IJCV)*. 78(2-3): 121–141.

[137] Corporation, M. M. (1998). "Mitsubishi motors develops 'new driver support system'". https://www.mitsubishi-motors.com/en/corporate/pressrelease/corporate/detail429.html. Online: accessed 17-May-2019.

[138] Costea, A. D., A. Petrovai, and S. Nedevschi (2018). "Fusion scheme for semantic and instance-level segmentation". In: *Proc. IEEE Conf. on Intelligent Transportation Systems (ITSC)*. 3469–3475.

[139] Costea, A. D., R. Varga, and S. Nedevschi (2017). "Fast boosting based detection using scale invariant multimodal multiresolution filtered features". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 993–1002.

[140] Crandall, D. J., A. Owens, N. Snavely, and D. Huttenlocher (2011). "Discrete-continuous optimization for large-scale structure from motion". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3001–3008.

[141] Cummins, M. and P. Newman (2008). "FAB-MAP: Probabilistic localization and mapping in the space of appearance". *International Journal of Robotics Research (IJRR)*. 27(6): 647–665.

[142] Curless, B. and M. Levoy (1996). "A volumetric method for building complex models from range images". In: *ACM Trans. on Graphics*. 303–312.

[143] Cvišic, I., J. Cesic, I. Markovic, and I. Petrovic (2017). "Soft-slam: Computationally efficient stereo visual slam for autonomous uavs". *Journal of Field Robotics (JFR)*.

[144] Cvisic, I. and I. Petrovic (2015). "Stereo odometry based on careful feature selection and tracking". *Proc. European Conf. on Mobile Robotics (ECMR)*. 35(4): 578–595.

[145] Dai, A. and M. Nießner (2018). "3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 458–474.

[146] Dai, J., K. He, Y. Li, S. Ren, and J. Sun (2016a). "Instance-sensitive fully convolutional networks". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 534–549.

[147] Dai, J., K. He, and J. Sun (2015). "Convolutional feature masking for joint object and stuff segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3992–4000.

[148] Dai, J., K. He, and J. Sun (2016b). "Instance-aware semantic segmentation via multi-task network cascades". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[149] Dai, J., Y. Li, K. He, and J. Sun (2016c). "R-FCN: Object detection via region-based fully convolutional networks". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[150] Dai, J., H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei (2017). "Deformable convolutional networks". *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[151] Daimler AG (2019). "Bosch and Daimler. Metropolis in California to become a pilot city for automated driving". https://www.daimler.com/innovation/case/autonomous/pilot-city-for-automated-driving.html. Online: accessed 18-October-2019.

[152] Dalal, N. and B. Triggs (2005). "Histograms of oriented gradients for human detection". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[153] Dame, A., V. Prisacariu, C. Ren, and I. Reid (2013). "Dense reconstruction using 3D object shape priors". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[154] DARPA (2014). "The DARPA grand challenge: Ten years later". https://www.darpa.mil/news-events/2014-03-13. Online: accessed 18-June-2019.

[155] Dehghan, A., S. M. Assari, and M. Shah (2015a). "GMMCP Tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[156] Dehghan, A., Y. Tian, P. H. S. Torr, and M. Shah (2015b). "Target identity-aware network flow for online multiple target tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[157] Deigmoeller, J. and J. Eggert (2016). "Stereo visual odometry without temporal filtering". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[158] Delaunoy, A. and M. Pollefeys (2014). "Photometric bundle adjustment for dense multi-view 3D modeling". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[159] Delaunoy, A. and E. Prados (2011). "Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3D reconstruction problems dealing with visibility". *International Journal of Computer Vision (IJCV)*. 95(2): 100–123.

[160] Dellaert, F., D. Fox, W. Burgard, and S. Thrun (1999). "Monte carlo localization for mobile robots". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[161] Dellaert, F. and M. Kaess (2006). "Square root SAM: Simultaneous localization and mapping via square root information smoothing". *International Journal of Robotics Research (IJRR)*. 25(12): 1181–1203.

[162] Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei (2009). "Imagenet: A large-scale hierarchical image database". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[163] Department of Motor Vehicles CA (2019). "Report of traffic collision involving an autonomous vehicle". https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/autonomousveh_ol316. Online: accessed 17-May-2019.

[164] Desai, C., D. Ramanan, and C. C. Fowlkes (2011). "Discriminative models for multi-class object layout". *International Journal of Computer Vision (IJCV)*. 95(1): 1–12.

[165] Deschaud, J. (2018). "IMLS-SLAM: Scan-to-model matching based on 3D data". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 2480–2485.

[166] Dewan, A., T. Caselitz, G. D. Tipaldi, and W. Burgard (2016a). "Motion-based detection and tracking in 3D LiDAR scans". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[167] Dewan, A., T. Caselitz, G. D. Tipaldi, and W. Burgard (2016b). "Rigid scene flow for 3D LiDAR scans". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[168] Dickmanns, E. D., R. Behringer, D. Dickmanns, T. Hildebrandt, M. Maurer, F. Thomanek, and J. Schiehlen (1994). "The seeing passenger car 'VaMoRs-P'". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[169] Dickmanns, E. D. and B. D. Mysliwetz (1992). "Recursive 3-D road and relative ego-state recognition". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 14(2): 199–213.

[170] Dickmanns, E. D. (1995). "Dynamic machine vision". http://dyna-vision.de/. Online: accessed 18-June-2019.

[171] Dickmanns, E. D. and V. Graefe (1988). "Dynamic monocular machine vision". *Machine Vision and Applications (MVA)*. 1(4): 223–240.

[172] Dickmanns, E. D., B. D. Mysliwetz, and T. Christians (1990). "An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles". *IEEE Trans. on Systems, Man and Cybernetics (TSMC)*. 20(6): 1273–1284.

[173] Dollar, P., C. Wojek, B. Schiele, and P. Perona (2011). "Pedestrian detection: An evaluation of the state of the art". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 34(4): 743–761.

[174] Dollár, P., R. Appel, S. J. Belongie, and P. Perona (2014). "Fast feature pyramids for object detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 36(8): 1532–1545.

[175] Dollár, P., Z. Tu, P. Perona, and S. Belongie (2009). "Integral channel features". In: *Proc. of the British Machine Vision Conf. (BMVC)*. BMVC Press.

[176] Dollar, P., C. Wojek, B. Schiele, and P. Perona (2009). "Pedestrian detection: A benchmark". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[177] Dollár, P., C. Wojek, B. Schiele, and P. Perona (2012). "Pedestrian detection: An evaluation of the state of the art". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 34(4): 743–761.

[178] Dosovitskiy, A., P. Fischer, E. Ilg, P. Haeusser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox (2015). "FlowNet: Learning optical flow with convolutional networks". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[179] Dosovitskiy, A., G. Ros, F. Codevilla, A. Lopez, and V. Koltun (2017a). "CARLA". http://carla.org/. Online: accessed 18-October-2019.

[180] Dosovitskiy, A., G. Ros, F. Codevilla, A. Lopez, and V. Koltun (2017b). "CARLA: An open urban driving simulator". In: *Proc. Conf. on Robot Learning (CoRL)*.

[181] Drory, A., C. Haubold, S. Avidan, and F. A. Hamprecht (2014). "Semi-global matching: A principled derivation in terms of message passing". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[182] Du, X., M. H. Ang, S. Karaman, and D. Rus (2018). "A general pipeline for 3D detection of vehicles". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 3194–3200.

[183] Duan, L. and F. Lafarge (2016). "Towards large-scale city reconstruction from satellites". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[184] Dubé, R., D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena (2017). "SegMatch: Segment based place recognition in 3D point clouds". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. IEEE. 5266–5272.

[185] Engel, J., J. Sturm, and D. Cremers (2013). "Semi-dense visual odometry for a monocular camera". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[186] Engel, J., V. Koltun, and D. Cremers (2018). "Direct sparse odometry". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 40: 611–625.

[187] Engel, J., T. Schöps, and D. Cremers (2014). "LSD-SLAM: Large-scale direct monocular SLAM". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[188] Engel, J., J. Stückler, and D. Cremers (2015). "Large-scale direct SLAM with stereo cameras". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS).*

[189] Engelcke, M., D. Rao, D. Z. Wang, C. H. Tong, and I. Posner (2017). "Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).* 1355–1361.

[190] Enzweiler, M. and D. M. Gavrila (2009). "Monocular pedestrian detection: Survey and experiments". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 31: 2179–2195.

[191] Enzweiler, M. and D. Gavrila (2008). "A mixed generative-discriminative framework for pedestrian classification". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[192] Enzweiler, M. and D. M. Gavrila (2011). "A multilevel mixture-of-experts framework for pedestrian classification". *IEEE Trans. on Image Processing (TIP).* 20(10): 2967–2979.

[193] Erbs, F., B. Schwarz, and U. Franke (2012). "Stixmentation – Probabilistic stixel based traffic scene labeling". In: *Proc. of the British Machine Vision Conf. (BMVC).*

[194] Erbs, F., B. Schwarz, and U. Franke (2013). "From stixels to objects – A conditional random field based approach". In: *Proc. IEEE Intelligent Vehicles Symposium (IV).*

[195] Ess, A., B. Leibe, K. Schindler, and L. Van Gool (2008). "A mobile vision system for robust multi-person tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[196] Ess, A., B. Leibe, K. Schindler, and L. Van Gool (2009a). "Robust multi-person tracking from a mobile platform". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 31: 1831–1846.

[197] Ess, A., T. Mueller, H. Grabner, and L. van Gool (2009b). "Segmentation-based urban traffic scene understanding". In: *Proc. of the British Machine Vision Conf. (BMVC).*

[198] Everingham, M., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman (2010). "The pascal visual object classes (VOC) challenge". *International Journal of Computer Vision (IJCV).* 88(2): 303–338.

[199] Fanani, N., M. Ochs, H. Bradler, and R. Mester (2016). "Keypoint trajectory estimation using propagation based tracking". In: *Proc. IEEE Intelligent Vehicles Symposium (IV).* 933–939.

[200] Fanani, N., A. Sturck, M. Barnada, and R. Mester (2017a). "Multimodal scale estimation for monocular visual odometry". In: *Proc. IEEE Intelligent Vehicles Symposium (IV).* 1714–1721.

[201] Fanani, N., A. Sturck, M. Ochs, H. Bradler, and R. Mester (2017b). "Predictive monocular odometry (PMO): What is possible without RANSAC and multiframe bundle adjustment?" *Image and Vision Computing (IVC).* 68: 3–13.

[202] Farneback, G. (2003). "Two-frame motion estimation based on polynomial expansion". In: *Scandinavian Conference on Image Analysis (SCIA).*

[203] Faugeras, O. D. and R. Keriven (1998). "Variational principles, surface evolution, PDEs, level set methods, and the stereo problem". *IEEE Trans. on Image Processing (TIP).* 7(3): 336–344.

[204] Feichtenhofer, C., A. Pinz, and A. Zisserman (2017). "Detect to track and track to detect". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[205] Felzenszwalb, P. F., D. A. McAllester, and D. Ramanan (2008). "A discriminatively trained, multiscale, deformable part model". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[206] Felzenszwalb, P. and D. Huttenlocher (2006). "Efficient belief propagation for early vision". *International Journal of Computer Vision (IJCV).* 70(1): 41–54.

[207] Ferryman, J. and A. Shahrokni (2009). "PETS2009: Dataset and challenge". In: *Performance Evaluation of Tracking and Surveillance.* 1–6.

[208] Floros, G. and B. Leibe (2012). “Joint 2D-3D temporally consistent semantic segmentation of street scenes”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[209] Frahm, J.-M., P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys (2010a). “Building Rome on a cloudless day”. In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[210] Frahm, J.-M., M. Pollefeys, S. Lazebnik, D. Gallup, B. Clipp, R. Raguram, C. Wu, C. Zach, and T. Johnson (2010b). “Fast robust large-scale mapping from video and internet photo collections”. *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS).* 65: 538–550.

[211] Franke, U., S. Mehring, A. Suissa, and S. Hahn (1994). “The Daimler-Benz steering assistant: A spin-off from autonomous driving”. In: *Proc. IEEE Intelligent Vehicles Symposium (IV).*

[212] Franke, U., D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, and C. Wöhler (1998). “Autonomous driving goes downtown”. *Intelligent Systems (IS).* 13(6): 40–48.

[213] Franke, U., C. Rabe, H. Badino, and S. Gehrig (2005). “6D-Vision: Fusion of stereo and motion for robust environment perception”. In: *Proc. of the DAGM Symposium on Pattern Recognition (DAGM).*

[214] Fraundorfer, F. and D. Scaramuzza (2011). “Visual odometry: Part II – Matching, robustness, and applications”. *Robotics and Automation Magazine (RAM).*

[215] Fritsch, J., T. Kuehnl, and A. Geiger (2013). “A new performance measure and evaluation benchmark for road detection algorithms”. In: *Proc. IEEE Conf. on Intelligent Transportation Systems (ITSC).*

[216] Frossard, D. and R. Urtasun (2018). “End-to-end learning of multi-sensor 3D tracking by detection”. In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).*

[217] Frost, D. P., O. Kähler, and D. W. Murray (2016). “Object-aware bundle adjustment for correcting monocular scale drift”. In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).*

[218] Früh, C., S. Jain, and A. Zakhor (2005). "Data processing algorithms for generating textured 3D building facade meshes from laser scans and camera images". *International Journal of Computer Vision (IJCV)*. 61(2): 159–184.

[219] Fuhrmann, S., F. Langguth, and M. Goesele (2014). "MVE – A multi-view reconstruction environment". In: *Eurographics Workshop on Graphics and Cultural Heritage (GCH)*. 11–18.

