Report of Dagstuhl Seminar 15461

Vision for Autonomous Vehicles and Probes

Edited by

Andrés Bruhn[1], Atsushi Imiya[2], Aleš Leonardis[3], and Tomas Pajdla[4]

1 **Universität Stuttgart, DE,** bruhn@vis.uni-stuttgart.de
2 **Chiba University, JP,** imiya@faculty.chiba-u.jp
3 **University of Birmingham, GB,** a.leonardis@cs.bham.ac.uk
4 **Czech Technical University Prague, CZ,** pajdla@cmp.fell.cvut.cz

Abstract

The vision-based autonomous driving and navigation of vehicles has a long history. In 2013, Daimler succeeded autonomous driving on a public drive way. Today, the Curiosity mars rover is sending video views from Mars to Earth. Computer vision plays a key role in advanced driver assistance systems (ADAS) as well as in exploratory and service robotics. Continuing topics of interest in computer vision are scene and environmental understanding using single- and multiple-camera systems, which are fundamental techniques for autonomous driving, navigation in unknown environments and remote visual exploration. Therefore, we strictly focuses on mathematical, geometrical and computational aspects of autonomous vehicles and autonomous vehicular technology which make use of computer vision and pattern recognition as the central component for autonomous driving and navigation and remote exploration.

Seminar November 8–13, 2015 – http://www.dagstuhl.de/15461
1998 ACM Subject Classification I.2.9 Robotics, I.2.10 Vision and Scene Understanding, I.4.5 Reconstruction, I.4.6 Segmentation, I.4.8 Scene Analysis
Keywords and phrases Vision-based autonomous driving and navigation, Exploratory rovers, Dynamic 3D scene understanding, Simultaneous localization and mapping, On-board algorithms
Digital Object Identifier 10.4230/DagRep.5.11.36
Edited in cooperation with Akihiro Torii (Tokyo Institute of Technology, JP) and Hayato Itoh (Chiba University, JP)

1 Executive Summary

Andrés Bruhn
Atsushi Imiya

Computer vision plays a key role in advanced driver assistance systems (ADAS) as well as in exploratory and service robotics. Visual odometry, trajectory planning for Mars exploratory rovers and the recognition of scientific targets in images are examples of successful applications. In addition, new computer vision theory focuses on supporting autonomous driving and navigation as applications to unmanned aerial vehicles (UAVs) and underwater robots. From the viewpoint of geometrical methods for autonomous driving, navigation and exploration, the on-board calibration of multiple cameras, simultaneous localisation and mapping (SLAM) in non-human-made environments and the processing of non-classical features are some of

Vision for Autonomous Vehicles and Probes, *Dagstuhl Reports*, Vol. 5, Issue 11, pp. 36–61
Editors: André Bruhn, Atsushi Imiya, Aleš Leonardis and Tomas Pajdla
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

current problems. Furthermore, the adaptation of algorithms to long image sequences, image pairs with large displacements and image sequences with changing illumination is desired for robust navigation and exploration. Moreover, the extraction of non-verbal and graphical information from environments to remote driver assistance is required.

Based on these wide range of theoretical interests from computer vision for new possibility of practical applications of computer vision and robotics, 38 participants (excluding organisers) attended from variety of countries: 4 from Australia, 3 from Austria, 3 from Canada, 1 from Denmark, 11 from Germany, 1 from Greece, 1 from France, 3 from Japan, 4 from Spain, 2 from Sweden, 4 from Switzerland and 3 from the US.

The seminar was workshop style. The talks are 40 mins and 30 mins for young researchers and for presenters in special sessions. The talks have been separated into sessions on aerial vehicle vision, under water and space vision, map building, three-dimensional scene and motion understanding as well as a dedicated session on robotics. In these tasks, various types of autonomous systems such as autonomous aerial vehicles, under water robots, field and space probes for remote exploration and autonomous driving cars were presented. Moreover, applications of state-of-the-art computer vision techniques such as global optimization methods, deep learning approaches as well as geometrical methods for scene reconstruction and understanding were discussed. Finally, with Seminar 15462 a joint session on autonomous driving with leading experts in the field was organised.

The working groups are focused on "Sensing," "Interpretation and Map building" and "Deep leaning." Sensing requires fundamental methodologies in computer vision. Low-level sensing is a traditional problem in computer vision. For applications of computer-vision algorithms to autonomous vehicles and probes, reformulation of problems for various conditions are required. Map building is a growing area including applications to autonomous robotics and urban computer vision. Today, application to autonomous map generation involves classical SLAM and large-scale reconstruction from indoor to urban sizes. Furthermore, for SLAM on-board and on–line computation is required. Deep learning, which goes back its origin to '70s, is a fundamental tool for image pattern recognition and classification. Although the method showed significant progress in image pattern recognition and discrimination, for applications to spatial recognition and three-dimensional scene understanding, we need detailed discussion and developments.

Through talks-and-discussion and working-group discussion, the seminar clarified that for designing of platforms for visual interpretation and understanding of three-dimensional world around the system, machine vision provides fundamental and essential methodologies. There is the other methodology which uses computer vision as a sensing system for the acquisition of geometrical data and analysis of motion around cars. For these visual servo systems, computer vision is a part of the platform for intelligent visual servo system. The former methodology is a promising one to provide a fundamental platform which is common to both autonomous vehicles, which are desired for consumer intelligence, and probes, which are used for remote exploration.

2 Table of Contents

3 Overview of Talks

35 talks have been categorised as follows.

