Automated Valet Parking and Charging for e-Mobility
Results of the V-Charge Project*

Ulrich Schwesinger[+1], Mathias Bürki[1], Julian Timpner[2], Stephan Rottmann[2], Lars Wolf[2],
Lina Maria Paz[3], Hugo Grimmett[3], Ingmar Posner[3], Paul Newman[3], Christian Häne[4],
Lionel Heng[5], Gim Hee Lee[6], Torsten Sattler[4], Marc Pollefeys[4], Marco Allodi[7], Francesco Valenti[7],
Keiji Mimura[8], Bernd Goebelsmann[8], Wojciech Derendarz[9], Peter Mühlfellner[9], Stefan Wonneberger[9],
Rene Waldmann[9], Sebastian Grysczyk[9], Carsten Last[9], Stefan Brüning[10], Sven Horstmann[10],
Marc Bartholomäus[10], Clemens Brummer[10], Martin Stellmacher[11], Fabian Pucks[11], Marcel Nicklas[11]
and Roland Siegwart[1]

Abstract— Automated valet parking services provide great potential to increase the attractiveness of electric vehicles by mitigating their two main current deficiencies: reduced driving ranges and prolonged refueling times. The European research project V-Charge aims at providing this service on designated parking lots using close-to-market sensors only. For this purpose the project developed a prototype capable of performing fully automated navigation in mixed traffic on designated parking lots and GPS-denied parking garages with cameras and ultrasonic sensors only. This paper summarizes the work of the project, comprising advances in network communication and parking space scheduling, multi-camera calibration, semantic mapping concepts, visual localization and motion planning. The project pushed visual localization, environment perception and automated parking to centimetre precision. The developed infrastructure-based camera calibration and semi-supervised semantic mapping concepts greatly reduce maintenance efforts. Results are presented from extensive month-long field tests.

Index Terms— Automated driving, autonomous vehicle, calibration, visual localization, motion planning

I. INTRODUCTION

The fight against climate change will among other measures require new mobility concepts. More widespread use of electric vehicles is one of the key measures that can help to reduce CO_2 emissions, which - to a significant amount - are produced by combustion engine cars. Reducing the traffic related to search of parking spots is another relevant aspect.

The V-Charge project tackles these issues by introducing a new mobility concept mitigating the two current deficiencies of electric vehicles: reduced driving ranges and increased refuelling duration. Driverless parking and charging is promoted to ease the traveller's transfer from individual short-range traffic with his or her electric vehicle to long haul public traffic by train or airplane. Relieving the traveller from the time-consuming task of searching for a parking space will render public transportation more appealing.

Fig. 1: The V-Charge VW e-Golf prototype performs fully automated navigation on parking lots using solely close-to-market sensors such as cameras and ultrasonic sensors.

This concept requires fully automated driving in indoor and outdoor parking areas. The V-Charge project sets itself the challenging task to achieve such automated driving capabilities with close-to-market, low-cost sensors only. Utilising a sensor setup only consisting of four monocular fisheye cameras, two stereo cameras and stock ultrasonic sensors helps to reduce the vehicle-side hardware costs, yet requires increased scientific and engineering efforts related to processing the sensor data obtained. Thanks to the automated valet parking service, charging stations can be freed up each time the charging process has finished - all without human intervention. This, as well as the possibility of high density parking, offers substantial benefits to parking lot operators. In order keep the installation costs low, V-Charge only requires a communication channel and a server - but no further modifications to the parking lot infrastructure.

From a scientific and engineering point of view the realisation of the V-Charge concept requires progress in the state of the art in various research domains. Automated indoor navigation without modifications to infrastructure and environment requires GPS independent localisation with on-board sensors only. Operation in parking lots utilised by both automated and manually operated cars requires detection, classification and estimation of the intents of other traffic participants. Fully automated parking in tight spaces calls for precise environment perception and control of the vehicle. And last but not least a convenient interface for the user is required making drop-off and pick-up of the vehicle as easy as possible, despite the complex scheduling algorithms for parking space and charging station assignment operating in the background.