[220] Furgale, P. T., U. Schwesinger, M. Rufli, W. Derendarz, H. Grimmett, P. Mühlfellner, S. Wonneberger, J. Timpner, S. Rottmann, B. Li, B. Schmidt, T. Nguyen, E. Cardarelli, S. Cattani, S. Bruning, S. Horstmann, M. Stellmacher, H. Mielenz, K. Köser, M. Beermann, C. Hane, L. Heng, G. H. Lee, F. Fraundorfer, R. Iser, R. Triebel, I. Posner, P. Newman, L. C. Wolf, M. Pollefeys, S. Brosig, J. Effertz, C. Pradalier, and R. Siegwart (2013). "Toward automated driving in cities using close-to-market sensors: An overview of the V-Charge Project". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[221] Furukawa, Y. and J. Ponce (2010). "Accurate, dense, and robust multi-view stereopsis". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 32(8): 1362–1376.

[222] Gadde, R., V. Jampani, M. Kiefel, D. Kappler, and P. V. Gehler (2016). "Superpixel convolutional networks using bilateral inceptions". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[223] Gadde, R., V. Jampani, R. Marlet, and P. V. Gehler (2018). "Efficient 2D and 3D facade segmentation using auto-context". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 40(5): 1273–1280.

[224] Gadot, D. and L. Wolf (2016). "PatchBatch: A batch augmented loss for optical flow". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[225] Gaidon, A., Q. Wang, Y. Cabon, and E. Vig (2016a). "Virtual KITTI". https://europe.naverlabs.com/research/computer-vision/proxy-virtual-worlds. Online: accessed 18-October-2019.

[226] Gaidon, A., Q. Wang, Y. Cabon, and E. Vig (2016b). "Virtual worlds as proxy for multi-object tracking analysis". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[227] Gallego, G., H. Rebecq, and D. Scaramuzza (2018). "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[228] Galleguillos, C., A. Rabinovich, and S. J. Belongie (2008). "Object categorization using co-occurrence, location and appearance". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[229] Galliani, S., K. Lasinger, and K. Schindler (2015). "Massively parallel multiview stereopsis by surface normal diffusion". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[230] Gallup, D., J. M. Frahm, P. Mordohai, and M. Pollefeys (2008). "Variable baseline/resolution stereo". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[231] Gallup, D., J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys (2007). "Real-time plane-sweeping stereo with multiple sweeping directions". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[232] Gallup, D., J.-M. Frahm, and M. Pollefeys (2010). "Piecewise planar and non-planar stereo for urban scene reconstruction". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[233] Gálvez-López, D. and J. D. Tardós (2012). "Bags of binary words for fast place recognition in image sequences". *IEEE Trans. on Robotics.* 28(5): 1188–1197.

[234] García, Á. A., J. A. Álvarez-García, and L. M. Soria-Morillo (2018). "Evaluation of deep neural networks for traffic sign detection systems". *Neurocomputing.* 316: 332–344.

[235] Garcia-Garcia, A., F. Gomez-Donoso, J. G. Rodríguez, S. Orts-Escolano, M. Cazorla, and J. A. López (2016). "PointNet: A 3D convolutional neural network for real-time object class recognition". In: *International Joint Conference on Neural Networks (IJCNN)*.

[236] Gast, J. and S. Roth (2018). "Lightweight probabilistic deep networks". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3369–3378.

[237] Gavrila, D. M. and S. Munder (2007). "Multi-cue pedestrian detection and tracking from a moving vehicle". *International Journal of Computer Vision (IJCV)*. 73: 41–59.

[238] Gehrig, D., H. Rebecq, G. Gallego, and D. Scaramuzza (2018). "Asynchronous, photometric feature tracking using events and frames". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[239] Gehrig, S. K., F. Eberli, and T. Meyer (2009). "A real-time low-power stereo vision engine using semi-global matching". In: *Proc. of the International Conf. on Computer Vision Systems (ICVS)*.

[240] Geiger, A. (2009). "Monocular road mosaicing for urban environments". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[241] Geiger, A., M. Lauer, F. Moosmann, B. Ranft, H. Rapp, C. Stiller, and J. Ziegler (2012a). "Team AnnieWAY's entry to the grand cooperative driving challenge 2011". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 13(3): 1008–1017.

[242] Geiger, A., M. Lauer, C. Wojek, C. Stiller, and R. Urtasun (2014). "3D traffic scene understanding from movable platforms". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 36(5): 1012–1025.

[243] Geiger, A., P. Lenz, C. Stiller, and R. Urtasun (2013). "Vision meets robotics: The KITTI dataset". *International Journal of Robotics Research (IJRR)*. 32(11): 1231–1237.

[244] Geiger, A., P. Lenz, and R. Urtasun (2012b). “Are we ready for autonomous driving? The KITTI vision benchmark suite”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[245] Geiger, A., F. Moosmann, O. Car, and B. Schuster (2012c). “Automatic calibration of range and camera sensors using a single shot”. In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).*

[246] Geiger, A., M. Roser, and R. Urtasun (2010). “Efficient large-scale stereo matching”. In: *Proc. of the Asian Conf. on Computer Vision (ACCV).*

[247] Geiger, A., J. Ziegler, and C. Stiller (2011). “StereoScan: Dense 3D reconstruction in real-time”. In: *Proc. IEEE Intelligent Vehicles Symposium (IV).*

[248] Geronimo, D., A. M. Lopez, A. D. Sappa, and T. Graf (2010). “Survey on pedestrian detection for advanced driver assistance systems”. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 32(7): 1239–1258.

[249] Geyer, C. and K. Daniilidis (2000). “A unifying theory for central panoramic systems and practical implications”. In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[250] Ghiasi, G. and C. C. Fowlkes (2016). “Laplacian pyramid reconstruction and refinement for semantic segmentation”. In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[251] Giebel, J., D. Gavrila, and C. Schnörr (2004). “A Bayesian framework for multi-cue 3D object tracking”. In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[252] Girshick, R. B. (2015). “Fast R-CNN”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[253] Girshick, R., J. Donahue, T. Darrell, and J. Malik (2014). “Rich feature hierarchies for accurate object detection and semantic segmentation”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[254] Gkioxari, G., B. Hariharan, R. Girshick, and J. Malik (2014). "Using k-poselets for detecting people and localizing their keypoints". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[255] González, A., D. Vázquez, A. M. Lóopez, and J. Amores (2016). "On-board object detection: Multicue, multimodal, and multiview random forest of local experts". *IEEE Trans. on Cybernetics*. 47(11): 3980–3990.

[256] González-Barbosa, J. and S. Lacroix (2005). "Fast dense panoramic stereovision". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 1210–1215.

[257] Gräter, J., A. Wilczynski, and M. Lauer (2018). "LIMO: Lidar-monocular visual odometry". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*. 7872–7879.

[258] Grimmett, H., M. Bürki, L. M. Paz, P. Pinies, P. T. Furgale, I. Posner, and P. Newman (2015). "Integrating metric and semantic maps for vision-only automated parking". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[259] Grisleri, P. and I. Fedriga (2010). "The BRAiVE platform". In: *Proc. of the IFAC Symposium on Intelligent Autonomous Vehicles (IFAC)*.

[260] Guindel, C., D. Martín, and J. M. Armingol (2018). "Fast joint object detection and viewpoint estimation for traffic scene understanding". *IEEE Intelligent Transportation Systems Magazine (ITSM)*. 10(4): 74–86.

[261] Guivant, J. and E. Nebot (2006). "Victoria park dataset". http://www-personal.acfr.usyd.edu.au/nebot/victoria_park.htm. Online: accessed 8-April-2019.

[262] Güney, F. and A. Geiger (2015). "Displets: Resolving stereo ambiguities using object knowledge". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[263] Güney, F. and A. Geiger (2016). "Deep discrete flow". In: *Proc. of the Asian Conf. on Computer Vision (ACCV)*.

[264] Günyel, B., R. Benenson, R. Timofte, and L. J. Van Gool (2012). "Stixels motion estimation without optical flow computation". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[265] Haala, N. and K.-H. Anders (1997). "Acquisition of 3D urban models by analysis of aerial images, digital surface models, and existing 2D building information". In: *Proc. of the SPIE Conf. Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision IIII*. Vol. 3072. International Society for Optics and Photonics. 212–222.

[266] Hackel, T., N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys (2017). "Semantic3D.net: A new large-scale point cloud classification benchmark". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences (APRS)*.

[267] Hackel, T., J. D. Wegner, and K. Schindler (2016). "Fast semantic segmentation of 3D point clouds with strongly varying density". *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences (APRS)*. III-3: 177–184.

[268] Haene, C., N. Savinov, and M. Pollefeys (2014). "Class specific 3D object shape priors using surface normals". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[269] Haene, C., C. Zach, A. Cohen, R. Angst, and M. Pollefeys (2013). "Joint 3D scene reconstruction and class segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[270] Haene, C., C. Zach, B. Zeisl, and M. Pollefeys (2012). "A patch prior for dense 3D reconstruction in man-made environments". In: *Proc. of the International Conf. on 3D Digital Imaging, Modeling, Data Processing, Visualization and Transmission (THREEDIMPVT)*.

[271] Han, X., T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg (2015). "MatchNet: Unifying feature and metric learning for patch-based matching". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[272] Häne, C., L. Heng, G. H. Lee, A. Sizov, and M. Pollefeys (2014). "Real-time direct dense matching on fisheye images using plane-sweeping stereo". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[273] Häne, C., T. Sattler, and M. Pollefeys (2015). “Obstacle detection for self-driving cars using only monocular cameras and wheel odometry”. In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[274] Hanson, A. (1978). *Computer Vision Systems*. Elsevier.

[275] Hariharan, B., P. A. Arbeláez, R. B. Girshick, and J. Malik (2014). “Simultaneous detection and segmentation”. In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[276] Hariharan, B., P. A. Arbeláez, R. B. Girshick, and J. Malik (2015). “Hypercolumns for object segmentation and fine-grained localization”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 447–456.

[277] Hariharan, B., P. Arbelaez, L. D. Bourdev, S. Maji, and J. Malik (2011). “Semantic contours from inverse detectors”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 991–998.

[278] Hartley, R. I. (1995). “In defence of the 8-point algorithm”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[279] Hartley, R. I. and A. Zisserman (2004). *Multiple View Geometry in Computer Vision*. Second edition. Cambridge University Press.

[280] Hartmann, W., S. Galliani, M. Havlena, L. Van Gool, and K. Schindler (2017). “Learned multi-patch similarity”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[281] Hayder, Z., X. He, and M. Salzmann (2017). “Boundary-aware instance segmentation”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 587–595.

[282] Hays, J. and A. A. Efros (2008). “im2gps: Estimating geographic information from a single image”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[283] Hazirbas, C., L. Ma, C. Domokos, and D. Cremers (2016). “FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture”. In: *Proc. of the Asian Conf. on Computer Vision (ACCV)*.

[284] He, K., G. Gkioxari, P. Dollár, and R. B. Girshick (2017). "Mask R-CNN". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 2980–2988.

[285] He, K., X. Zhang, S. Ren, and J. Sun (2014). "Spatial pyramid pooling in deep convolutional networks for visual recognition". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[286] He, K., X. Zhang, S. Ren, and J. Sun (2016a). "Deep residual learning for image recognition". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[287] He, K., X. Zhang, S. Ren, and J. Sun (2016b). "Identity mappings in deep residual networks". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 630–645.

[288] He, X., R. S. Zemel, and M. A. Carreira-Perpinan (2004). "Multiscale conditional random fields for image labeling". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[289] He, X., R. S. Zemel, and D. Ray (2006). "Learning and incorporating top-down cues in image segmentation". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[290] Heeger, D. J. (1988). "Optical flow using spatiotemporal filters". *International Journal of Computer Vision (IJCV)*. 1(4): 279–302.

[291] Heikkila, J. and O. Silven (1997). "A four-step camera calibration procedure with implicit image correction". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[292] Henaff, M., A. Canziani, and Y. LeCun (2019). "Model-predictive policy learning with uncertainty regularization for driving in dense traffic". arXiv: 1901.02705 [cs.LG].

[293] Heng, L., P. T. Furgale, and M. Pollefeys (2015). "Leveraging image-based localization for infrastructure-based calibration of a multi-camera rig". *Journal of Field Robotics (JFR)*. 32(5): 775–802.

[294] Heng, L., B. Li, and M. Pollefeys (2013). "CamOdoCal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[295] Henschel, R., L. Leal-Taixé, D. Cremers, and B. Rosenhahn (2018). "Fusion of head and full-body detectors for multi-object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*. 1428–1437.

[296] Henschel, R., Y. Zhou, and B. Rosenhahn (2019). "Fusion of head and full-body detectors for multi-object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[297] Hirschmüller, H. and D. Scharstein (2007). "Evaluation of cost functions for stereo matching". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 1–8.

[298] Hirschmüller, H. (2008). "Stereo processing by semiglobal matching and mutual information". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 30(2): 328–341.

[299] Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". *Neural Computation*. 9(8): 1735–1780.

[300] Hoiem, D., A. Efros, and M. Hebert (2008). "Putting objects in perspective". *International Journal of Computer Vision (IJCV)*. 80: 3–15.

[301] Hoiem, D., A. A. Efros, and M. Hebert (2007). "Recovering surface layout from an image". *International Journal of Computer Vision (IJCV)*. 75(1): 151–172.

[302] Horn, B. K. P. and B. G. Schunck (1981). "Determining optical flow". *Artificial Intelligence (AI)*. 17(1–3): 185–203.

[303] Houben, S., J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel (2013). "Detection of traffic signs in real-world images: The German traffic sign detection benchmark". In: *International Joint Conference on Neural Networks (IJCNN)*. 1–8.

[304] Hu, H., D. Munoz, J. A. Bagnell, and M. Hebert (2013). "Efficient 3-D scene analysis from streaming data". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 2297–2304.

[305] Hu, H., Q. Cai, D. Wang, J. Lin, M. Sun, P. Krähenbühl, T. Darrell, and F. Yu (2019). "Joint monocular 3D vehicle detection and tracking". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[306] Hu, H.-N., Y.-C. Lin, M.-Y. Liu, H.-T. Cheng, Y.-J. Chang, and M. Sun (2017). “Deep 360 pilot: Learning a deep agent for piloting through 360° sports video”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[307] Hu, X., X. Xu, Y. Xiao, H. Chen, S. He, J. Qin, and P.-A. Heng (2018). “SINet: A scale-insensitive convolutional neural network for fast vehicle detection”. *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 20(3): 1010–1019.