Vision for Mapping, Reconstruction and SLAM

Hayko Riemenschneider	Efficient Multi-view Semantic Segmentation
Michal Havlena	Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition
Antonios Gasteratos	Semantic Maps for High Level Robot Navigation
Daniel Cremers	Dense and Direct Methods for 3D Reconstruction and Visual SLAM
Vladyslav Usenko	Direct SLAM Techniques for Vehicle Localization and Autonomous Navigation
Akihiko Torii	Large-scale visual place recognition and online 3D reconstruction
Yasutaka Furukawa	Structured Indoor Modeling and/or Uncanny Valley for 3D Reconstruction
Torsten Sattler	The Limits of Pose Estimation in Very Large Maps

Vision for Aerial, Space and Underwater Robotics

Friedrich Fraundorfer	Drone Vision – Computer vision algorithms for drones
Davide Scaramuzza	From Frames to Events: Vision for High-speed Robotics
Takashi Kubota	Image based Navigation for Exploration Probe
Lazaros Nalpantidis	Stereo Vision for Future Autonomous Space Exploration Robots
Ben Huber	Planetary Robotic Vision Processing for NASA and ESA Rover Missions
Rafael Garcia	Underwater Vision: Robots that “see” beneath the surface

Vision for Scene Understanding

Jürgen Sturm	Tracking and Mapping in Project Tango
Raquel Urtasun	3D Scene Understanding for Autonomous Driving
Andreas Geiger	High-level Knowledge in Low-level Vision
Bernt Schiele	Towards 3D Scene Understanding

Vision for Motion Analysis

Cédric Demonceaux	Pose Estimation and 3D Segmentation using 3D Knowledge in Dynamic Environments
Florian Becker	Recursive Joint Estimation of Dense Scene Structure and Camera Motion in an Automotive Scenario
Johannes Berger	Second-Order Recursive Filtering on the Rigid-Motion Group $SE(3)$ Based on Nonlinear Observations from Monocular Videos
Michael Felsberg	Learning to Drive
Mikael Persson	Structure and Motion -Challenges and solutions for real time geometric estimation from video
Reihard Koch	Model-based Object and Deformation Tracking with Robust Global Optimization

Vision for Autonomous Driving and Robotics

Heiko Hirschmueller	Visual-Inertial Navigation for Mobile Robots
Sven Behnke	Semantic RGB-D Perception for Cognitive Robots
Darius Burschka	Robust Coupling of Perception to Actuation in Dynamic Environments
Juan Andrade-Cetto	Perception for Mobile Robotics
Niko Sünderhauf	Deep Learning for Visual Place Recognition and Online 3D Reconstruction
David Vázquez Bermudez	Learning See in a Virtual World
José Alvaerz	Real-world Semantic Segmentation
Steven Beauchemin	Vehicular Instrumentation for the Study of Driver Intent and Related Applications
Danil Prokhorov	Toward Highly Intelligent Automobiles
Andres Wendel	Realizing Self-Driving Car
Thomas Pock	Efficient Block Optimization Methods for Computer Vision

4 Talks Abstracts

4.1 Real-world Semantic Segmentation

José M. Alvarez (NICTA – Canberra, AU)

Semantic segmentation is a key low-level task to fully understand the environment of vehicle. Ideal semantic segmentation algorithms have four desirable properties: fast, robust, accurate and compact. Semantic segmentation methods must be fast to enable real-time high-level reasoning; Robust to operate at any-time in any weather conditions; Accurate enough to be reliable and, compact to facilitate functionalities in embedded platforms where power and resources are relevant. In this talk we present our recent work towards fast, robust and accurate semantic segmentation in embedded platforms.

4.2 Perception for Mobile Robotics

Juan Andrade-Cetto (UPC – Barcelona, ES)

Joint work of Andrade-Cetto, Juna; Andreasson, Henrik; Corominas, Andreu; Fleuret, Francois; Ila, Viorela; Lippiello, Vincenzo; Moreno-Noguer, Françesc; Peñate-Sanchez, Adriàn; Porta, Josep M.; Sanfeliu, Alberto; Teniente, Ernesto H.; Saarinen, Jari; Santamaria-Navarro, Ángel; Valencia, Rafael; Solá, Joan; Vallvè, Joan; Villamizar, Michael

In this talk I addressed several challenges on perception for mobile robotics. First I overview a pair of object detection and pose recognition algorithms that have the property of being very fast to compute. The first exploits the bootstrapping of very simple features on a boosting classifier. For the second one we propose the use of 3D annotated features. This allows camera pose estimation from low quality monocular images of a previously learned scene. Further in the talk I described our research on visual servoing for UAV manipulation and UAV odometry estimation using tight sensor data fusion. I then proceeded talking about mapping for mobile robots, and in particular about using principled information theoretic metrics to keep the map size tractable. These very same information metrics can also be used for optimal navigation, exploration and optimal sensor placement as explained also in the talk. I concluded the talk with the application of these research results for the autonomous driving of trucks and heavy load AGVs in cargo container terminals.

References

1 V. Ila, J.M. Porta and J. Andrade-Cetto, Information-based compact Pose SLAM, IEEE Transactions on Robotics, 26(1), 78–93, 2010.

2 V. Ila, J.M. Porta and J. Andrade-Cetto, Amortized constant time state estimation in Pose SLAM and hierarchical SLAM using a mixed Kalman-information filter, Robotics and Autonomous Systems, 59(5), 310–318, 2011.

3 A. Peñate-Sanchez, J. Andrade-Cetto and F. Moreno-Noguer, Exhaustive linearization for robust camera pose and focal length estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(10), 2387–2400, 2013.

4 A. Peñate-Sanchez, F. Moreno-Noguer, J. Andrade-Cetto and F. Fleuret, LETHA: Learning from high quality inputs for 3D pose estimation in low quality images, Proc. of 2nd Int'l Conf. on 3D Vision:Tokyo, 517–524, 2014.

5 Á. Santamaria-Navarro and J. Andrade-Cetto, Uncalibrated image-based visual servoing, Proc. of 2013 IEEE Int'l Conf. on Robotics and Automation, Karlsruhe, 5247–5252, 2013.

6 Á. Santamaria-Navarro, V. Lippiello and J. Andrade-Cetto, Task priority control for aerial manipulation, Proceedings of 2014 IEEE Int'l Symp. on Safety, Security and Rescue Robotics, Toyako-cho, 1–6, 2014.

7 Á. Santamaria-Navarro, J. Solá and J. Andrade-Cetto, High-frequency MAV state estimation using low-cost inertial and optical flow measurement units, Proc. of 2015 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems:Hamburg, 1864–1871, 2015.

8 E. H. Teniente and J. Andrade-Cetto, HRA*: Hybrid randomized path planning for complex 3D environments, Proc. of 2013 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, Tokyo, 1766–1771, 2013.

9 R. Valencia, M. Morta, J. Andrade-Cetto and J. M. Porta. Planning reliable paths with pose SLAM, IEEE Transactions on Robotics, 29(4), 1050-1059, 2013.

10 R. Valencia, J. Saarinen, H. Andreasson, J. Vallvè, J. Andrade-Cetto and A. Llilienthal, Localization in highly dynamic environments using dual-timescale NDT-MCL, Proc. of 2014 IEEE Int'l Conf. on Robotics and Automation, Hong Kong, 956–3962, 2014.