[+]Corresponding author: ulrich.schwesinger@mavt.ethz.ch, [1]Autonomous Systems Lab, ETH Zurich, [2]IBR, Technische Universität Braunschweig, [3]University of Oxford, [4]Computer Vision and Geometry Group, Department of Computer Science, ETH Zurich, [5]DSO National Laboratories, [6]National University of Singapore, [7]VisLab, Università degli Studi di Parma, [8]Robert Bosch GmbH, Corporate Research, [9]Volkswagen AG, [10]Carmeq GmbH, [11]IAV GmbH

In the course of the project's four years, three fully functional prototypes capable of automated operation on outdoor parking lots and indoor parking garages were developed. Two combustion engine VW Golf VI platforms and one fully electric VW e-Golf prototypes were equipped with close-to-market cameras and navigation software developed in the project. The prototypes were successfully demonstrated at various locations including several indoor parking garages as well as one medium-scale outdoor parking lot. The fully automated valet parking functionality was shown in the low-speed domain of up to 10 km/h. The intermediate results of the project were presented in [1]. The following chapters summarize the project's ultimate state and report on the final results.

II. PLATFORM AND SENSOR SETUP

The V-Charge prototype depicted in Fig. 1 is based on a VW e-Golf platform. It is equipped with front- and rear facing stereo cameras with a horizontal field of view (FOV) of $45°$ and $120°$ respectively. Four $1.3\,\mathrm{Mpx}$ monocular fisheye cameras with a nominal FOV of $185°$ synchronously triggered at $12.5\,\mathrm{Hz}$ provide $360°$ surround view. The prototype wide-angle rear stereo camera uses a $5\,\mathrm{cm}$ wide baseline and the same fisheye lenses as the monocular cameras. 12 ultrasonic sensors are used for close-range obstacle detection. The sensor setup is depicted in Fig. 4 and complemented by stock odometers. More details can be found in [1].

III. MULTI-CAMERA CALIBRATION

Map creation, visual localisation, and obstacle detection all rely on image data provided by the four monocular fisheye cameras mounted on the car. Consequently, these cameras need to be calibrated both intrinsically and extrinsically wrt. the odometry frame of the car. Projection of a three-dimensional (3D) landmark to an image point requires the following calibration parameters in addition to the vehicle pose: the intrinsic parameters for each camera and the transformation between each camera and the vehicle. Imprecise calibration parameters lead to inaccurate (back-)projections of landmarks, yielding suboptimal results for multi-camera-based algorithms such as motion estimation, pose estimation, and dense reconstruction. We developed the "CamOdoCal" software to calibrate such a novel fisheye multi-camera setup and made this software publicly available [2].

Environmental changes, wear and tear, and vibrations cause calibration parameters to slowly change, calling for frequent re-calibration. To avoid operator fatigue, the V-Charge project developed unsupervised, highly accurate calibration methods for the surround view camera system. The calibration method makes use of natural features in the environment to minimise infrastructure setup costs. In [3], we introduced a self-calibration method which leverages multi-camera SLAM to simultaneously build a map of the environment and estimate the calibration parameters.

A typical calibration takes several hours. To minimise turnaround time, we developed an infrastructure-based calibration method [4] that uses the self-calibration method as a one-time bootstrapping step for generating a map of
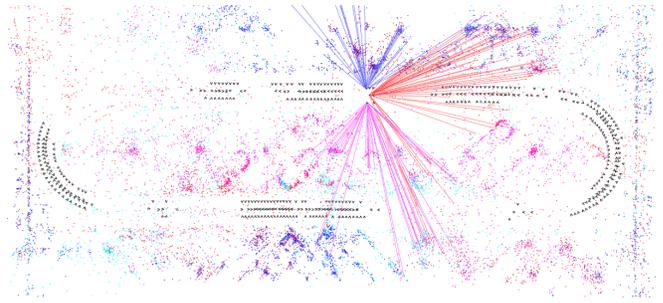

Fig. 2: Multi-colored points represent already observed 3D landmarks from the map, colored according to the camera they were first observed in. Small black triangles correspond to camera poses estimated by image-based localisation. Lines connect currently observed landmarks to the current camera poses.

a calibration area. Image-based localisation (see Fig. 2)) is used to estimate camera poses which are then used to estimate the calibration parameters. This infrastructure-based calibration enables calibration within minutes rather than hours. Both calibration methods make use of natural landmarks and do not require artificial calibration targets such as checkerboards or markers (except when computing the camera intrinsics for the first time). Further details can be found in [5].