[308] Huang, C., B. Wu, and R. Nevatia (2008). “Robust object tracking by hierarchical association of detection responses”. In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[309] Huang, G., Z. Liu, L. van der Maaten, and K. Q. Weinberger (2017a). “Densely connected convolutional networks”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2261–2269.

[310] Huang, G., Z. Liu, L. van der Maaten, and K. Q. Weinberger (2017b). “Densely connected convolutional networks”. https://www.youtube.com/watch?v=-W6y8xnd--U. Online: accessed 18-February-2020.

[311] Huang, J. and S. You (2016). “Point cloud labeling using 3D convolutional neural network”. In: *Proc. of the International Conf. on Pattern Recognition (ICPR)*.

[312] Huang, P., K. Matzen, J. Kopf, N. Ahuja, and J. Huang (2018a). “DeepMVS: Learning multi-view stereopsis”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2821–2830.

[313] Huang, X., X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang (2018b). “The apolloscape dataset for autonomous driving”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[314] Huguet, F. and F. Devernay (2007). “A variational method for scene flow estimation from stereo sequences”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[315] Hui, T., X. Tang, and C. C. Loy (2018). “LiteFlowNet: A lightweight convolutional neural network for optical flow estimation”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[316] Hur, J. and S. Roth (2017). “MirrorFlow: Exploiting symmetries in joint optical flow and occlusion estimation”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[317] Ilg, E., Ö. Çiçek, S. Galesso, A. Klein, O. Makansi, F. Hutter, and T. Brox (2018). “Uncertainty estimates and multi-hypotheses networks for optical flow”. In: *Proc. of the European Conf. on Computer Vision (ECCV), Part VII.* 677–693.

[318] Ilg, E., N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox (2017). “FlowNet 2.0: Evolution of optical flow estimation with deep networks”. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[319] Ioffe, S. and C. Szegedy (2015). “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *Proc. of the International Conf. on Machine Learning (ICML).*

[320] Irschara, A., C. Zach, J. Frahm, and H. Bischof (2009). “From structure-from-motion point clouds to fast location recognition”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 2599–2606.

[321] Jampani, V., M. Kiefel, and P. V. Gehler (2016). “Learning sparse high dimensional filters: Image filtering, dense CRFs and bilateral neural networks”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[322] Janai, J., F. Güney, A. Ranjan, M. Black, and A. Geiger (2018). “Unsupervised learning of multi-frame optical flow with occlusions”. In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[323] Janai, J., F. Güney, J. Wulff, M. Black, and A. Geiger (2017). “Slow flow: Exploiting high-speed cameras for accurate and diverse optical flow reference data”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[324] Jancosek, M. and T. Pajdla (2011). "Multi-view reconstruction preserving weakly-supported surfaces". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[325] Jégou, S., M. Drozdzal, D. Vázquez, A. Romero, and Y. Bengio (2017). "The one hundred layers Tiramisu: Fully convolutional DenseNets for semantic segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*. 1175–1183.

[326] Jensen, R. R., A. L. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs (2014). "Large scale multi-view stereopsis evaluation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[327] Ji, K., H. Chen, H. Di, J. Gong, G. Xiong, J. Qi, and T. Yi (2018). "CPFG-SLAM: A robust simultaneous localization and mapping based on LIDAR in off-road environment". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. 650–655.

[328] Ji, M., J. Gall, H. Zheng, Y. Liu, and L. Fang (2017). "SurfaceNet: An end-to-end 3D neural network for multiview stereopsis". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[329] Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell (2014). "Caffe: Convolutional architecture for fast feature embedding". In: *Proc. of the International Conf. on Multimedia (ICM)*.

[330] Jiang, H., S. Fels, and J. J. Little (2007). "A linear programming approach for multiple object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[331] Jin, J., K. Fu, and C. Zhang (2014). "Traffic sign recognition with hinge loss trained convolutional neural networks". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 15(5): 1991–2000.

[332] Julier, S. J. and J. K. Uhlmann (2001). "A counter example to the theory of simultaneous localization and map building". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[333] Jung, S. and K. Hong (2017). "Deep network aided by guiding network for pedestrian detection". *Pattern Recognition Letters*. 90: 43–49.

[334] Kaess, M., H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert (2012). "iSAM2: Incremental smoothing and mapping using the Bayes tree". *International Journal of Robotics Research (IJRR)*. 31(2): 217–236.

[335] Kaess, M., A. Ranganathan, and F. Dellaert (2008). "iSAM: Incremental smoothing and mapping". *IEEE Trans. on Robotics*. 24(6): 1365–1378.

[336] Kaminsky, R. S., N. Snavely, S. M. Seitz, and R. Szeliski (2009). "Alignment of 3D point clouds to overhead images". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[337] Kang, K., H. Li, T. Xiao, W. Ouyang, J. Yan, X. Liu, and X. Wang (2017). "Object detection in videos with tubelet proposal networks". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[338] Kang, K., W. Ouyang, H. Li, and X. Wang (2016). "Object detection from video tubelets with convolutional neural networks". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[339] Kang, Y., H. Yin, and C. Berger (2019). "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments". *Proc. IEEE Transactions on Intelligent Vehicles (T-IV)*. 4(2): 171–185.

[340] Kar, A., C. Häne, and J. Malik (2017). "Learning a multi-view stereo machine". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[341] Kassir, A. and T. Peynot (2010). "Reliable automatic camera-laser calibration". In: *Proc. IEEE Australasian Conf. on Robotics and Automation (ACRA)*.

[342] Keller, C. G., M. Enzweiler, M. Rohrbach, D. F. Llorca, C. Schnörr, and D. M. Gavrila (2011). "The benefits of dense stereo for pedestrian detection". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 12(4): 1096–1106.

[343] Kendall, A., Y. Gal, and R. Cipolla (2018a). "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 7482–7491.

[344] Kendall, A., M. Grimes, and R. Cipolla (2015). "PoseNet: A convolutional network for real-time 6-DOF camera relocalization". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[345] Kendall, A., J. Hawke, D. Janz, P. Mazur, D. Reda, J. M. Allen, V. D. Lam, A. Bewley, and A. Shah (2018b). "Learning to drive in a day". arXiv: 1807.00412 [cs.LG].

[346] Kendall, A., H. Martirosyan, S. Dasgupta, and P. Henry (2017). "End-to-end learning of geometry and context for deep stereo regression". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[347] Kerl, C., J. Sturm, and D. Cremers (2013). "Robust odometry estimation for RGB-D cameras". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 3748–3754.

[348] Kesten, R., M. Usman, J. Houston, T. Pandya, K. Nadhamuni, A. Ferreira, M. Yuan, B. Low, A. Jain, P. Ondruska, S. Omari, S. Shah, A. Kulkarni, A. Kazakova, C. Tao, L. Platinsky, W. Jiang, and V. Shet (2019). "Lyft level 5 AV dataset 2019". https://level5.lyft.com/dataset/. Online: accessed 25-March-2020.

[349] Keuper, M., S. Tang, B. Andres, T. Brox, and B. Schiele (2018). "Motion segmentation & multiple object tracking by correlation co-clustering". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 42(1): 140–153.

[350] Kim, C., F. Li, A. Ciptadi, and J. M. Rehg (2015). "Multiple hypothesis tracking revisited". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[351] Kim, C., F. Li, and J. M. Rehg (2018). "Multi-object tracking with neural gating using bilinear LSTM". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[352] Kirillov, A., R. B. Girshick, K. He, and P. Dollár (2019a). "Panoptic feature pyramid networks". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[353] Kirillov, A., K. He, R. B. Girshick, C. Rother, and P. Dollár (2019b). "Panoptic segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 9404–9413.

[354] Kirillov, A., E. Levinkov, B. Andres, B. Savchynskyy, and C. Rother (2017). "InstanceCut: From edges to instances with MultiCut". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 7322–7331.

[355] Kitt, B., A. Geiger, and H. Lategahn (2010). "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme". In: *Proc. IEEE Intelligent Vehicles Symposium (IV).*

[356] Klein, G. and D. W. Murray (2007). "Parallel tracking and mapping for small AR workspaces". In: *Proc. of the International Symposium on Mixed and Augmented Reality (ISMAR)*. 225–234.

[357] Klette, R. (2015). "Vision-based driver assistance systems". *Tech. rep.* CITR, Auckland, New Zealand.

[358] Klingner, B. M., D. Martin, and J. Roseborough (2013). "Street view motion-from-structure-from-motion". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 953–960.

[359] Knapitsch, A., J. Park, Q.-Y. Zhou, and V. Koltun (2017). "Tanks and temples: Benchmarking large-scale scene reconstruction". *ACM Trans. on Graphics*. 36(4).

[360] Kohli, P., L. Ladicky, and P. H. S. Torr (2009). "Robust higher order potentials for enforcing label consistency". *International Journal of Computer Vision (IJCV)*. 82(3): 302–324.

[361] Kolmogorov, V. (2006). "Convergent tree-reweighted message passing for energy minimization". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 28(10): 1568–1583.

[362] Konda, V. R. and J. N. Tsitsiklis (1999). "Actor-critic algorithms". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 1008–1014.

[363] Kondermann, C., D. Kondermann, B. Jähne, and C. S. Garbe (2007). "An adaptive confidence measure for optical flows based on linear subspace projections". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*. 132–141.

[364] Kondermann, C., R. Mester, and C. S. Garbe (2008). "A statistical confidence measure for optical flows". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[365] Kondermann, D., R. Nair, K. Honauer, K. Krispin, J. Andrulis, A. Brock, B. Gussefeld, M. Rahimimoghaddam, S. Hofmann, C. Brenner, and B. Jahne (2016). "The HCI benchmark suite: Stereo and flow ground truth with uncertainties for urban autonomous driving". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[366] Krähenbühl, P. and V. Koltun (2011). "Efficient inference in fully connected CRFs with Gaussian edge potentials". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[367] Krešo, I. and S. Šegvić (2015). "Improving the egomotion estimation by correcting the calibration bias". In: *Proc. of the Conf. on Computer Vision Theory and Applications (VISAPP)*.

[368] Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). "ImageNet classification with deep convolutional neural networks". In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[369] Kroeger, T., R. Timofte, D. Dai, and L. Van Gool (2016). "Fast optical flow using dense inverse search". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 471–488.

[370] Ku, J., M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander (2018). "Joint 3D proposal generation and object detection from view aggregation". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*. 1–8.

[371] Ku, J., A. D. Pon, and S. L. Waslander (2019). "Monocular 3D object detection leveraging accurate proposals and shape reconstruction". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[372] Kuehnl, T., F. Kummert, and J. Fritsch (2012). "Spatial ray features for real-time ego-lane extraction". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[373] Kuettel, D., M. D. Breitenstein, L. Van Gool, and V. Ferrari (2010). "What's going on?: Discovering spatio-temporal dependencies in dynamic scenes". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[374] Kuhn, A., H. Hirschmüller, D. Scharstein, and H. Mayer (2017). "A TV prior for high-quality scalable multi-view stereo reconstruction". *International Journal of Computer Vision (IJCV)*. 124(1): 2–17.

[375] Kumar, S. and M. Hebert (2005). "A hierarchical field framework for unified context-based classification". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[376] Kümmerle, R., G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard (2011). "G^2o: A general framework for graph optimization". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 3607–3613.

[377] Kundu, A., Y. Li, F. Dellaert, F. Li, and J. Rehg (2014). "Joint semantic segmentation and 3D reconstruction from monocular video". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[378] Kundu, A., Y. Li, and J. M. Rehg (2018). "3D-RCNN: Instance-level 3D object reconstruction via render-and-compare". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3559–3568.

[379] Kundu, A., V. Vineet, and V. Koltun (2016). "Feature space optimization for semantic video segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[380] Kuschk, G. and D. Cremers (2013). "Fast and accurate large-scale stereo reconstruction using variational methods". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV) Workshops*.

[381] Kutulakos, K. N. and S. M. Seitz (2000). "A theory of shape by space carving". *International Journal of Computer Vision (IJCV)*. 38(3): 199–218.

[382] Kybic, J. and C. Nieuwenhuis (2011). "Bootstrap optical flow confidence and uncertainty measure". *Computer Vision and Image Understanding (CVIU)*. 115(10): 1449–1462.

[383] Labatut, P., J. Pons, and R. Keriven (2007). "Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).* 1–8.

[384] Laddha, A., M. K. Kocamaz, L. E. Navarro-Serment, and M. Hebert (2016). "Map-supervised road detection". In: *Proc. IEEE Intelligent Vehicles Symposium (IV).*

[385] Ladicky, L., C. Russell, P. Kohli, and P. Torr (2009). "Associative hierarchical CRFs for object class image segmentation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[386] Ladicky, L., C. Russell, P. Kohli, and P. H. Torr (2010). "Graph cut based inference with co-occurrence statistics". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[387] Ladicky, L., C. Russell, P. Kohli, and P. H. S. Torr (2014). "Associative hierarchical random fields". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 36(6): 1056–1077.

[388] Lafarge, F., X. Descombes, J. Zerubia, and M. P. Deseilligny (2010). "Structural approach for building reconstruction from a single DSM". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 32(1): 135–147.

[389] Lafarge, F., R. Keriven, M. Bredif, and H.-H. Vu (2013). "A hybrid multiview stereo algorithm for modeling urban scenes". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 35(1): 5–17.

[390] Lafarge, F. and C. Mallet (2012). "Creating large-scale city models from 3D-point clouds: A robust approach with hybrid representation". *International Journal of Computer Vision (IJCV).* 99(1): 69–85.

[391] Laskey, M., J. Lee, R. Fox, A. D. Dragan, and K. Goldberg (2017). "DART: Noise injection for robust imitation learning". In: *Proc. Conf. on Robot Learning (CoRL).* 143–156.