11 J. Vallvè and J. Andrade-Cetto, Dense entropy decrease estimation for mobile robot exploration, Proceedings of 2014 IEEE Int'l Conf. on Robotics and Automation, Hong Kong, 6083-6089, 2014.

12 J. Vallvè and J. Andrade-Cetto, Potential information fields for mobile robot exploration, Robotics and Autonomous Systems, 69, 68–79, 2015.

13 J. Vallvè and J. Andrade-Cetto, Active Pose SLAM with RRT*, Proc. of 2015 IEEE Int'l Conf. on Robotics and Automation, Seattle, 2167–2173, 2015.

14 M. Villamizar, F. Moreno-Noguer, J. Andrade-Cetto and A. Sanfeliu, Efficient rotation invariant object detection using boosted random Ferns, Proc. of 2010 IEEE CS Conf. on Computer Vision and Pattern Recognition:San Francisco, 1038–1045, 2010.

15 M. Villamizar, J. Andrade-Cetto, A. Sanfeliu and F. Moreno-Noguer, Bootstrapping boosted random Ferns for discriminative and efficient object classification, Pattern Recognition, 45(9): 3141–3153, 2012.

4.3 Vehicular Instrumentation for the Study of Driver Intent and Related Applications

Steven S. Beauchemin (University of Western Ontario – London, CA)

In this contribution we describe a vehicular instrumentation for the study of driver intent. Our instrumented vehicle is capable of recording the 3D gaze of the driver and relating it to the frontal depth map obtained with a stereo system in real-time, including the sum of vehicular parameters actuator motion, speed, and other relevant driving parameters. Additionally, we describe other real-time algorithms that are implemented in the vehicle, such as a frontal vehicle recognition system, a free lane space estimation method, and a GPS position-correcting technique using lane recognition as land marks.

4.4 Recursive Joint Estimation of Dense Scene Structure and Camera Motion in an Automotive Scenario

Florian Becker (Sony – Stuttgart, DE)

The optical flow induced by a camera moving through a static 3d scene contains valuable information on the geometry. We present an approach to jointly estimating camera motion and a depth map from real-life monocular image sequences which parametrize a flow field. Temporal consistency is exploited in a recursive manner which allows to reduce the estimation task to a series of two-frame problems complemented by an additional temporal smoothness prior. Results for image sequences recorded in a real world traffic scenario are presented.

4.5 Semantic RGB-D Perception for Cognitive Robots

Sven Behnke (Universität Bonn, DE)

Cognitive robots need to understand their surroundings not only in terms of geometry, but they also need to categorize surfaces, detect objects, estimate their pose, etc. In the talk, I will report on efficient methods to address these tasks, which are based on RGB-D sensors. We learn semantic segmentation using random forests and aggregate the surface category in 3D by RGB-D SLAM. We use deep learning methods to categorize surfaces, to recognize objects and to estimate their pose. Efficient RGB-D registration methods are the basis for the manipulation of known objects. They have been extended to non-rigid registration, which allows for transferring manipulation skills to novel objects.

4.6 Second-Order Recursive Filtering on the Rigid-Motion Group SE(3) Based on Nonlinear Observations from Monocular Videos

Johannes Berger (Universität Heidelberg, DE)

Joint camera motion and depth map estimation from observed scene features is a key task in order to reconstruct 3D scene structure using low-cost monocular video sensors. Due to the nonlinear measurement equations that connect ego-motion with the high-dimensional depth map and optical flow, the task of stochastic state-space filtering is intractable.

After introducing the overall problem, the talk focuses on a novel second-order minimum energy approximation that exploits the geometry of $SE(3)$ and recursively estimates the state based on a higher-order kinematic model and the nonlinear measurements. Experimental results for synthetic and real sequences (e.g. KITTI benchmark) demonstrate that our approach achieves the accuracy of modern visual odometry methods.

4.7 Robust Coupling of Perception to Actuation in Dynamic Environments

Darius Burschka (TU München, DE)

I will present methods to represent the state of dynamic environments at a level that is least sensitive to errors in calibration parameters of the sensors. The method is used for a fail-safe implementation of instinctive behaviours on vehicles, like obstacle avoidance. The presented method allows a monitoring of large areas around the vehicle, where a Cartesian representation is not appropriate due to the with distance increasing error in the reconstruction of the three-dimensional information. I will present also the first implementations of this system on a car.

4.8 Direct and Dense Methods for 3D Reconstruction and Visual SLAM

Daniel Cremers (TU München, DE)

Joint work of Cremers, Daniel; Engel, Jakob; Kerl, Christian; Oswald, Martin R; Stuckler, Jorg; Sturm, Jürgen; Stuhmer, Jan; Usenko, Vladyslav

The reconstruction of the 3D world from images is among the central challenges in computer vision. Having started in the 2000s, researchers have pioneered algorithms which can reconstruct camera motion and sparse feature-points in real-time. In my talk, I will present spatially dense methods for camera tracking and reconstruction. They do not require feature point estimation, they exploit all available input data and they recover dense geometry rather than sparse point clouds.

4.9 Pose Estimation and 3D Segmentation using 3D Knowledge in Dynamic Environments

Cédric Demonceaux (University of Bourgogne, FR)

When 2D and 3D cameras observe the same scene, their measurements are usually complementary to each other for scene reconstruction and understanding. A classic example includes 2D cameras capturing high quality texture information and 3D cameras providing accurate location of the scene points. Fusing these complementary measurements has many potential applications such as change detection, scene gaps filling, camera pose correction, and visual odometry. In this talk, we show that the 3D localization of 2D cameras can be improved knowing 3D structure of the scene. Thus, we develop two methods using 3D information on the scene for camera pose estimation. The first one doesn't require 3D feature extraction and doesn't need any geometric hypothesis but it converges in a local optimum. The second one supposes that the scene is composed of planar patches and converges to the global optimum. Then, this 3D information will be used conjointly with 2D images for extracting and reconstructing the background and the dynamic objects of the scene.