IV. OFFLINE MAPPING

To enable operation in GPS-denied indoor parking garages, the V-Charge vehicle uses visual localisation with respect to a map containing visual features. We build the map offline following the standard structure from motion (SfM) framework with images collected from our multi-camera system while driving through the areas to be mapped.
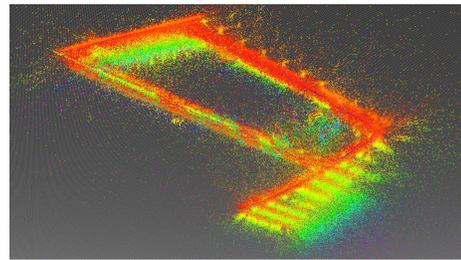

Fig. 3: Example of a loop-closed offline map (3D visual landmarks colored by height) of our testing site at Mobile Life Campus, Wolfsburg, Germany.

First, we perform motion estimation for all acquired images. Since the multi-camera setup results in a non-perspective camera (light rays from individual cameras do not intersect in a single centre of projection), motion estimation with epipolar geometry [6] becomes impossible. Instead, we model our multi-camera system as a generalised camera where all light rays are defined with respect to a common coordinate frame. This leads to the generalised epipolar geometry for motion estimation [7]. A naive implementation of this approach is computationally too expensive for practical use. We leveraged the fact that our multi-camera system is mounted rigidly onto a car to constrain the motion, thus greatly reducing the computational complexity [8], [9]. The

Fisheye Mono Cameras
FoV: 185°

Rear Stereo Camera
Range: 5 m / FoV: 120°

Front Stereo Camera
Range: 12,5 - 25 m / FoV: 45
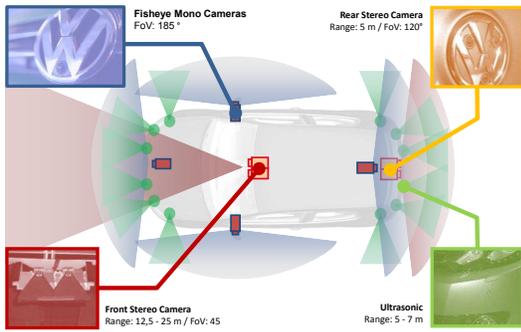
Ultrasonic
Range: 5 - 7 m

Fig. 4: The V-Charge sensor setup consists of two stereo cameras, twelve ultrasonic sensors and four monocular cameras.

3D map points are obtained from triangulation of feature correspondences after estimating all the camera poses. Each 3D point is additionally associated with descriptors from all images that observed this point for visual localisation.

To get a globally consistent map, a visual place recogniser based on the bag-of-words approach was implemented to identify loop-closures. We proposed a robust optimisation technique [10] to minimise loop-closure errors against wrong loop-closures identified by the visual place recogniser. Fig. 3 shows an exemplary offline map created using this pipeline.

V. Perception

Perceiving obstacles in the vehicle's environment is essential for safe navigation. The V-Charge sensor setup depicted in Fig. 4 is designed to give complete and redundant coverage of the vehicle's surrounding. Stereo cameras and ultrasonic sensors provide obstacle information to the front and the rear, yet this information is often not enough. For additional coverage of the vehicle sides, a SfM pipeline for the monocular cameras was developed. All sensor information is fused in an occupancy grid map (OGM).

A. Motion Stereo/Structure from Motion

When driving along a row of parked cars searching for an empty parking space, it is difficult to see inside a free parking space with only front/rear facing cameras. Likewise, while driving in narrow passages, obstacle information towards the side is required. Our pipeline outputs static obstacle information using the monocular fish eye cameras and wheel odometry [11]. The pipeline first computes a depth map using stereo matching on several consecutive images from a single camera. In a second step, the depth map is projected to the two-dimensional (2D) plane and obstacles above ground plane are extracted.

For stereo matching we used the well known approach of plane sweeping [12], [13]. To have a baseline for stereo matching from the monocular cameras we use several images captured in sequence while moving (only static obstacles will be reconstructed). The camera poses are computed from the wheel odometry and the extrinsic calibration of the fish eye cameras. To fully benefit from the wide FOV of the fish eye cameras, we avoid prior rectification [14]. This allows us to compute depth maps as shown in Fig. 5 in real-time using a
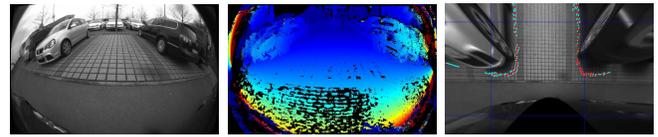

Fig. 5: From left to right: Input image, computed depth map (red close, blue far), extracted obstacles (red: obstacle position, cyan: uncertainty interval)

high end GPU. We made our plane sweeping implementation publicly available[1].