[392] Lategahn, H., A. Geiger, and B. Kitt (2011). "Visual SLAM for autonomous ground vehicles". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).*

[393] Lauer, M. (2011). "Grand cooperative driving challenge 2011". *Proc. IEEE Intelligent Transportation Systems Magazine (ITSM)*. 3(3): 38–40.

[394] Law, H. and J. Deng (2018). "CornerNet: Detecting objects as paired keypoints". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 765–781.

[395] Le, N., A. Heili, and J.-M. Odobez (2016). "Long-term time-sensitive costs for CRF-based tracking by detection". In: *Proc. of the European Conf. on Computer Vision (ECCV) Workshops*.

[396] Leal-Taixé, L., C. Canton-Ferrer, and K. Schindler (2016). "Learning by tracking: Siamese CNN for robust target Association". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[397] Leal-Taixé, L., A. Milan, I. D. Reid, S. Roth, and K. Schindler (2015). "MOTChallenge 2015: Towards a benchmark for multi-target tracking". arXiv: 1504.01942 [cs.CV].

[398] Lee, G. H., F. Fraundorfer, and M. Pollefeys (2013a). "Motion estimation for self-driving cars with a generalized camera". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[399] Lee, G. H., F. Fraundorfer, and M. Pollefeys (2013b). "Structureless pose-graph loop-closure with a multi-camera system on a self-driving car". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[400] Lee, G. H., M. Pollefeys, and F. Fraundorfer (2014). "Relative pose estimation for a multi-camera system with known vertical direction". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[401] Lee, S., J. Kim, J. S. Yoon, S. Shin, O. Bailo, N. Kim, T. Lee, H. S. Hong, S. Han, and I. S. Kweon (2017). "VPGNet: Vanishing point guided network for lane and road marking detection and recognition". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 1965–1973.

[402] Leibe, B., N. Cornelis, K. Cornelis, and L. Van Gool (2007). "Dynamic 3D scene analysis from a moving vehicle". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[403] Leibe, B., A. Leonardis, and B. Schiele (2008a). "Robust object detection with interleaved categorization and segmentation". *International Journal of Computer Vision (IJCV).* 77(1-3): 259–289.

[404] Leibe, B., K. Schindler, N. Cornelis, and L. Van Gool (2008b). "Coupled detection and tracking from static cameras and moving vehicles". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 30(10): 1683–1698.

[405] Lenac, K., J. Cesic, I. Markovic, and I. Petrovic (2018). "Exactly sparse delayed state filter on Lie groups for long-term pose graph SLAM". *International Journal of Robotics Research (IJRR).* 37(6): 585–610.

[406] Lenz, P., A. Geiger, and R. Urtasun (2015). "FollowMe: Efficient online min-cost flow tracking with bounded memory and computation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[407] Leutenegger, S., P. T. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart (2013). "Keyframe-based visual-inertial SLAM using nonlinear optimization". In: *Proc. Robotics: Science and Systems (RSS).*

[408] Levi, D., N. Garnett, and E. Fetaya (2015). "StixelNet: A deep convolutional network for obstacle detection and road segmentation". In: *Proc. of the British Machine Vision Conf. (BMVC).*

[409] Levinkov, E., J. Uhrig, S. Tang, M. Omran, E. Insafutdinov, A. Kirillov, C. Rother, T. Brox, B. Schiele, and B. Andres (2017). "Joint graph decomposition & node labeling: Problem, algorithms, Applications". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 1904–1912.

[410] Levinson, J., M. Montemerlo, and S. Thrun (2007). "Map-based precision vehicle localization in urban environments". In: *Proc. Robotics: Science and Systems (RSS).*

[411] Levinson, J. and S. Thrun (2010). "Robust vehicle localization in urban environments using probabilistic maps". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[412] Li, B., T. Zhang, and T. Xia (2016). "Vehicle detection from 3D lidar using fully convolutional network". In: *Proc. Robotics: Science and Systems (RSS)*.

[413] Li, G., M. Mueller, V. Casser, N. Smith, D. L. Michels, and B. Ghanem (2018a). "Teaching UAVs to race with observational imitation learning". arXiv: 1803.01129 [cs.CV].

[414] Li, J., X. Mei, D. V. Prokhorov, and D. Tao (2017a). "Deep neural network for structural prediction and lane detection in traffic scene". *IEEE Trans. on Neural Networks and Learning Systems*. 28(3): 690–703.

[415] Li, Q., A. Arnab, and P. H. S. Torr (2018b). "Weakly- and semi-supervised panoptic segmentation". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 106–124.

[416] Li, X., H. Zhao, L. Han, Y. Tong, and K. Yang (2019). "GFF: Gated fully fusion for semantic segmentation". arXiv: 1904.01803 [cs.CV].

[417] Li, Y., N. Snavely, and D. P. Huttenlocher (2010). "Location recognition using prioritized feature matching". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[418] Li, Y., H. Qi, J. Dai, X. Ji, and Y. Wei (2017b). "Fully convolutional instance-aware semantic segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 4438–4446.

[419] Li, Y. and R. S. Zemel (2014). "Mean-field networks". *Proc. of the International Conf. on Machine Learning (ICML) Workshops*.

[420] Li, Y., D. J. Crandall, and D. P. Huttenlocher (2009). "Landmark classification in large-scale image collections". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[421] Li, Y., N. Snavely, D. Huttenlocher, and P. Fua (2012). "Worldwide pose estimation using 3D point clouds". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[422] Liang, X., T. Wang, L. Yang, and E. Xing (2018). "CIRL: Controllable imitative reinforcement learning for vision-based self-driving". arXiv: 1807.03776 [cs.CV].

[423] Lin, T., S. J. Belongie, and J. Hays (2013). "Cross-view image geolocalization". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 891–898.

[424] Lin, T., Y. Cui, S. J. Belongie, and J. Hays (2015). "Learning deep representations for ground-to-aerial geolocalization". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[425] Lin, T., P. Goyal, R. B. Girshick, K. He, and P. Dollár (2017). "Focal loss for dense object detection". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 2999–3007.

[426] Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick (2014). "Microsoft COCO: Common objects in context". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[427] Liu, L., H. Li, and Y. Dai (2017a). "Efficient Global 2D-3D matching for camera localization in a large-scale 3D map". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 2391–2400.

[428] Liu, S.-C. and T. Delbruck (2010). "Neuromorphic sensory systems". *Current Opinion in Neurobiology*. 20(3): 288–295.

[429] Liu, S., J. Jia, S. Fidler, and R. Urtasun (2017b). "SGN: Sequential grouping networks for instance segmentation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 3516–3524.

[430] Liu, S., L. Qi, H. Qin, J. Shi, and J. Jia (2018). "Path aggregation network for instance segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 8759–8768.

[431] Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C. Fu, and A. C. Berg (2016). "SSD: Single shot MultiBox detector". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[432] Long, J., E. Shelhamer, and T. Darrell (2015). "Fully convolutional networks for semantic segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[433] Longuet-Higgins, H. (1981). "A computer algorithm for reconstructing a scene from two projections". *Nature*. 293: 133–135.

[434] Loper, M., N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black (2015). "SMPL: A skinned multi-person linear model". *ACM Trans. on Graphics*.

[435] Lorensen, W. E. and H. E. Cline (1987). "Marching cubes: A high resolution 3D surface construction algorithm". In: *ACM Trans. on Graphics*. 163–169.

[436] Lowe, D. G. (2004). "Distinctive image features from scale-invariant keypoints". *International Journal of Computer Vision (IJCV)*. 60(2): 91–110.

[437] Lowry, S. M., N. Sünderhauf, P. Newman, J. J. Leonard, D. D. Cox, P. I. Corke, and M. J. Milford (2016). "Visual place recognition: A survey". *IEEE Trans. on Robotics*. 32(1): 1–19.

[438] Lu, C., H. Uchiyama, D. Thomas, A. Shimada, and R. Taniguchi (2018). "Sparse cost volume for efficient stereo matching". *Remote Sensing (RS)*. 10(11): 1844.

[439] Luiten, J., T. Fischer, and B. Leibe (2019). "Track to reconstruct and reconstruct to track". arXiv: 1910.00130 [cs.CV].

[440] Luo, W., A. Schwing, and R. Urtasun (2016). "Efficient deep learning for stereo matching". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[441] Lv, Z., C. Beall, P. Alcantarilla, F. Li, Z. Kira, and F. Dellaert (2016). "A continuous optimization approach for efficient and accurate scene flow". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[442] Lynen, S., T. Sattler, M. Bosse, J. A. Hesch, M. Pollefeys, and R. Siegwart (2015). "Get out of my lab: Large-scale, real-time visual-inertial localization". In: *Proc. Robotics: Science and Systems (RSS)*.

[443] Ma, L., S. Tang, M. J. Black, and L. Van Gool (2018). "Customized multi-person tracker". In: *Proc. of the Asian Conf. on Computer Vision (ACCV)*.

[444] Ma, W., S. Wang, R. Hu, Y. Xiong, and R. Urtasun (2019). "Deep rigid instance scene flow". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[445] Mac Aodha, O., A. Humayun, M. Pollefeys, and G. J. Brostow (2013). "Learning a confidence measure for optical flow". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 35(5): 1107–1120.

[446] Maddern, W., G. Pascoe, C. Linegar, and P. Newman (2016). "1 year, 1000 km: The Oxford RobotCar dataset". *International Journal of Robotics Research (IJRR).* 36(1): 3–15.

[447] Maldonado-Bascón, S., S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gómez-Moreno, and F. López-Ferreras (2007). "Road-sign detection and recognition based on support vector machines". *IEEE Trans. on Intelligent Transportation Systems (T-ITS).* 8(2): 264–278.

[448] Manhardt, F., W. Kehl, and A. Gaidon (2019). "ROI-10D: Monocular lifting of 2D detection to 6D pose and metric shape". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[449] Mansinghka, V., T. Kulkarni, Y. Perov, and J. Tenenbaum (2013). "Approximate Bayesian image interpretation using generative probabilistic graphics programs". In: *Advances in Neural Information Processing Systems (NeurIPS).*

[450] Maqueda, A. I., A. Loquercio, G. Gallego, N. N. García, and D. Scaramuzza (2018). "Event-based vision meets deep learning on steering prediction for self-driving cars". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[451] Martinović, A., J. Knopp, H. Riemenschneider, and L. Van Gool (2015). "3D all the way: Semantic segmentation of urban scenes from start to end in 3D". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[452] Mathias, M., A. Martinovic, and L. Van Gool (2016). "ATLAS: A three-layered approach to facade parsing". *International Journal of Computer Vision (IJCV).* 118(1): 22–48.

[453] Mathias, M., R. Timofte, R. Benenson, and L. Van Gool (2013). "Traffic sign recognition – How far are we from the solution?" In: *International Joint Conference on Neural Networks (IJCNN)*. IEEE. 1–8.

[454] Matthies, L. H., R. Szeliski, and T. Kanade (1988). "Incremental estimation of dense depth maps from image sequences". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 366–374.

[455] Mattyus, G., S. Wang, S. Fidler, and R. Urtasun (2015). "Enhancing road maps by parsing aerial images around the world". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[456] Mattyus, G., S. Wang, S. Fidler, and R. Urtasun (2016). "HD maps: Fine-grained road segmentation by parsing ground and aerial images". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[457] Maturana, D. and S. Scherer (2015). "VoxNet: A 3D convolutional neural network for real-time object recognition". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[458] Mayer, N., E. Ilg, P. Haeusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox (2016). "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[459] McCormac, J., A. Handa, A. J. Davison, and S. Leutenegger (2017). "SemanticFusion: Dense 3D semantic mapping with convolutional neural networks". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 4628–4635.

[460] Mehta, A., A. Subramanian, and A. Subramanian (2018). "Learning end-to-end autonomous driving using guided auxiliary supervision". arXiv: 1808.10393 [cs.LG].

[461] Mei, C. and P. Rives (2007). "Single view point omnidirectional camera calibration from planar grids". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[462] Meister, S., J. Hur, and S. Roth (2018). "UnFlow: Unsupervised learning of optical flow with a bidirectional census loss". In: *Proc. of the Conf. on Artificial Intelligence (AAAI)*.

[463] Menze, M. and A. Geiger (2015). "Object scene flow for autonomous vehicles". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[464] Menze, M., C. Heipke, and A. Geiger (2015a). "Discrete optimization for optical flow". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[465] Menze, M., C. Heipke, and A. Geiger (2015b). "Joint 3D estimation of vehicles and scene flow". In: *Proc. of the ISPRS Workshop on Image Sequence Analysis (ISA)*.

[466] Menze, M., C. Heipke, and A. Geiger (2018). "Object scene flow". *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*. 140: 60–76.

[467] Merrell, P., A. Akbarzadeh, L. Wang, P. Mordohai, J. Frahm, R. Yang, D. Nistér, and M. Pollefeys (2007). "Real-time visibility-based fusion of depth maps". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 1–8.

[468] Metz, C. (2018). "A Toaster on wheels to deliver groceries? Self-driving tech tests practical uses". https://www.nytimes.com/2018/12/18/technology/driverless-mini-car-deliver-groceries.html. Online: accessed 18-October-2019.

[469] Micusik, B. and J. Kosecka (2009). "Piecewise planar city 3D modeling from street view panoramic sequences". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[470] Milan, A., S. Roth, and K. Schindler (2014). "Continuous energy minimization for multitarget tracking". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 36(1): 58–72.

[471] Milan, A., L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler (2016). "MOT16: A benchmark for multi-object tracking". arXiv: 1603.00831 [cs.CV].

[472] Milan, A., S. H. Rezatofighi, A. R. Dick, I. D. Reid, and K. Schindler (2017). "Online multi-target tracking using recurrent neural networks". In: *Proc. of the Conf. on Artificial Intelligence (AAAI)*. 4225–4232.