4.10 Learning to Drive

Michael Felsberg (Linköping University, SE)

Driving a car is a prototypical example for graded autonomy, where the human driver and the assistance system co-operate. There are various legal, technological, and practical reasons why the human driver is kept in the loop and should be able to override the system's decisions. However, the co-operation, and thus graded autonomy, should be more than just taking over power of command. It is desirable that the assistance system seamlessly acquires new capabilities during the driver's manual intervention, in order to increase the level of autonomy in subsequent operation. Thus, the task for the system is to use input, or precepts, as observed by the human user and output, or actions, as provided by the human user to extend the systems capabilities on the fly. We address this task in the present work and propose a novel approach to online perception-action learning, which lifts human-machine interaction clearly beyond current methods based on switching command. We evaluate our approach on the problem of road following with a model RC-car on reconfigurable tracks indoors, outdoors, and at night. The system's capabilities are continuously extended by interchangeably applying supervised learning (by demonstration), instantaneous reinforcement learning, and unsupervised learning (self-reinforcement learning). The resulting autonomous capabilities go beyond those of existing methods and of the human driver with respect to speed, accuracy, and functionality.

4.11 Drone Vision – Computer Vision Algorithms for Drones

Friedrich Fraundorfer (TU Graz, AT)

Drone Vision – Computer Drones are small scale flying robots and it is predicted that the drone market will see a major growth in the near future. Computer vision will play a major role in controlling and developing autonomous drones. My proposal is to utilize tight IMU-vision coupling for ego-motion estimation of drones. This will result in a new class of fundamental algorithms for ego-motion estimation, being more robust and lots faster. IMU measurements can be used to transform the complex estimation problems into a simpler formulation of vision algorithms for drones.

4.12 Structured Indoor Modeling and/or Uncanny Valley for 3D Reconstruction

Yasutaka Furukawa (Washington University – St. Louis, US)

Depending on how our discussion will go, I will give a talk on one of the following two topics or a mix.

Structured indoor reconstruction. We propose a novel indoor scene reconstruction algorithm. The approach produces a structured 3D model in a top-down manner. The reconstruction algorithm is driven by a indoor structure grammar. The new model representation enables many new applications such as novel indoor scene visualization, inverse CAD, floorplan generation, and tunable reconstruction.

Uncanny Valley for 3D Reconstruction. Accurate 3D reconstruction is usually the key to high quality visualization applications. However, very often, improving reconstruction accuracy degrades the quality of visualization. This issue is little known to researchers, yet very important in practice.

4.13 Underwater Vision: Robots that "see" beneath the surface

Rafael Garcia (University of Girona, ES)

Using vision underwater is a difficult endeavor due to the transmission properties of the medium. Light is absorbed and scattered by the particles suspended in the water column, producing degraded images with limited range, blurring, low contrast and weak colours, among other effects. Moreover, artificial lighting tends to provide non-uniform illumination and introduces shadows in the scene, generating a motion flow that does not obey the dominant motion of the camera.

However, with the adequate processing pipeline, vision can be a powerful tool for underwater robots to explore the ocean.

In this talk we will explain how computer vision techniques can be adapted to the underwater environment if we understand and deal with the different associated problems. An approach to create accurate three-dimensional textured models of the seafloor will be presented. The method de-hazes the images to improve signal-to-noise ratio, then generates a dense cloud of 3D points and is able to compute a meshed surface being robust to common defects in underwater imaging such as high percentage of outliers (due to light backscatter) and point-cloud noise (due to the blurring of forward scattering).

4.14 Semantic Maps for High Level Robot Navigation

Antonios Gasteratos (Democritus University of Thrace – Xanthi, GR)

In the near future domestic robots should be equipped with the potential of producing meaningful internal retalks of their own environment, allowing them to cope a wide range of real-life tasks. Intense research efforts occur to build cognitive robots able to perceive and understand their surroundings in a human-centred manner. Semantic mapping in mobile robotics can constitute a definite solution for this challenge. The semantic map is an augmented representation of the environment that –supplementary to the geometrical knowledge – encapsulates characteristics compatible with human understanding. It provides several algorithmic opportunities for innovative development of applications that will eventually lead to the human robot interaction. This talk will describe the construction of accurate and consistent semantic maps facilitating adequate robot deployment in domestic environments.

4.15 High-level Knowledge in Low-level Vision

Andreas Geiger (MPI für Intelligente Systeme – Tübingen, DE)

1. Stereo techniques have witnessed tremendous progress over the last decades, yet some aspects of the problem still remain challenging today. Striking examples are reflecting and textureless surfaces which cannot easily be recovered using traditional local regularizers. In this work, we therefore propose to regularize over larger distances using object-category specific disparity proposals (displets) which we sample using inverse graphics techniques based on a sparse disparity estimate and a semantic segmentation of the image. The proposed displets encode the fact that objects of certain categories are not arbitrarily shaped but typically exhibit regular structures. We integrate them as non-local regularizer for the challenging object class "car" into a superpixel based CRF framework and demonstrate its benefits on the KITTI stereo evaluation.
2. This work proposes a novel model and dataset for 3D scene flow estimation with an application to autonomous driving. Taking advantage of the fact that outdoor scenes often decompose into a small number of independently moving objects, we represent each element in the scene by its rigid motion parameters and each superpixel by a 3D plane

as well as an index to the corresponding object. This minimal representation increases robustness and leads to a discrete-continuous CRF where the data term decomposes into pairwise potentials between superpixels and objects. Moreover, our model intrinsically segments the scene into its constituting dynamic components. We demonstrate the performance of our model on existing benchmarks as well as a novel realistic dataset with scene flow ground truth. We obtain this dataset by annotating 400 dynamic scenes from the KITTI raw data collection using detailed 3D CAD models for all vehicles in motion. Our experiments also reveal novel challenges which can't be handled by existing methods.

4.16 Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition

Michal Havlena (ETH Zürich, CH)

Structure-based localization is the task of finding the absolute pose of a given query image w.r.t. a pre-computed 3D model. While this is almost trivial at small scale, special care must be taken as the size of the 3D model grows, because straight-forward descriptor matching becomes ineffective due to the large memory footprint of the model, as well as the strictness of the ratio test in 3D. Recently, several authors have tried to overcome these problems, either by a smart compression of the 3D model or by clever sampling strategies for geometric verification. Here we explore an orthogonal strategy, which uses all the 3D points and standard sampling, but performs feature matching implicitly, by quantization into a fine vocabulary. We show that although this matching is ambiguous and gives rise to 3D hyperpoints when matching each 2D query feature in isolation, a simple voting strategy, which enforces the fact that the selected 3D points shall be co-visible, can reliably find a locally unique 2D-3D point assignment. Experiments on two large-scale datasets demonstrate that our method achieves state-of-the-art performance, while the memory footprint is greatly reduced, since only visual word labels but no 3D point descriptors need to be stored.

4.17 Visual-Inertial Navigation for Mobile Robots

Heiko Hirschmuller (Roboception GmbH – München, DE)

Joint work of Hirschmuller, Heiko; Schmid, Korbinian; Suppa, Michael; Lutz, Philipp; Tomic, Teodor; Mair Elmar

Navigation of mobile robots is still challenging, especially in environments that are not prepared for robotics, like at home or outdoors. At the German Aerospace Center (DLR) we have developed passive stereo-vision based navigation that is supported by an inertial measurement unit (IMU). The ego-motion estimation is precise due to visual odometry, robust due to the IMU and fulfils hard-real time constraints for using it directly in the control loop of robots, like for autonomously flying highly agile quadcopters. Several experiments with rovers and flying systems in mixed indoor/outdoor settings proved the concept.

In the DLR spin-off Roboception GmbH we are going to bring such technology as plug and play device into the marked. The talk covers the main concepts and new developments. One of them is the extension of the Semi-Global Matching method for delivering not just disparity, but also error and confidence values for each pixel. The error is given in disparities and has a value of 0.5 for most of the pixels, but can go up to 2. The confidence is the probability that the true disparity is within a three times error interval around the measured disparity. Thus, the error is seen as the standard deviation of an Gaussian error, while the confidence is the probability that the measured value is not an outlier.

References

1 Heiko Hirschmuller, Korbinian Schmid and Michael Suppa, Computer vision for mobile robot navigation, Proceedings of Photogrammetric Week 2015:Stuttgart, 143–154, 2015.

2 Korbinian Schmid, Philipp Lutz, Teodor Tomic, Elmar Mair and Heiko Hirschmuller, Autonomous vision-based micro air vehicle for indoor and outdoor navigation, Journal of Field Robotics, Special Issue on Low Altitude Flight of UAVs, 31(4), 537–570 2014.

3 Heiko Hirschmuller, Stereo Processing by semi-global matching and mutual information, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2), 328–341, 2008.

4.18 Planetary Robotic Vision Processing for Rover Missions

Ben Huber (Joanneum Research – Graz, AT)

The international community of planetary science and exploration has successfully launched, landed and operated about thirty human and robotic missions to the planets and the Moon. They have collected differing numbers of surface imagery that have only been partially utilized throughout these missions and thereafter for further scientific application purposes. The data for most of these missions including meta-data is publicly available. Many of the mentioned missions rely on stereo imagery for navigation and offer huge datasets that can be reconstructed in 3D and put into a common coordinate context. By doing this we generate datasets with a huge benefit for scientific geological analysis in rendered 3D space based on mission data that has been almost forgotten in planetary data archives.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 312377 PRoViDE.

4.19 Model-based Object and Deformation Tracking with Robust Global Optimization

Reinhard Koch (Universität Kiel, DE)

Model-based analysis, also termed Analysis-by-Synthesis or Analysis from Generative models, is a powerful tool to solve ill-posed problems like 3D object surface tracking from images. A parametric model of the object in focus is generated and the visual appearance and motion of the object is synthesized from the model and compared with the visual input by a cost or fitness function. Adapting the model parameters according to the cost function solves the tracking problem. In order to cope with possibly high-dimensional parameter space, efficient and robust non-local stochastic estimators are needed. In my talk I will outline the AbS principle and discuss the optimizer and applications. Examples are tracking of multiple animals in confined housing, deformation of thin plate models for human user interaction, and others.

4.20 Image based Navigation for Exploration Probe

Takashi Kubota (ISAS/JAXA, JP)

This talk firstly introduces future lunar or planetary exploration plans, which consist of lunar, Mars and asteroid exploration. Then robotics technology is shown for lunar or planetary exploration. Vision system makes important roles in deep space exploration for efficient and safe exploration. This talk presents the intelligent system for navigation, path planning, sampling, etc. Especially image based navigation schemes are presented in detail.

4.21 Stereo Vision for Future Autonomous Space Exploration Robots

Lazaros Nalpantidis (Aalborg University Copenhagen, DK)

Space exploration rovers need to be highly autonomous because the vehicle should spend as much of its traverse time as possible moving, rather than waiting for delayed and often interrupted teleoperation commands. Autonomous behavior can be supported by vision systems that provide wide views to enable navigation and 3D reconstruction, as well as close-up views ensuring safety and providing reliable odometry data.

This talk presents the design, development and testing of such a stereo vision system for a space exploration rover. This system was designed with the intention of being efficient, low-cost, accurate and was ultimately implemented on an FPGA platform. We are discussing our experiences with this system and highlight useful lessons learned.

4.22 Structure and Motion – Challenges and solutions for real time geometric estimation from video

Mikael Persson (Linköping University, SE)

In this talk I introduce the cv4x SfM system, which achieved state of the art results on the challenging KITTI odometry benchmark earlier this year(2015). I motivate the choice in reconstruction method, the bootstrap tracking by matching scheme used and how perceptual aliasing was addressed. Visual odometry systems such as cv4x, in particular if fused with a strong loop closure system, achieve excellent and more importantly sufficient results, but several challenges remain: The trajectory prediction of independently moving objects is of particular interest in autonomous driving and is principally, if not practically, the same problem as IMU-free VO and as such my current focus of research.

4.23 Efficient Block Optimization Methods for Computer Vision

Thomas Pock (TU Graz, AT)

In this talk I will discuss recent advances in block optimization methods for minimizing non-smooth optimization problems in computer vision and image processing. It turns out that a large class of 2D and 3D total-variation regularized problems can be reduced to an algorithm that computes exact solutions with respect to certain subsets of the variables in each iteration. For example, if the subsets are 1D total variation problems, we can efficiently compute their solutions based on dynamic programming. Furthermore, we can make use of gradient acceleration techniques to additionally speed up the algorithms. I will show applications to computing globally optimal minimizers of total variation regularized stereo problems.