[473] Milan, A., K. Schindler, and S. Roth (2013). "Detection- and trajectory-level exclusion in multiple object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[474] Mirabdollah, M. H. and B. Mertsching (2014). "On the second order statistics of essential matrix elements". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[475] Mirabdollah, M. H. and B. Mertsching (2015). "Fast techniques for monocular visual odometry". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[476] Mirowski, P., A. Banki-Horvath, K. Anderson, D. Teplyashin, K. M. Hermann, M. Malinowski, M. K. Grimes, K. Simonyan, K. Kavukcuoglu, A. Zisserman, and R. Hadsell (2019). "The streetlearn environment and dataset". arXiv: 1903.01292 [cs.AI].

[477] Mitzel, D. and B. Leibe (2012). "Taking mobile multi-object tracking to the next level: People, unknown objects, and carried items". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[478] Mohan, R. (2014). "Deep deconvolutional networks for scene parsing". *arXiv.org*. 1411.4101.

[479] Montemerlo, M., S. Thrun, D. Koller, and B. Wegbreit (2002). "FastSLAM: A factored solution to the simultaneous localization and mapping problem". In: *Artificial Intelligence (AI)*. 593–598.

[480] Montoya, J., J. D. Wegner, L. Ladicky, and K. Schindler (2015). "Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques". In: *ISPRS Conf. Photogrammetric Image Analysis (PIA)*.

[481] Moosmann, F. and C. Stiller (2013). "Joint self-localization and tracking of generic objects in 3D range data". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[482] Moulon, P., P. Monasse, R. Marlet, *et al.* (2012). "OpenMVG. An open multiple view geometry library". https://github.com/openMVG/openMVG. Online: accessed 23-April-2019.

[483] Mousavian, A., D. Anguelov, J. Flynn, and J. Košecká (2017). "3D bounding box estimation using deep learning and geometry". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 5632–5640.

[484] Mueggler, E., G. Gallego, and D. Scaramuzza (2015). "Continuous-time trajectory estimation for event-based vision sensors". In: *Proc. Robotics: Science and Systems (RSS)*.

[485] Mueggler, E., H. Rebecq, G. Gallego, T. Delbrück, and D. Scaramuzza (2017). "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM". In: *International Journal of Robotics Research (IJRR)*.

[486] Müller, M., A. Dosovitskiy, B. Ghanem, and V. Koltun (2018). "Driving policy transfer via modularity and abstraction". *arXiv.org*. abs/1804.09364.

[487] Munoz, D., J. A. Bagnell, and M. Hebert (2010). "Stacked hierarchical labeling". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[488] Munoz, D., J. A. Bagnell, N. Vandapel, and M. Hebert (2009). "Contextual classification with functional max-margin Markov networks". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[489] Mur-Artal, R., J. M. M. Montiel, and J. D. Tardós (2015). "ORB-SLAM: A versatile and accurate monocular SLAM system". *IEEE Trans. on Robotics*. 31(5): 1147–1163.

[490] Musialski, P., P. Wonka, D. G. Aliaga, M. Wimmer, L. J. Van Gool, and W. Purgathofer (2013). "A survey of urban reconstruction". *Computer Graphics Forum*. 32(6): 146–177.

[491] Myeong, H., J. Y. Chang, and K. M. Lee (2012). "Learning object relationships via graph-based context model". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[492] Neoral, M. and J. Šochman (2017). "Object scene flow with temporal consistency". In: *Proc. of the Computer Vision Winter Workshop (CVWW)*.

[493] Neoral, M., J. Sochman, and J. Matas (2018). "Continual occlusions and optical flow estimation". In: *Proc. of the Asian Conf. on Computer Vision (ACCV)*.

[494] Neuhaus, F., T. Koß, R. Kohnen, and D. Paulus (2018). "MC2-SLAM: Real-time inertial lidar odometry using two-scan motion compensation". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*. 60–72.

[495] Neuhold, G., T. Ollmann, S. Rota Bulò, and P. Kontschieder (2017). "The mapillary vistas dataset for semantic understanding of street scenes". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[496] Newcombe, R. A., S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon (2011a). "KinectFusion: Real-time dense surface mapping and tracking". In: *Proc. of the International Symposium on Mixed and Augmented Reality (ISMAR)*.

[497] Newcombe, R. A., S. Lovegrove, and A. J. Davison (2011b). "DTAM: Dense tracking and mapping in real-time". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[498] Nießner, M., M. Zollhöfer, S. Izadi, and M. Stamminger (2013). "Real-time 3D reconstruction at scale using voxel hashing". In: *ACM Trans. on Graphics*.

[499] Nistér, D. (2004). "An efficient solution to the five-point relative pose problem". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 26(6): 756–777.

[500] Oh, S. M., S. Tariq, B. N. Walker, and F. Dellaert (2004). "Map-based priors for localization". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[501] Ohta, Y. (1985). *Knowledge-Based Interpretation of Outdoor Natural Color Scenes*. Vol. 4. Morgan Kaufmann.

[502] Oliveira, G., W. Burgard, and T. Brox (2016). "Efficient deep methods for monocular road segmentation". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[503] Oliveira, V. M. de, V. Santos, A. D. Sappa, P. Dias, and A. P. Moreira (2016). "Incremental scenario representations for autonomous driving using geometric polygonal primitives". *Robotics and Autonomous Systems (RAS)*. 83: 312–325.

[504] Ortelt, B., C. Herrmann, D. Willersinn, and J. Beyerer (2018). "Foveal vision for instance segmentation of road images". In: *Proc. of the International Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*. 371–378.

[505] Osep, A., W. Mehner, M. Mathias, and B. Leibe (2017). "Combined image- and world-space tracking in traffic scenes". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[506] Pan, X., Y. You, Z. Wang, and C. Lu (2017). "Virtual to real reinforcement learning for autonomous driving". In: *Proc. of the British Machine Vision Conf. (BMVC)*.

[507] Pandey, G., J. R. McBride, and R. M. Eustice (2011). "Ford campus vision and lidar data set". *International Journal of Robotics Research (IJRR)*. 30: 1543–1552.

[508] Pang, J., W. Sun, J. S. J. Ren, C. Yang, and Q. Yan (2017). "Cascade residual learning: A two-stage convolutional neural network for stereo matching". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 878–886.

[509] Papageorgiou, C. and T. A. Poggio (2000). "A trainable system for object detection". *International Journal of Computer Vision (IJCV)*. 38(1): 15–33.

[510] Paschalidou, D., A. O. Ulusoy, C. Schmitt, L. van Gool, and A. Geiger (2018). "RayNet: Learning volumetric 3D reconstruction with ray potentials". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[511] Paskin, M. A. (2003). "Thin junction tree filters for simultaneous localization and mapping". In: *Proc. of the International Joint Conf. on Artificial Intelligence (IJCAI)*. 1157–1166.

[512] Paul, R. and P. Newman (2010). "FAB-MAP 3D: Topological mapping with spatial and visual appearance". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[513] Pellegrini, S., A. Ess, K. Schindler, and L. J. Van Gool (2009). "You'll never walk alone: Modeling social behavior for multi-target tracking". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[514] Pepik, B., M. Stark, P. V. Gehler, and B. Schiele (2015). "Multi-view and 3D deformable part models". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 37(11): 2232–2245.

[515] Pereira, F., J. Luft, G. Ilha, A. Sofiatti, and A. Susin (2017). "Backward motion for estimation enhancement in sparse visual odometry". In: *Workshop of Computer Vision (WVC)*. 61–66.

[516] Perera, A. G. A., C. Srinivas, A. Hoogs, G. Brooksby, and W. Hu (2006). "Multi-object tracking through simultaneous long occlusions and split-merge conditions". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[517] Persson, M., T. Piccini, M. Felsberg, and R. Mester (2015). "Robust stereo visual odometry from monocular techniques". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[518] Pfeiffer, D. and U. Franke (2010). "Efficient representation of traffic scenes by means of dynamic stixels". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[519] Pfeiffer, D. and U. Franke (2011). "Towards a global optimal multi-layer stixel representation of dense 3D data". In: *Proc. of the British Machine Vision Conf. (BMVC)*.

[520] Pham, C. C. and J. W. Jeon (2017). "Robust object proposals re-ranking for object detection in autonomous driving using convolutional neural networks". *Signal Processing: Image Communication (SPIC)*. 53: 110–122.

[521] Pillai, S., R. Ambrus, and A. Gaidon (2019). "SuperDepth: Self-supervised, super-resolved monocular depth estimation". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 9250–9256.

[522] Pinggera, P., U. Franke, and R. Mester (2015). "High-performance long range obstacle detection using stereo vision". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[523] Pinggera, P., S. Ramos, S. Gehrig, U. Franke, C. Rother, and R. Mester (2016). "Lost and found: Detecting small road hazards for self-driving vehicles". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[524] Pinheiro, P. H. O., R. Collobert, and P. Dollár (2015). "Learning to segment object candidates". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 1990–1998.

[525] Pinheiro, P. O., T. Lin, R. Collobert, and P. Dollár (2016). "Learning to refine object segments". In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 75–91.

[526] Pirsiavash, H., D. Ramanan, and C. C. Fowlkes (2011). "Globally-optimal greedy algorithms for tracking a variable number of objects". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[527] Pishchulin, L., E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. V. Gehler, and B. Schiele (2016). "DeepCut: Joint subset partition and labeling for multi person pose estimation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[528] Pishchulin, L., A. Jain, M. Andriluka, T. Thormaehlen, and B. Schiele (2012). "Articulated people detection and pose estimation: Reshaping the future". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[529] Plotkin, L. (2015). "PyDriver: Entwicklung eines Frameworks für räumliche Detektion und Klassifikation von Objekten in Fahrzeugumgebung". *MA thesis*. Karlsruhe Institute of Technology.

[530] Pohlen, T., A. Hermans, M. Mathias, and B. Leibe (2017). "Full-resolution residual networks for semantic segmentation in street scenes". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3309–3318.

[531] Pollard, T. and J. L. Mundy (2007). "Change detection in a 3-d world". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[532] Pollefeys, M. (2008). "Detailed real-time urban 3D reconstruction from video". *International Journal of Computer Vision (IJCV)*. 78(2-3): 143–167.

[533] Pomerleau, D. and T. Jochem (1996). "Rapidly adapting machine vision for automated vehicle steering". *IEEE Expert*. 11(2): 19–27.

[534] Pomerleau, D. (1988). "ALVINN: An autonomous land vehicle in a neural network". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 305–313.

[535] Pomerleau, D. and T. Jochem (2015). "Look, Ma, No Hands". https://www.cmu.edu/news/stories/archives/2015/july/look-ma-no-hands.html. Online: accessed 18-June-2019.

[536] Pont-Tuset, J., P. Arbelaez, J. T. Barron, F. Marqués, and J. Malik (2017). "Multiscale combinatorial grouping for image segmentation and object proposal generation". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 39(1): 128–140.

[537] Qi, C. R., W. Liu, C. Wu, H. Su, and L. J. Guibas (2018). "Frustum pointnets for 3D object detection from RGB-D data". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[538] Qin, Z. and C. R. Shelton (2012). "Improving multi-target tracking via social grouping". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[539] Qiu, W., F. Zhong, Y. Zhang, S. Qiao, Z. Xiao, T. S. Kim, Y. Wang, and A. Yuille (2017). "UnrealCV: Virtual worlds for computer vision". *ACM Multimedia Open Source Software Competition*: 1221–1224.

[540] Quam, L. H. (1984). "Hierarchical warp stereo". *DARPA Image Understanding Workshop*: 149–155.

[541] Rabe, C., T. Mueller, A. Wedel, and U. Franke (2010). "Dense, robust, and accurate motion field estimation from stereo image sequences in real-time". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[542] Rabinovich, A., A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie (2007). "Objects in context". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[543] Radwan, N., A. Valada, and W. Burgard (2018). "VLocNet++: Deep multitask learning for semantic visual localization and odometry". *IEEE Robotics and Automation Letters (RA-L)*. 3(4): 4407–4414.

[544] Ranftl, R., K. Bredies, and T. Pock (2014). "Non-local total generalized variation for optical flow estimation". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[545] Ranftl, R., T. Pock, and H. Bischof (2013). "Minimizing TGV-based variational models with non-convex data terms". In: *Proc. of the International Conf. on Scale Space and Variational Methods in Computer Vision (SSVM).*

[546] Ranjan, A. and M. Black (2017). "Optical flow estimation using a spatial pyramid network". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[547] Rebecq, H., T. Horstschaefer, G. Gallego, and D. Scaramuzza (2016). "EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time". In: *IEEE Robotics and Automation Letters (RA-L).*

[548] Redmon, J., S. K. Divvala, R. B. Girshick, and A. Farhadi (2016). "You only look once: Unified, real-time object detection". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 779–788.

[549] Redmon, J. and A. Farhadi (2017). "YOLO9000: Better, faster, stronger". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[550] Reilly, V., H. Idrees, and M. Shah (2010). "Detection and tracking of large number of targets in wide area surveillance". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[551] Ren, J., X. Chen, J. Liu, W. Sun, J. Pang, Q. Yan, Y.-W. Tai, and L. Xu (2017a). "Accurate single stage detector using recurrent rolling convolution". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[552] Ren, S., K. He, R. B. Girshick, and J. Sun (2015). "Faster R-CNN: Towards real-time object detection with region proposal networks". In: *Advances in Neural Information Processing Systems (NeurIPS).*

[553] Ren, S., K. He, R. B. Girshick, and J. Sun (2017b). "Faster R-CNN: Towards real-time object detection with region proposal Networks". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 39(6): 1137–1149.

[554] Ren, S., K. He, R. B. Girshick, X. Zhang, and J. Sun (2017c). "Object detection networks on convolutional feature maps". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 39(7): 1476–1481.

[555] Ren, Z., D. Sun, J. Kautz, and E. B. Sudderth (2017d). "Cascaded scene flow prediction using semantic segmentation". In: *Proc. of the International Conf. on 3D Vision (3DV)*. 225–233.

[556] Restrepo, M. I., A. O. Ulusoy, and J. L. Mundy (2014). "Evaluation of feature-based 3-D registration of probabilistic volumetric scenes". *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*.

[557] Revaud, J., P. Weinzaepfel, Z. Harchaoui, and C. Schmid (2015). "EpicFlow: Edge-preserving interpolation of correspondences for optical flow". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[558] Richardt, C., H. Kim, L. Valgaerts, and C. Theobalt (2016). "Dense wide-baseline scene flow from two handheld video cameras". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[559] Richter, S. R., Z. Hayder, and V. Koltun (2017). "Playing for benchmarks". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[560] Richter, S. R., V. Vineet, S. Roth, and V. Koltun (2016a). "Playing for data: Ground truth from computer games". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[561] Richter, S. R., V. Vineet, S. Roth, and V. Koltun (2016b). "Playing for data: Ground truth from computer games". https://download.visinf.tu-darmstadt.de/data/from_games/. Online: accessed 18-October-2019.

[562] Riegler, G., A. O. Ulusoy, H. Bischof, and A. Geiger (2017a). "OctNetFusion: Learning depth fusion from data". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[563] Riegler, G., A. O. Ulusoy, and A. Geiger (2017b). "OctNet: Learning deep 3D representations at high resolutions". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[564] Riemenschneider, H., A. Bódis-Szomorú, J. Weissenberg, and L. Van Gool (2014). "Learning where to classify in multi-view semantic segmentation". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[565] Ristani, E., F. Solera, R. Zou, R. Cucchiara, and C. Tomasi (2016). "Performance measures and a data set for multi-target, multi-camera tracking". In: *Proc. of the European Conf. on Computer Vision (ECCV) Workshops*.

[566] Roberts, L. G. (1963). "Machine perception of three-dimensional solids". *PhD thesis*. Massachusetts Institute of Technology.

[567] Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 9351. *Lecture Notes in Computer Science*. Springer. 234–241.

[568] Ros, G., L. Sellart, J. Materzynska, D. Vazquez, and A. Lopez (2016a). "SYNTHIA dataset". http://synthia-dataset.net/. Online: accessed 18-October-2019.

[569] Ros, G., L. Sellart, J. Materzynska, D. Vazquez, and A. Lopez (2016b). "The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[570] Ross, S. and D. Bagnell (2010). "Efficient reductions for imitation learning". In: *Conference on Artificial Intelligence and Statistics AISTATS*. Vol. 9. *JMLR Proceedings*. 661–668.

[571] Rother, C., V. Kolmogorov, and A. Blake (2004). "GrabCut: Interactive foreground extraction using iterated graph cuts". In: *ACM Trans. on Graphics*. Vol. 23. No. 3. 309–314.

[572] Rublee, E., V. Rabaud, K. Konolige, and G. Bradski (2011). "ORB: An efficient alternative to SIFT or SURF". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 2564–2571.

[573] Rudin, L. I., S. Osher, and E. Fatemi (1992). "Nonlinear total variation based noise removal algorithms". *Physica D: Nonlinear Phenomena*. 60(1–4): 259–268.

[574] Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). "Learning representations by back-propagating errors". *Nature*. 323: 533–536.

[575] Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei (2015). "ImageNet large scale visual recognition challenge". *International Journal of Computer Vision (IJCV)*. 115(3): 211–252.

[576] Sadeghian, A., A. Alahi, and S. Savarese (2017). "Tracking the untrackable: Learning to track multiple cues with long-term dependencies". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[577] Saha, S., G. Varma, and C. V. Jawahar (2018). "Improved visual relocalization by discovering anchor points". In: *Proc. of the British Machine Vision Conf. (BMVC)*. 164.

[578] Santana, E. and G. Hotz (2016). "Learning a driving simulator". arXiv: 1608.01230 [cs.LG].

[579] Sarlin, P., C. Cadena, R. Siegwart, and M. Dymczyk (2018). "From coarse to fine: Robust hierarchical localization at large scale". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[580] Sattler, T., B. Leibe, and L. Kobbelt (2016). "Efficient effective prioritized matching for large-scale image-based localization". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 39(9): 1744–1756.

[581] Sattler, T., M. Havlena, F. Radenovic, K. Schindler, and M. Pollefeys (2015). "Hyperpoints and fine vocabularies for large-scale location recognition". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[582] Sattler, T., W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, F. Kahl, and T. Pajdla (2018). "Benchmarking 6DOF outdoor visual localization in changing conditions". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 8601–8610.

[583] Sattler, T., Q. Zhou, M. Pollefeys, and L. Leal-Taixé (2019). "Understanding the limitations of CNN-based absolute camera pose regression". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[584] Sauer, A., N. Savinov, and A. Geiger (2018). "Conditional affordance learning for driving in urban environments". In: *Proc. Conf. on Robot Learning (CoRL)*.

[585] Savva, M., A. X. Chang, and P. Hanrahan (2015). "Semantically-enriched 3D models for common-sense knowledge". *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[586] Saxena, R., R. Schuster, O. Wasenmüller, and D. Stricker (2019). "PWOC-3D: Deep occlusion-aware end-to-end scene flow estimation". *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[587] Scaramuzza, D. and R. Siegwart (2008). "Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles". *IEEE Trans. on Robotics*. 24(5): 1015–1026.

[588] Scaramuzza, D. and F. Fraundorfer (2011). "Visual odometry [tutorial]". *Robotics and Automation Magazine (RAM)*. 18(4): 80–92.

[589] Scaramuzza, D., F. Fraundorfer, and R. Siegwart (2009). "Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[590] Scaramuzza, D. and A. Martinelli (2006). "A toolbox for easily calibrating omnidirectional cameras". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[591] Scharstein, D., H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling (2014). "High-resolution stereo datasets with subpixel-accurate ground truth". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[592] Scharstein, D. and R. Szeliski (2002). "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms". *International Journal of Computer Vision (IJCV)*. 47: 7–42.

[593] Scharstein, D. and R. Szeliski (2003). "High-accuracy stereo depth maps using structured light". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[594] Schauwecker, K., R. Klette, and A. Zell (2012). "A new feature detector and stereo matching method for accurate high-performance sparse stereo matching". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*. 5171–5176.

[595] Scheidegger, S., J. Benjaminsson, E. Rosenberg, A. Krishnan, and K. Granström (2018). "Mono-camera 3D multi-object tracking using deep learning detections and PMBM filtering". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. 433–440.

[596] Schneider, L., M. Cordts, T. Rehfeld, D. Pfeiffer, M. Enzweiler, U. Franke, M. Pollefeys, and S. Roth (2016). "Semantic stixels: Depth is not enough". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[597] Schönbein, M. and A. Geiger (2014). "Omnidirectional 3D reconstruction in augmented manhattan worlds". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[598] Schönbein, M., T. Strauss, and A. Geiger (2014). "Calibrating and centering quasi-central catadioptric cameras". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[599] Schönberger, J. L. and J.-M. Frahm (2016). "Structure-from-motion revisited". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[600] Schönberger, J. L., E. Zheng, M. Pollefeys, and J.-M. Frahm (2016). "Pixelwise view selection for unstructured multi-view stereo". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[601] Schönberger, J., M. Pollefeys, A. Geiger, and T. Sattler (2018). "Semantic visual localization". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[602] Schöps, T., J. Schönberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, and A. Geiger (2017). "A multi-view stereo benchmark with high-resolution images and multi-camera videos". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[603] Schreiber, M., C. Knöppel, and U. Franke (2013). "LaneLoc: Lane marking based localization using highly accurate maps". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.
[604] Schulter, S., P. Vernaza, W. Choi, and M. Chandraker (2017). "Deep network flow for multi-object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
[605] Scott, S. (2019). "Meet Scout". https://blog.aboutamazon.com/transportation/meet-scout. Online: accessed 18-October-2019.
[606] Seff, A. and J. Xiao (2016). "Learning from maps: Visual common sense for autonomous driving". arXiv: 1611.08583 [cs.CV].
[607] Seitz, S. M., B. Curless, J. Diebel, D. Scharstein, and R. Szeliski (2006). "A comparison and evaluation of multi-view stereo reconstruction algorithms". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
[608] Seki, A. and M. Pollefeys (2016). "Patch based confidence prediction for dense disparity map". In: *Proc. of the British Machine Vision Conf. (BMVC)*.
[609] Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra (2017). "Grad-CAM: Visual explanations from deep networks via gradient-based localization". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 618–626.
[610] Sengupta, S., E. Greveson, A. Shahrokni, and P. H. Torr (2013). "Urban 3D semantic modelling using stereo vision". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.
[611] Sengupta, S., P. Sturgess, L. Ladicky, and P. H. S. Torr (2012). "Automatic dense visual semantic mapping from street-level imagery". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.
[612] Sermanet, P., D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun (2014). "OverFeat: Integrated recognition, localization and detection using convolutional networks". In: *Proc. of the International Conf. on Learning Representations (ICLR)*.
[613] Sermanet, P., K. Kavukcuoglu, S. Chintala, and Y. LeCun (2013). "Pedestrian detection with unsupervised multi-stage feature learning". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[614] Sermanet, P. and Y. LeCun (2011). "Traffic sign recognition with multi-scale convolutional networks". In: *International Joint Conference on Neural Networks (IJCNN)*. 2809–2813.

[615] Sevilla-Lara, L., D. Sun, V. Jampani, and M. J. Black (2016). "Optical flow with semantic segmentation and localized layers". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[616] Shafique, K., M. W. Lee, and N. Haering (2008). "A rank constrained continuous formulation of multi-frame multi-target tracking problem". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[617] Shan, Q., C. Wu, B. Curless, Y. Furukawa, C. Hernández, and S. M. Seitz (2014). "Accurate geo-registration by ground-to-aerial image matching". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[618] Sharifzadeh, S., I. Chiotellis, R. Triebel, and D. Cremers (2016). "Learning to drive using inverse reinforcement learning and deep Q-networks". In: *Advances in Neural Information Processing Systems (NeurIPS) Workshops*.

[619] Sharma, S., J. A. Ansari, J. K. Murthy, and K. M. Krishna (2018). "Beyond pixels: Leveraging geometry and shape cues for online multi-object tracking". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[620] Shashua, A., Y. Gdalyahu, and G. Hayun (2004). "Pedestrian detection for driving assistance systems: Single-frame classification and system level performance". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*.

[621] Shen, H., L. Huang, C. Huang, and W. Xu (2018). "Tracklet association tracker: An end-to-end learning-based association approach for multi-object tracking". arXiv: 1808.01562 [cs.CV].

[622] Sheng, H., Y. Zhang, J. Chen, Z. Xiong, and J. Zhang (2019). "Heterogeneous association graph fusion for target association in multiple object tracking". *IEEE Transactions on Circuits and Systems for Video Technology*. 29(11): 3269–3280.

[623] Shotton, J., J. Winn, C. Rother, and A. Criminisi (2009). "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context". *International Journal of Computer Vision (IJCV)*. 81: 2–23.

[624] Shotton, J., B. Glocker, C. Zach, S. Izadi, A. Criminisi, and A. Fitzgibbon (2013). "Scene coordinate regression forests for camera relocalization in RGB-D images". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[625] Shu, G. (2012). "Part-based multiple-person tracking with partial occlusion handling". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[626] Simoncelli, E. P., E. H. Adelson, and D. J. Heeger (1991). "Probability distributions of optical flow". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[627] Simonyan, K. and A. Zisserman (2015). "Very deep convolutional networks for large-scale image recognition". In: *Proc. of the International Conf. on Learning Representations (ICLR)*.

[628] Smith, M., I. Baldwin, W. Churchill, R. Paul, and P. Newman (2009). "The new college vision and laser data set". *International Journal of Robotics Research (IJRR)*. 28: 595–599.

[629] Smith, R., M. Self, and P. Cheeseman (1987). "Estimating uncertain spatial relationships in robotics". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[630] Snavely, N., S. M. Seitz, and R. Szeliski (2006). "Photo tourism: Exploring photo collections in 3D". In: *ACM Trans. on Graphics*. 835–846.

[631] Snavely, N., S. M. Seitz, and R. Szeliski (2008). "Modeling the world from internet photo collections". *International Journal of Computer Vision (IJCV)*. 80(2): 189–210.

[632] Song, S. and M. Chandraker (2014). "Robust scale estimation in real-time monocular SFM for autonomous driving". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[633] Song, X., X. Zhao, H. Hu, and L. Fang (2018). "EdgeStereo: A context integrated residual pyramid network for stereo matching". In: *Proc. of the Asian Conf. on Computer Vision (ACCV)*.

[634] Stallkamp, J., M. Schlipsing, J. Salmen, and C. Igel (2011). "The German traffic sign recognition benchmark: A multi-class classification competition". In: *International Joint Conference on Neural Networks (IJCNN)*. 1453–1460.

[635] Stiefelhagen, R., K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, and P. Soundararajan (2007). "The CLEAR 2006 evaluation". In: *CLEAR*. Vol. 4122. *Lecture Notes in Computer Science*. Springer. 1–44.

[636] Strasdat, H., A. J. Davison, J. M. M. Montiel, and K. Konolige (2011). "Double window optimisation for constant time visual SLAM". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 2352–2359.

[637] Strasdat, H., J. M. M. Montiel, and A. J. Davison (2010). "Scale drift-aware large scale monocular SLAM". In: *Proc. Robotics: Science and Systems (RSS)*.

[638] Strecha, C., W. von Hansen, L. J. Van Gool, P. Fua, and U. Thoennessen (2008). "On benchmarking camera calibration and multi-view stereo for high resolution imagery". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[639] Sturm, J., N. Engelhard, F. Endres, W. Burgard, and D. Cremers (2012). "A benchmark for the evaluation of RGB-D SLAM systems". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*. 573–580.

[640] Suard, F., A. Rakotomamonjy, A. Bensrhair, and A. Broggi (2006). "Pedestrian detection using infrared images and histograms of oriented gradients". In: *Proc. IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 206–212.