4.24 Toward Highly Intelligent Automobiles

Danil V. Prokhorov (Toyota Research Institute North America – Ann Arbor, US)

Intelligent automobiles a.k.a. self-driving, autonomous or highly automated cars are capturing people's imagination while opening up new opportunities for research in many areas including robotics, machine learning and vision. In my talk I overview the state of art in highly automated cars and discuss an example of near-production AHDA car of Toyota, as well as my personal experience with ACC-equipped vehicles. Then I discuss an example of on-going research project of a car capable of making a variety of autonomous decisions on public roads, with the focus on the roundabout maneuver. This maneuver illustrates an importance of the holistic perception-action system approach, rather than module-by-module considerations still prevalent in this field. In conclusion I offer my view of open challenges for intelligent automobiles.

4.25 Efficient Multi-view Semantic Segmentation

Hayko Riemenschneider(ETH Zürich, CH)

There is an increasing interest in semantically annotated 3D models, e.g. of cities. The typical approaches start with the slow semantic labelling of all the images used for the 3D model. The inherent redundancy among the overlapping images calls for more efficient solutions. This work deals with two an alternative approaches. First, we exploit the geometry of a 3D mesh model to predict the best view before the actual labelling. For this we find the single image part that bests supports the correct semantic labelling of each face of the underlying 3D mesh. Second, we directly use the 3D point cloud itself, skipping any image processing entirely. This pure 3D approach relies solely on 3D surface features (and a bit of classic RGB) and provides state-of-the-art results without any heavy 2D image features. In both works we show how to significantly speedup the semantic segmentation and even increase the accuracy – leaving the question – how much of 2D images is needed for 3D semantic segmentation?

4.26 The Limits of Pose Estimation in Very Large Maps

Torsten Sattler (ETH Zürich, CH)

In many applications of autonomous vehicles, we can safely assume that there exists a 3D map of the scene the vehicle operates in, which can be used for navigation and localization. One major problem of maps covering a very large area is that they contain many structures with globally repeating appearance, causing problems when trying to match structures between a query image and the map. Common large-scale localization approaches operate under the assumption that we can recover the pose of the query image as long as we find enough good matches and as long as the pose estimation process is robust enough to large quantities of wrong matches. Unfortunately, current pose estimation strategies have problems dealing with too many matches. In this talk, I will discuss a truly scalable pose estimation strategy. Using this strategy, we will show that there are limits to the approach of just using more and more matches and hoping that pose estimation will be able to recover the correct pose.

4.27 From Frames to Events: Vision for High-speed Robotics

Davide Scaramuzza (Universität Zürich, CH)

Autonomous micro drones will soon play a major role in search-and-rescue and remote-inspection missions, where a fast response is crucial. They can navigate quickly through unstructured environments, enter and exit buildings through narrow gaps, and fly through collapsed buildings. However, their speed and maneuverability are still far from those of birds. Indeed, agile navigation through unknown, indoor environments poses a number of challenges for robotics research in terms of perception, state estimation, planning, and control. In this talk, I will give an overview of my research activities on visual inertial navigation of quadrotors, from slow navigation (using standard frame-based cameras) to agile flight (using event-based cameras).

4.28 Towards 3D Scene Understanding

Bernt Schiele (MPI für Informatik – Saarbrücken, DE)

Inspired by the ability of humans to interpret and understand 3D scenes nearly effortlessly, the problem of 3D scene understanding has long been advocated as the "holy grail" of computer vision. In the early days this problem was addressed in a bottom-up fashion without enabling satisfactory or reliable results for scenes of realistic complexity. In recent years there has been considerable progress on many sub-problems of the overall 3D scene understanding problem. As the performance for these sub-tasks starts to achieve remarkable performance levels we argue that the problem to automatically infer and understand 3D scenes should be addressed again.

This talk highlights recent progress on some essential components (such as object recognition and person detection), on our attempt towards 3D scene understanding, as well as on our work towards activity recognition and the ability to describe video content with natural language. These efforts are part of a longer-term agenda towards visual scene understanding. While visual scene understanding has long been advocated as the “holy grail” of computer vision, we believe it is time to address this challenge again, based on the progress in recent years.

4.29 Tracking and Mapping in Project Tango

Jürgen Sturm (Google – München, DE)

Google’s Project Tango aims to provide a mobile solution for visual-inertial 6-DOF motion estimation and dense 3D reconstruction. In my talk, I will give a technical presentation of the algorithms underlying the Tango API, including visual-inertial odometry, SLAM, loop closure detection, re-localization and 3D reconstruction. During my talk, I will present several live demos on a Tango tablet.

4.30 Deep Learning for Visual Place Recognition and Online 3D Reconstruction

Niko Sünderhauf (Queensland University of Technology – Brisbane, AU)

In the first part of this talk I will summarize our recent work on visual place recognition in changing environments using deep convolutional network features.

In the second part I talk about some lessons learned when applying ConvNets in robotics (e.g. for object detection on a mobile robot) and the gaps between the computer vision community and robotics in that particular area. I hope to induce a discussion on what we as a community can do to bridge this gap.

4.31 Large-scale Visual Place Recognition – Current challenges

Akihiko Torii (Tokyo Institute of Technology, JP)

Large-scale visual place recognition (VPR) takes an important role for localization of robots and autonomous cars, e.g. rough initial localization. In this seminar, we first compare key properties of compact image descriptors – Bag of Visual Words (BoVW) and Vector of Locally Aggregated Descriptors (VLAD) – popularly used in VPR. On top of the decent analysis of these image retalks, we discuss challenges in VPR, e.g. repetition, illumination changes, change of seasons, and aging that give major appearance changes among testing-query and database images. We show that the adaptive soft-assignment scheme on BoVW is effective on the street-level visual place recognition. We also show dense feature detection followed by VLAD representation gives a significant improvement in localization performance and expanding the database by view synthesis gives an additional gain on the challenging datasets.

4.32 3D Scene Understanding for Autonomous Driving

Raquel Urtasun (University of Toronto, CA)

Developing autonomous systems that are able to assist humans in everyday's tasks is one of the grand challenges in modern computer science. Notable examples are personal robotics for the elderly and people with disabilities, as well as autonomous driving systems which can help decrease fatalities caused by traffic accidents. In order to perform tasks such as navigation, recognition and manipulation of objects, these systems should be able to efficiently extract 3D knowledge of their environment. In this talk, I'll show how graphical models provide a great mathematical formalism to extract this knowledge. In particular, I'll focus on a few examples, including 3D reconstruction, 3D object and layout estimation and self-localization.