[641] Suleymanov, T., L. M. Paz, P. Piniés, G. Hester, and P. Newman (2016). "The path less taken: A fast variational approach for scene segmentation used for closed loop control". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[642] Sun, D., S. Roth, and M. J. Black (2014). "A quantitative analysis of current practices in optical flow estimation and the principles behind them". *International Journal of Computer Vision (IJCV)*. 106(2): 115–137.

[643] Sun, D., X. Yang, M. Liu, and J. Kautz (2018a). "Models matter, so does training: An empirical study of CNNs for optical flow estimation". arXiv: 1809.05571 [cs.CV].

[644] Sun, D., X. Yang, M.-Y. Liu, and J. Kautz (2018b). "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[645] Sun, M. and S. Savarese (2011). "Articulated part-based model for joint object detection and pose estimation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[646] Sun, P., H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov (2019). "Scalability in perception for autonomous driving: Waymo open dataset". arXiv: 1912.04838 [cs.CV].

[647] Svarm, L., O. Enqvist, F. Kahl, and M. Oskarsson (2017). "City-scale localization for cameras with known vertical direction". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 39(7): 1455–1461.

[648] Svärm, L., O. Enqvist, M. Oskarsson, and F. Kahl (2014). "Accurate localization and pose estimation for large 3D models". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 532–539.

[649] Svoboda, T. and T. Pajdla (2002). "Epipolar geometry for central catadioptric cameras". *International Journal of Computer Vision (IJCV)*. 49(1): 23–37.

[650] Sweeney, C. (2016). "Theia multiview geometry library: Tutorial & reference". http://theia-sfm.org. Online: accessed 23-April-2019.

[651] Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich (2015). "Going deeper with convolutions". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[652] Szeliski, R. (2011). *Computer Vision – Algorithms and Applications. Texts in Computer Science*. Springer.

[653] Tang, P., C. Wang, X. Wang, W. Liu, W. Zeng, and J. Wang (2019a). "Object detection in videos by high quality object linking". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 42(5): 1272–1278.

[654] Tang, S., B. Andres, M. Andriluka, and B. Schiele (2015). "Subgraph decomposition for multi-target tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[655] Tang, S., B. Andres, M. Andriluka, and B. Schiele (2016). "Multi-person tracking by multicut and deep matching". In: *Proc. of the European Conf. on Computer Vision (ECCV) Workshops*.

[656] Tang, S., M. Andriluka, B. Andres, and B. Schiele (2017). "Multiple people tracking by lifted multicut and person re-identification". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3701–3710.

[657] Tang, T. Y., D. J. Yoon, and T. D. Barfoot (2019b). "A white-noise-on-jerk motion prior for continuous-time trajectory estimation on SE(3)". *IEEE Robotics and Automation Letters (RA-L)*. 4(2): 594–601.

[658] Taniai, T., S. N. Sinha, and Y. Sato (2017). "Fast multi-frame stereo scene flow with motion segmentation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 6891–6900.

[659] Tesla (2014). "Tesla autopilot". https://www.tesla.com/autopilot. Online: accessed 18-October-2019.

[660] Tesla (2018). "Introducing software version 9.0". https://www.tesla.com/blog/introducing-software-version-9?redirect=no. Online: accessed 8-June-2019.

[661] Thorpe, C., M. H. Hebert, T. Kanade, and S. A. Shafer (1988). "Vision and navigation for the Carnegie-Mellon Navlab". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 10(3): 362–372.

[662] Thrun, S., W. Burgard, and D. Fox (2005). *Probabilistic Robotics*. The MIT Press.

[663] Tian, W., M. Lauer, and L. Chen (2019). "Online multi-object tracking using joint domain information in traffic scenarios". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 21(1): 374–384.

[664] Tian, Y., A. Dehghan, and M. Shah (2018). "On detection, data association and segmentation for multi-target tracking". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 41(9): 2146–2160.

[665] TIME USA (1925). "Science: Radio auto". http://content.time.com/time/magazine/article/0,9171,720720,00.html. Online: accessed 18-October-2019.

[666] Timofte, R. and L. Van Gool (2015). "Sparse flow: Sparse matching for small to large displacement optical flow". In: *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*.

[667] Tombari, F., S. Salti, and L. di Stefano (2010). "Unique signatures of histograms for local surface description". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[668] Topfer, D., J. Spehr, J. Effertz, and C. Stiller (2015). "Efficient road scene understanding for intelligent vehicles using compositional hierarchical models". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 16(1): 441–451.

[669] Torii, A., R. Arandjelović, J. Sivic, M. Okutomi, and T. Pajdla (2015). "24/7 place recognition by view synthesis". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[670] Tsai, A., A. Yezzi Jr., W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky (2003). "A shape-based approach to the segmentation of medical imagery using level sets". *Medical Imaging*. 22(2): 137–154.

[671] Tu, Z. and X. Bai (2010). "Auto-context and its application to high-level vision tasks and 3D brain image segmentation". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 32(10): 1744–1757.

[672] Tulyakov, S., A. Ivanov, and F. Fleuret (2018). "Practical deep stereo (PDS): Toward applications-friendly deep stereo matching". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 5875–5885.

[673] Uber (2015). "Advanced technologies group". https://www.uber.com/de/de/atg/. Online: accessed 18-October-2019.

[674] Uhrig, J., M. Cordts, U. Franke, and T. Brox (2016). "Pixel-level encoding and depth layering for instance-level semantic labeling". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*. 14–25.

[675] Uijlings, J. R., K. E. van de Sande, T. Gevers, and A. W. Smeulders (2013). "Selective search for object recognition". *International Journal of Computer Vision (IJCV)*. 104(2): 154–171.

[676] Ulusoy, A. O., M. Black, and A. Geiger (2017). "Semantic multi-view stereo: Jointly estimating objects and voxels". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[677] Ulusoy, A. O., A. Geiger, and M. J. Black (2015). "Towards probabilistic volumetric reconstruction using ray potentials". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[678] Uras, S., F. Girosi, A. Verri, and V. Torre (1988). "A computational approach to motion perception". *Biological Cybernetics*. 60(2): 79–87.

[679] Usenko, V. C., J. Engel, J. Stückler, and D. Cremers (2016). "Direct visual-inertial odometry with stereo cameras". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 1885–1892.

[680] Valada, A., R. Mohan, and W. Burgard (2018). "Self-supervised model adaptation for multimodal semantic segmentation". arXiv: 1808.03833 [cs.CV].

[681] Valada, A., J. Vertens, A. Dhall, and W. Burgard (2017). "AdapNet: Adaptive semantic segmentation in adverse environmental conditions". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 4644–4651.

[682] Valentin, J. P., S. Sengupta, J. Warrell, A. Shahrokni, and P. H. Torr (2013). "Mesh based semantic modelling for indoor and outdoor scenes". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[683] Vedula, S., S. Baker, P. Rander, R. Collins, and T. Kanade (1999). "Three-dimensional scene flow". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[684] Verbeek, J. J. and B. Triggs (2007). "Scene segmentation with CRFs learned from partially labeled images". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 1553–1560.

[685] Verdie, Y. and F. Lafarge (2014). "Detecting parametric objects in large scenes by Monte Carlo sampling". *International Journal of Computer Vision (IJCV)*. 106(1): 57–75.

[686] Vidal, A. R., H. Rebecq, T. Horstschaefer, and D. Scaramuzza (2018). "Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios". *IEEE Robotics and Automation Letters (RA-L)*.

[687] Vijayanarasimhan, S. and K. Grauman (2012). "Active frame selection for label propagation in videos". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[688] Vineet, V., O. Miksik, M. Lidegaard, M. Niessner, S. Golodetz, V. A. Prisacariu, O. Kahler, D. W. Murray, S. Izadi, P. Perez, and P. H. S. Torr (2015). "Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[689] Viola, P. A., M. J. Jones, and D. Snow (2005). "Detecting pedestrians using patterns of motion and appearance". *International Journal of Computer Vision (IJCV)*. 63(2): 153–161.

[690] Viola, P. A. and M. J. Jones (2004). "Robust real-time face detection". *International Journal of Computer Vision (IJCV)*. 57(2): 137–154.

[691] Vogel, C., S. Roth, and K. Schindler (2013). "An evaluation of data costs for optical flow". In: *Proc. of the German Conference on Pattern Recognition (GCPR)*.

[692] Vogel, C., K. Schindler, and S. Roth (2015). "3D scene flow estimation with a piecewise rigid scene model". *International Journal of Computer Vision (IJCV)*. 115(1): 1–28.

[693] Vosselman, G., S. Dijkman, *et al.* (2001). "3D building model reconstruction from point clouds and ground plans". *Proc. of the ISPRS Workshop Land Surface Mapping and Characterization Using Laser Altimetry*. 34(3/W4): 37–44.

[694] Walch, F., C. Hazirbas, L. Leal-Taixé, T. Sattler, S. Hilsenbeck, and D. Cremers (2017). "Image-based localization using LSTMs for structured feature correlation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 627–637.

[695] Walk, S., N. Majer, K. Schindler, and B. Schiele (2010). "New features and insights for pedestrian detection". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[696] Wang, D., C. Devin, Q. Cai, P. Krähenbühl, and T. Darrell (2019a). "Monocular plan view networks for autonomous driving". arXiv: 1905.06937 [cs.CV].

[697] Wang, D. Z. and I. Posner (2015). "Voting for voting in online point cloud object detection". In: *Proc. Robotics: Science and Systems (RSS)*.

[698] Wang, G., Y. Wang, H. Zhang, R. Gu, and J. Hwang (2018a). "Exploit the connectivity: Multi-object tracking with TrackletNet". arXiv: 1811.07258 [cs.CV].

[699] Wang, R., M. Schwörer, and D. Cremers (2017). "Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 3923–3931.

[700] Wang, S., S. Suo, W.-C. Ma, A. Pokrovsky, and R. Urtasun (2018b). "Deep parametric continuous convolutional neural networks". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[701] Wang, S., Y. Zhou, J. Yan, and Z. Deng (2018c). "Fully motion-aware network for video object detection". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[702] Wang, X., M. Yang, S. Zhu, and Y. Lin (2015). "Regionlets for generic object detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*. 37(10): 2071–2084.

[703] Wang, Y., Y. Yang, Z. Yang, L. Zhao, P. Wang, and W. Xu (2018d). "Occlusion aware unsupervised learning of optical flow". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[704] Wang, Z., B. Liu, S. Schulter, and M. Chandraker (2019b). "A parametric top-view representation of complex road scenes". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[705] Wannenwetsch, A. S., M. Keuper, and S. Roth (2017). "ProbFlow: Joint optical flow and uncertainty estimation". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*. 1182–1191.

[706] Waymo (2019). "Be an early rider". https://waymo.com/apply. Online: accessed 18-October-2019.

[707] Wedel, A., T. Brox, T. Vaudrey, C. Rabe, U. Franke, and D. Cremers (2011). "Stereoscopic scene flow computation for 3D motion understanding". *International Journal of Computer Vision (IJCV)*. 95(1): 29–51.

[708] Wedel, A., C. Rabe, H. Badino, H. Loose, U. Franke, and D. Cremers (2009). "B-spline modeling of road surfaces with an application to free space estimation". *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 10(4): 572–583.

[709] Wedel, A., C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers (2008). "Efficient dense scene flow from sparse or dense stereo data". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[710] Wegner, J. D., S. Branson, D. Hall, K. Schindler, and P. Perona (2016). "Cataloging public objects using aerial and street-level images - Urban trees". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[711] Wegner, J. D., J. A. Montoya-Zegarra, and K. Schindler (2013). "A higher-order CRF model for road network extraction". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[712] Wegner, J. D., J. A. Montoya-Zegarra, and K. Schindler (2015). "Road networks as collections of minimum cost paths". *ISPRS Journal of Photogrammetry and Remote Sensing (JPRS)*. 108 (Complete): 128–137.

[713] Wei, D., C. Liu, and W. Freeman (2014). "A data-driven regularization model for stereo and flow". In: *Proc. of the International Conf. on 3D Vision (3DV)*.

[714] Weinzaepfel, P., J. Revaud, Z. Harchaoui, and C. Schmid (2013). "DeepFlow: Large displacement optical flow with deep matching". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[715] Whelan, T., S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. J. Davison (2015). "ElasticFusion: Dense SLAM without A pose graph". In: *Proc. Robotics: Science and Systems (RSS)*.

[716] Winner, H., S. Hakuli, F. Lotz, C. Singer, A. Geiger, *et al.* (2015). *Handbook of Driver Assistance Systems*. Springer Vieweg.

[717] Wöhler, C. and J. K. Anlauf (1999). "A time delay neural network algorithm for estimating image-pattern shape and motion". *Image and Vision Computing (IVC)*.

[718] Wojek, C., S. Roth, K. Schindler, and B. Schiele (2010). "Monocular 3D scene modeling and inference: Understanding multi-object traffic scenes". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[719] Wojek, C., S. Walk, and B. Schiele (2009). "Multi-cue onboard pedestrian detection". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[720] Wojek, C. and B. Schiele (2008a). "A dynamic conditional random field model for joint labeling of object and scene classes". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[721] Wojek, C. and B. Schiele (2008b). "A performance evaluation of single and multi-feature people detection". In: *Proc. of the DAGM Symposium on Pattern Recognition (DAGM)*.

[722] Wojek, C., S. Walk, S. Roth, and B. Schiele (2011). "Monocular 3D scene understanding with explicit occlusion reasoning". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[723] Wojek, C., S. Walk, S. Roth, K. Schindler, and B. Schiele (2013). "Monocular visual scene understanding: Understanding multi-object traffic scenes". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 35(4): 882–897.

[724] Wojke, N., A. Bewley, and D. Paulus (2017). "Simple online and realtime tracking with a deep association metric". In: *Proc. IEEE International Conf. on Image Processing (ICIP).* 3645–3649.

[725] Wolff, M., R. T. Collins, and Y. Liu (2016). "Regularity-driven facade matching between aerial and street views". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[726] Woodford, O., P. Torr, I. Reid, and A. Fitzgibbon (2009). "Global stereo reconstruction under second-order smoothness priors". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 31: 2115–2128.