4.33 Direct SLAM Techniques for Vehicle Localization and Autonomous Navigation

Vladyslav Usenko (TU München, DE)

Joint work of Cremers, Daniel; Engel, Jakob; Jorg Stuckler, Jorg; von Stumberg, Lukas; Usenko, Vladyslav

Localization and mapping are two very important challenges for autonomous vehicles. Even though many different types of sensors can be used for this purpose, camera based solutions gain popularity because of the low costs, small weight and simple mechanical design. In my talk I present several extensions to the LSD-SLAM – camera based large-scale direct semi-dense slam method, that enable reliable operation on the real-world data from the vehicles. In particular, I present an extension of the method to the stereo-camera setup and tight integration with Inertial Measurement Unit, and demonstrate an autonomous exploration and control on a consumer grade flying robot.

4.34 Learning See in a Virtual World

David Vázquez Bermudez (Autonomous University of Barcelona, ES)

The ADAS group from the Computer Vision Center based at the Universitat Autònoma de Barcelona, has an extensive experience developing ADAS systems such as Lane Departure Warning, Collision Warning, Automatic Cruise Control, Pedestrian Protection, Headlights Control, etc. Currently ADAS is developing an Autonomous Vehicle based on relatively cheap sensors such cameras, IMU and GPS. In this talk we will give a short overview of the ADAS systems developed until now by the group and the Autonomous Vehicle project. Then we will explain in more detail two parts of the autonomous vehicle that has been awarded by the IEEE Intelligent Transportation Systems Society Spanish Chapter. The use of virtual images to training models that are able to operate in a real world (Best Ph.D. Thesis award) and a vehicle localization system based on GPS, IMU and cameras (Accesit M.Sc award).

4.35 Realizing Self-Driving Car

Andres Wendel (Google Inc. – Mountain View, US)

Self-driving vehicles are coming. They will save lives, save time and offer mobility to those who otherwise don't have it. Eventually they will reshape the world we live in. A dedicated team at Google has spent the last few years moving self-driving vehicles closer to reality. New algorithms, increased processing power, innovative sensors and massive amounts of data enable our vehicles to see further, understand more and handle a wide variety of challenging driving scenarios. Our vehicles have driven over a million miles on highways, suburban and urban streets. Through this journey, we've learned a lot; not just about how to drive, but about interacting with drivers, users and others on the road, and about what it takes to bring an incredibly complex system to fruition. In my talk, I share some insights in how the technology works, how we have rolled out our new prototype vehicles to public roads, and which edge case situations we have to solve.

5 Working groups

5.1 Sensing

Editor: Andrés Bruhn
Topic: Low-Level Sensing

5.1.1 Workgroup members in alphabetical order

Andrés Bruhn	(Universität Stuttgart, DE)
Florian Becker	(Sony – Stuttgart, DE)
Johannes Berger	(Universität Heidelberg, DE)
Darius Burschka	(TU München, DE)
Ben Huber	(Joanneum Research-Graz, AT)
Reinhard Koch	(Universität Kiel, DE)
Lazaros Nalpantidis	(Aalborg University Copenhagen, DK)
Thomas Pock	(TU Graz, AT)

5.1.2 Discussion Summary

This working group discussed aspects of low-level sensing methods for autonomous vehicles and probes. While most systems for autonomous driving rely on the same types of modules – e.g. algorithms for motion estimation, stereo reconstruction, and scene flow computation – those modules are typically designed and evaluated separately from the remaining system. Evidently, this makes it difficult to integrate feedback in terms of scene understanding, which would be likely to improve the robustness of such algorithms in difficult situations, i.e. under adverse weather conditions. Moreover, the learning of suitable models or model components for specific scenarios is becoming increasingly important as the integration of previously learned priors may improve the quality of the algorithms as well. Also from an evaluation viewpoint, there is a clear need for improvement. Currently there is a clear lack of suitable benchmarks to evaluate the quality of vision algorithms for autonomous driving. While there are at least some benchmarks that can be used to evaluate the performance of low-level algorithms separately (e.g. the KITTI Benchmark Suite), it remains unclear which accuracy and robustness demands complex systems for autonomous driving actually have w.r.t. to the performance of their underlying modules. Hence it is hardly possible to predict the performance and robustness of such methods for real applications such as autonomous driving. Finally, the use of different hardware for image acquisition may significantly improve both performance and speed of low-level algorithms. One the one hand, one may consider the use of high speed cameras to avoid ambiguous large displacements which still pose a problem for most applications. On the other hand, it may be worthwhile to investigate the usefulness of “differential” cameras that allow an reduction of the processed data by only providing information in terms of image changes. In detail, the following research questions have been discussed:

How can low-level models be further improved?

- flexible system design (modules) vs. robustness (high level knowledge).
- however: jointly solving strongly related tasks may improve performance.
- need of joint modelling and inference in terms of holistic approaches.
- moreover: learning of model components based on given application.

How can benchmarking be improved towards practical relevance?

- in practice: absolute accuracy not that important.
- algorithm must be "sufficiently accurate" for a certain application.
- evaluation as part of the entire vision system.
- robustness matters: determine breaking point under certain degradations.

How can the image acquisition process be improved?

- in general: higher frame rates desirable for motion estimation.
- simpler algorithms sufficient → faster computation.
- less complex motion, smaller displacements → higher accuracy.

What are suitable representations when extracting information?

- typically: not all pixels needed for making decisions.
- only consider locations that deviate from expected behaviour (e.g. intensity changes, deviations from high-level models).
- "differential" cameras (e.g. event cameras).

5.2 Mapping for Autonomous Vehicles and Probes

Editor: Hayko Riemenschneider
Topic: Offline Mapping

5.2.1 Workgroup members in alphabetical order

Yasutaka Furukawa	(Washington University St. Louis, US)
Antonios Gasteratos	(Democritus University of Thrace – Xanthi, GR)
Michal Havlena	(ETH Zürich, CH)
Hayko Riemenschneider	(ETH , Zürich, CH)
Torsten Sattler	(ETH Zürich, CH)
Akihiko Torii	(Tokyo Institute of Technology, JP)
Vladyslav Usenko	(TU München, DE)

5.2.2 Discussion Summary

The what, where, when, how, and who of mapping. This working group defined mapping as the process to create (offline/online) environment maps including road an urban environment, lane markings and traffic symbols as well as dynamic obstacles like pedestrians or weather conditions. One main topic is the distinction between offline prior mapping and online mapping. The group concluded that the hard cases, those which currently pose the most challenges (dynamic objects and up to date information), can only be solved in online mapping whereas offline mapping can provide a solid environment yet by definition will always be out of date. Hence, the question arises of the use cases for offline mapping, e.g. route navigation planning.