[727] Workman, S., R. Souvenir, and N. Jacobs (2015). "Wide-area image geolocalization with aerial reference imagery". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV).*

[728] Wu, B. and R. Nevatia (2007). "Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet part detectors". *International Journal of Computer Vision (IJCV).* 75(2): 247–266.

[729] Wu, C. (2011). "VisualSFM: A visual structure from motion system". Available at URL: http://www.cs.washington.edu/homes/ccwu/vsfm (accessed 23 April 2019).

[730] Wu, T. and A. Ranganathan (2012). "A practical system for road marking detection and recognition". In: *Proc. IEEE Intelligent Vehicles Symposium (IV).* 25–30.

[731] Wu, T., B. Li, and S. Zhu (2016). "Learning and-or model to represent context and occlusion for car detection and viewpoint estimation". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 38(9): 1829–1843.

[732] Wu, Z., T. H. Kunz, and M. Betke (2011). "Efficient track linking methods for track graphs using network-flow and set-cover techniques". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 1185–1192.

[733] Wu, Z., A. Thangali, S. Sclaroff, and M. Betke (2012). "Coupling detection and data association for multiple object tracking". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[734] Wu, Z., C. Shen, and A. van den Hengel (2019). "Wider or deeper: Revisiting the ResNet model for visual recognition". *Pattern Recognition*. 90: 119–133.

[735] Wulff, J. and M. J. Black (2015). "Efficient sparse-to-dense optical flow estimation using a learned basis and layers". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[736] Wymann, B., C. Dimitrakakis, A. Sumner, and C. Guionneau (2015). "TORCS: The open racing car simulator".

[737] Xiang, Y., A. Alahi, and S. Savarese (2015a). "Learning to track: Online multi-object tracking by decision making". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[738] Xiang, Y., W. Choi, Y. Lin, and S. Savarese (2015b). "Data-driven 3D voxel patterns for object category recognition". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[739] Xiang, Y., W. Choi, Y. Lin, and S. Savarese (2017). "Subcategory-aware convolutional neural networks for object proposals and detection". In: *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*. 924–933.

[740] Xiao, J., T. Fang, P. Zhao, M. Lhuillier, and L. Quan (2009). "Image-based street-side city modeling". *ACM Trans. on Graphics*. 28(5): 114:1–114:12.

[741] Xiao, J. and L. Quan (2009). "Multiple view semantic segmentation for street view images". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[742] Xie, J., M. Kiefel, M.-T. Sun, and A. Geiger (2016). "Semantic instance annotation of street scenes by 3D to 2D label transfer". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[743] Xing, J., H. Ai, and S. Lao (2009). "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[744] Xiong, X., D. Munoz, J. A. Bagnell, and M. Hebert (2011). "3-D scene analysis via sequenced predictions over points and regions". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA).* 2609–2616.

[745] Xiong, Y., R. Liao, H. Zhao, R. Hu, M. Bai, E. Yumer, and R. Urtasun (2019). "UPSNet: A unified panoptic segmentation network". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[746] Xu, H., Y. Gao, F. Yu, and T. Darrell (2017). "End-to-end learning of driving models from large-scale video datasets". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 3530–3538.

[747] Xu, Q. and W. Tao (2019). "Multi-scale geometric consistency guided multi-view stereo". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[748] Yamaguchi, K., D. McAllester, and R. Urtasun (2013). "Robust monocular epipolar flow estimation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[749] Yamaguchi, K., T. Hazan, D. McAllester, and R. Urtasun (2012). "Continuous Markov random fields for robust stereo estimation". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[750] Yamaguchi, K., D. McAllester, and R. Urtasun (2014). "Efficient joint segmentation, occlusion labeling, stereo and flow estimation". In: *Proc. of the European Conf. on Computer Vision (ECCV).*

[751] Yan, Y., Y. Mao, and B. Li (2018). "SECOND: Sparsely embedded convolutional detection". *Sensors.* 18(10): 3337.

[752] Yang, B., J. Yan, Z. Lei, and S. Z. Li (2016a). “Craft objects from images”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 6043–6051.

[753] Yang, B., C. Huang, and R. Nevatia (2011). “Learning affinities and dependencies for multi-target tracking using a CRF model”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[754] Yang, B. and R. Nevatia (2012). “An online learned CRF model for multi-target tracking”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[755] Yang, F., W. Choi, and Y. Lin (2016b). “Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[756] Yang, G., H. Zhao, J. Shi, Z. Deng, and J. Jia (2018a). “SegStereo: Exploiting semantic information for disparity estimation”. In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 660–676.

[757] Yang, N., R. Wang, J. Stückler, and D. Cremers (2018b). “Deep virtual stereo odometry: Leveraging deep depth prediction for monocular direct sparse odometry”. In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 835–852.

[758] Yang, T., X. Long, A. K. Sangaiah, Z. Zheng, and C. Tong (2018c). “Deep detection network for real-life traffic sign in vehicular networks”. *Computer Networks*. 136: 95–104.

[759] Yang, Y., S. Baker, A. Kannan, and D. Ramanan (2012). “Recognizing proxemics in personal photos”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 3522–3529.

[760] Yao, Y., Z. Luo, S. Li, T. Fang, and L. Quan (2018). “MVSNet: Depth inference for unstructured multi-view stereo”. *Proc. of the European Conf. on Computer Vision (ECCV)*.

[761] Yao, Y., Z. Luo, S. Li, T. Shen, T. Fang, and L. Quan (2019). “Recurrent MVSNet for high-resolution multi-view stereo depth inference”. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[762] Yin, Z., T. Darrell, and F. Yu (2019). "Hierarchical discrete distribution decomposition for match density estimation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 6044–6053.

[763] Yoon, J. H., C.-R. Lee, M.-H. Yang, and K.-J. Yoon (2016). "Online multi-object tracking via structural constraint event aggregation". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[764] Yoon, J. H., M. Yang, J. Lim, and K. Yoon (2015). "Bayesian multi-object tracking using motion context from multiple objects". In: *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*.

[765] Yu, F., W. Li, Q. Li, Y. Liu, X. Shi, and J. Yan (2016a). "POI: Multiple object tracking with high performance detection and appearance feature". In: *Proc. of the European Conf. on Computer Vision (ECCV) Workshops*.

[766] Yu, F. and V. Koltun (2016). "Multi-scale context aggregation by dilated convolutions". In: *Proc. of the International Conf. on Learning Representations (ICLR)*.

[767] Yu, F., W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell (2018). "BDD100K: A diverse driving video database with scalable annotation tooling". arXiv: 1805.04687 [cs.CV].

[768] Yu, F., W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell (2019). "Berkeley DeepDrive". https://bdd-data.berkeley.edu/. Online: accessed 05-June-2019.

[769] Yu, F., J. Xiao, and T. A. Funkhouser (2015). "Semantic alignment of LiDAR data at city scale". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 1722–1731.

[770] Yu, J. J., A. W. Harley, and K. G. Derpanis (2016b). "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[771] Zach, C., T. Pock, and H. Bischof (2007a). "A duality based approach for realtime TV-L1 optical flow". In: *Proc. of the DAGM Symposium on Pattern Recognition (DAGM)*. 214–223.

[772] Zach, C., T. Pock, and H. Bischof (2007b). “A duality based approach for realtime TV-L1 optical flow”. In: *Pattern Recognition Letters*. Berlin, Heidelberg: Springer. 214–223.

[773] Zach, C., T. Pock, and H. Bischof (2007c). “A globally optimal algorithm for robust TV-L1 range image integration”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[774] Zamir, A. R., A. Dehghan, and M. Shah (2012). “GMCP-tracker: Global multi-object tracking using generalized minimum clique graphs”. In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[775] Žbontar, J. and Y. LeCun (2016). “Stereo matching by training a convolutional neural network to compare image patches”. *Journal of Machine Learning Research (JMLR)*. 17(65): 1–32.

[776] Zeiler, M. D., G. W. Taylor, and R. Fergus (2011). “Adaptive deconvolutional networks for mid and high level feature learning”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[777] Zeisl, B., T. Sattler, and M. Pollefeys (2015). “Camera pose voting for large-scale image-based localization”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[778] Zhang, H., A. Geiger, and R. Urtasun (2013). “Understanding high-level semantics by modeling traffic patterns”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[779] Zhang, J., M. Kaess, and S. Singh (2014a). “Real-time depth enhanced monocular odometry”. In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[780] Zhang, J. and S. Singh (2014). “LOAM: Lidar odometry and mapping in real-time”. In: *Proc. Robotics: Science and Systems (RSS)*.

[781] Zhang, J. and S. Singh (2015). “Visual-lidar odometry and mapping: Low-drift, robust, and fast”. In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*.

[782] Zhang, L., Y. Li, and R. Nevatia (2008). “Global data association for multi-object tracking using network flows”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[783] Zhang, N., J. Donahue, R. B. Girshick, and T. Darrell (2014b). "Part-based R-CNNs for fine-grained category detection". In: *Proc. of the European Conf. on Computer Vision (ECCV)*.

[784] Zhang, Q. and R. Pless (2004). "Extrinsic calibration of a camera and laser range finder". In: *Proc. IEEE International Conf. on Intelligent Robots and Systems (IROS)*.

[785] Zhang, R., S. A. Candra, K. Vetter, and A. Zakhor (2015a). "Sensor fusion for semantic segmentation of urban scenes". In: *Proc. IEEE International Conf. on Robotics and Automation (ICRA)*. 1850–1857.

[786] Zhang, S., R. Benenson, M. Omran, J. H. Hosang, and B. Schiele (2016a). "How far are we from solving pedestrian detection?" In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[787] Zhang, S., X. Zhao, L. Fang, H. Fei, and H. Song (2018). "Led: Localization-quality estimation embedded detector". In: *Proc. IEEE International Conf. on Image Processing (ICIP)*. 584–588.

[788] Zhang, Y. and T. Chen (2012). "Efficient inference for fully-connected CRFs with stationarity". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[789] Zhang, Z., S. Fidler, and R. Urtasun (2016b). "Instance-level segmentation for autonomous driving with deep densely connected MRFs". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[790] Zhang, Z., A. G. Schwing, S. Fidler, and R. Urtasun (2015b). "Monocular object instance segmentation and depth ordering with CNNs". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[791] Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia (2017). "Pyramid scene parsing network". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 6230–6239.

[792] Zheng, S., S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr (2015). "Conditional random fields as recurrent neural networks". In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[793] Zheng, Y., M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T.-S. Chua, and H. Neven (2009). “Tour the world: Building a web-scale landmark recognition engine”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[794] Zhou, B., A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva (2014). “Learning deep features for scene recognition using places database”. In: *Advances in Neural Information Processing Systems (NeurIPS)*.

[795] Zhou, B., P. Krähenbühl, and V. Koltun (2019). “Does computer vision matter for action?” arXiv: 1905.12887 [cs.CV].

[796] Zhou, C., F. Güney, Y. Wang, and A. Geiger (2015). “Exploiting object similarity in 3D reconstruction”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[797] Zhou, Y. and O. Tuzel (2018). “VoxelNet: End-to-end learning for point cloud based 3D object detection”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[798] Zhu, H., K. V. Yuen, L. Mihaylova, and H. Leung (2017a). “Overview of environment perception for intelligent vehicles”. *IEEE Trans. on Intelligent Transportation Systems (T-ITS)*. 18(10): 2584–2601.

[799] Zhu, J., H. Yang, N. Liu, M. Kim, W. Zhang, and M. Yang (2018a). “Online multi-object tracking with dual matching attention networks”. In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 379–396.

[800] Zhu, J. (2017). “Image gradient-based joint direct visual odometry for stereo camera”. In: *Proc. of the International Joint Conf. on Artificial Intelligence (IJCAI)*. 4558–4564.

[801] Zhu, X., J. Dai, L. Yuan, and Y. Wei (2018b). “Towards high performance video object detection”. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[802] Zhu, X., Y. Wang, J. Dai, L. Yuan, and Y. Wei (2017b). “Flow-guided feature aggregation for video object detection”. In: *Proc. of the IEEE International Conf. on Computer Vision (ICCV)*.

[803] Zhu, X., Y. Xiong, J. Dai, L. Yuan, and Y. Wei (2017c). "Deep feature flow for video recognition". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).*

[804] Zhu, Y., J. Wang, C. Zhao, H. Guo, and H. Lu (2016a). "Scale-adaptive deconvolutional regression network for pedestrian detection". In: *Proc. of the Asian Conf. on Computer Vision (ACCV).*

[805] Zhu, Z., D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu (2016b). "Traffic-sign detection and classification in the wild". In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 2110–2118.

[806] Zhuang, Y., L. Tao, F. Yang, C. Ma, Z. Zhang, H. Jia, and X. Xie (2018a). "RelationNet: Learning deep-aligned representation for semantic image segmentation". In: *Proc. of the International Conf. on Pattern Recognition (ICPR).* 1506–1511.

[807] Zhuang, Y., F. Yang, L. Tao, C. Ma, Z. Zhang, Y. Li, H. Jia, X. Xie, and W. Gao (2018b). "Dense relation network: Learning consistent and context-aware representation for semantic image segmentation". In: *Proc. IEEE International Conf. on Image Processing (ICIP).* 3698–3702.

[808] Zia, M., M. Stark, B. Schiele, and K. Schindler (2013). "Detailed 3D representations for object recognition and modeling". *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI).* 35(11): 2608–2623.

[809] Zia, M., M. Stark, and K. Schindler (2015). "Towards scene understanding with detailed 3D object representations". *International Journal of Computer Vision (IJCV).* 112(2): 188–203.

[810] Ziegler, J., P. Bender, M. Schreiber, and H. Lategahnf (2014). "Making bertha drive - An autonomous journey on a historic route". *Proc. IEEE Intelligent Transportation Systems Magazine (ITSM).* 6(2): 8–20.

[811] Zimmer, H., A. Bruhn, and J. Weickert (2011). "Optic flow in harmony". *International Journal of Computer Vision (IJCV).* 93(3): 368–388.