How to technically create environment maps?

- offline mapping will benefits from the richness of all sensors (vision, LIDAR, etc).
- online mapping also, yet for real time purposes needs specializations (instant LIDAR results vs stixel like abstractions).
- transfer manual annotation from 2D to 3D or vice versa, needed for guarantee on quality.
- define levels of details for roads, surroundings and buildings.
- formalism of maps, structures, relationships of contents in there (roads are connecting).

When to create environment maps?

- temporally changing maps, need for continuously updating.
- integrating visual information acquired by different companies, people, cars, ...
- collecting data, by own cars, taxi, trams, community services.
- how fast should be the update vs on the drive will always be fastest.
- how to integrate multimodal data coming apart from visual data.
- long term changes, building construction.
- short term changes, e.g. parking cars, construction sites, people movement updating maps give more useful prior information for autonomous driving/routing, e.g. once construction signs found, we have no need to drive there.
- other important issues: security, redundancy, fall back of the map creation.

Who is responsible for the creation of maps?

- service levels agreement for quality.
- will there be a uniform/standard format of global map?
- consortium of car companies and users to define these standards no common maps since business model. coverage of the countries in terms of where are maps needed and to what detail.
- accuracy of the maps w.r.t. coverage, not everywhere is a cm/time accuracy needed.
- crowdsourcing for every car.
- human control and verification of structural changes (suggested by vision, LIDAR).

Where and what is included in the environment maps?

- we should know limitations of online/offline mapping!
- what is the difference between online/offline mapping?
- online: annotation, change detection → impossible to do online. only the surrounding areas.
- only does simple dynamic obstacles/events detection (no understanding what it is).
- what if non standard events happen. if the roads are covered by bus. deadlock situation.
- offline: semantics: soft not binary decision: continuous occlusion space and object classes.
- classifiers on actions/intent of others, to allow high level interpretation when breaking rules.

5.3 Beyond Deep Learning

Editor:	Michael Felsberg
Topic:	Deep Learning

5.3.1 Workgroup members in alphabetical order

Andreas Geiger	(MPI für Intelligente Systeme – Tübingen, DE)
Atsushi Imiya	(Chiba University, JP)
Bernt Schiele	(MPI für Informatik – Saarbrücken, DE)
José M. Alvarez	(NICTA – Canberra, AU)
Jürgen Sturm	(Google – München, DE)
Michael Felsberg	(Linköping University, SE)
Niko Sünderhauf	(Queensland University of Technology – Brisbane, AU)
Rafael Garcia	(University of Girona, ES)
Raquel Urtasun	(University of Toronto, CA)
Sven Behnke	(Universität Bonn, DE)

5.3.2 Discussion Summary

Recurrent and Dynamic Networks with Structural Models. Future work will have to address procedural fundamentals of the learning algorithm and the network:

- How to realize deep learning of recurrent networks?
- How to realize networks that learn layered dynamic processes?
- How to regularize with known structural and geometrical models?
- How to enforce invariance beyond shift invariance?
- How to inject hard constraints?

These issues establish an engineering – understanding trade-off. Advanced visualizations and modelling of solution manifolds are required for future progress.

Learning Process and Training Data. It has been reflected that previous research often evaluated sub-tasks, such as detection or recognition, instead of system-level performance. The latter will, in most cases, require embodiment and thus perception-action learning. Future challenges will include:

- How to perform reinforcement learning on deep networks?
- How to generate training data with sufficient volume and quality?
- How to synthesize and augment data?
- How to regularize learning with known stochastic models to avoid overfitting?
- How to assess performance on system-level tasks in relation to other techniques such as random forests?

These issues establish a major challenge on the empirical analysis of deep learning. An evaluation methodology has to be developed to assess progress properly.

Participants

- José M. Alvarez
NICTA – Canberra, AU
- Juan Andrade-Cetto
UPC – Barcelona, ES
- Steven S. Beauchemin
University of Western Ontario – London, CA
- Florian Becker
Sony – Stuttgart, DE
- Sven Behnke
Universität Bonn, DE
- Johannes Berger
Universität Heidelberg, DE
- Andrés Bruhn
Universität Stuttgart, DE
- Darius Burschka
TU München, DE
- Daniel Cremers
TU München, DE
- Krzysztof Czarnecki
University of Waterloo, CA
- Cédric Demonceaux
University of Bourgogne, FR
- Michael Felsberg
Linköping University, SE
- Friedrich Fraundorfer
TU Graz, AT
- Yasutaka Furukawa
Washington University – St. Louis, US
- Rafael Garcia
University of Girona, ES
- Antonios Gasteratos
Democritus Univ. of Thrace – Xanthi, GR
- Andreas Geiger
MPI für Intelligente Systeme – Tübingen, DE
- Michal Havlena
ETH Zürich, CH
- Heiko Hirschmuller
Roboception GmbH – München, DE
- Ben Huber
Joanneum Research – Graz, AT
- Atsushi Imiya
Chiba University, JP
- Reinhard Koch
Universität Kiel, DE
- Takashi Kubota
ISAS/JAXA – Sagamihara, JP
- Lazaros Nalpantidis
Aalborg Univ. Copenhagen, DK
- Mikael Persson
Linköping University, SE
- Thomas Pock
TU Graz, AT
- Danil V. Prokhorov
Toyota Research Institute North America – Ann Arbor, US
- Sebastian Ramos
Daimler AG-Boblingen, DE
- Hayko Riemenschneider
ETH – Zürich, CH
- Torsten Sattler
ETH Zürich, CH
- Davide Scaramuzza
Universität Zürich, CH
- Bernt Schiele
MPI für Informatik – Saarbrücken, DE
- Jürgen Sturm
Google – München, DE
- Niko Sünderhauf
Queensland University of Technology – Brisbane, AU
- Akihiko Torii
Tokyo Institute of Technology, JP
- Raquel Urtasun
University of Toronto, CA
- Vladyslav Usenko
TU München, DE
- David Vázquez Bermudez
Autonomus University of Barcelona, ES
- Andreas Wendel
Google Inc. – Mountain View, US
- Christian Winkens
Universitat Koblenz-Landau, DE