With this pipeline, accurate measurements of parking space sizes and distances to obstacles are achieved, while running in real-time on multiple cameras concurrently [11]. The depth maps computed can also be used to generate dense height maps of the environment [1], [15].

B. Occupancy Grid Map Fusion

The sensors we employ each have different capabilities and limitations, like limited FOVs and precision and varying false-positive/false-negative rates. With the goal of combining the sensors' strengths while eliminating their weaknesses, our OGM fuses the individual sensor outputs on a metric lattice in the local surroundings of the vehicle at ten samples per second. In the following, we evaluate the *completeness* and *precision* of our grid fusion.

The OGM's *completeness* is evaluated on several scenes containing elements that are typically difficult for the sensors in our suite. Obstacles with repetitive texture, such as wooden and wire-link fences (which are notoriously difficult for the stereo and SfM based obstacle sensors) as well as objects with high ultrasonic absorbency and small obstacles (such as traffic cones and poles) are considered. Based on snapshots of the fused OGM, we determine whether objects are missing and whether safe driving would be possible (see Tab. I).

The overall system is able to detect most types of obstacles. Exceptions are the arms of the parking lot gates, which are not perceived by a single sensor; this requires the parking lot server to communicate the opening/closing state of the gate arms to the vehicle. While the weakly/repetitively textured obstacles do not pose a big problem to the camera sensors, the stereo camera is not able to reliably detect the small traffic cones, due to its limited vertical FOV. The wire link fence is only perceived reliably by the ultrasonic sensors. The contribution of SfM to detect parking spaces at the side of the vehicle at a higher distance than the ultrasonic sensors while driving is essential.

The *precision* of the OGM is evaluated via the procedure shown in Fig. 6 and results are given in Tab. II. Cars parked on both sides of the parking space restrict its width.

The accuracy for both stereo sensors lies in the range of 11-21 centimetres. The ultrasonic sensors are not accurate in the early stages of the parking manoeuvre, due to their limited range. Finally, SfM has the most indirect measurement principle; it is therefore the sensor most susceptible to calibration-, synchronisation- and pose-estimation errors.

[1] http://cvg.ethz.ch/research/planeSweepLib/

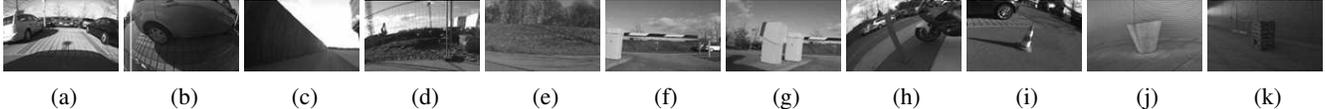| Scenario	Figure	Stereo Front	Stereo Rear	Dense SfM	Ultrasonic	Fused Grid
Parking Spots While Driving	a	-	-	+	-	+
Parking Spots While Parking	b	-	-	+	+	+
Wooden Fence	c	+	+	+	+	+
Wire Fence	d	-	-	-	+	+
Earth Wall	e	+	+	+	-	+
Horizontal Gate Arm	f	-	-	-	-	-
Gate Housing	g	+	+	+	+	+
Poles	h	+	+	+	-	+
Cones	i	-	-	+	+	+
Ashtray	j	+	+	+	+	+
Foam Block	k	+	+	+	+	+

(a)　(b)　(c)　(d)　(e)　(f)　(g)　(h)　(i)　(j)　(k)

TABLE I: Qualitative evaluation of the completeness of the occupancy grid map with regard to various obstacle types. A + indicates that we asses the sensor to be able to perceive this obstacle type; - indicates that it is not.

Mean Error (m) (Std. Dev. (m)) / Phase	Stereo Front	Stereo Rear	Dense SfM	Ultrasonic	Fused Grid
Phase A	-0.1 (0.21)		-0.5 (0.10)		0.01 (0.20)
Phase B	-0.1 (0.21)		0.09 (0.05)	0.66 (0.19)	0.12 (0.05)
Phase C	-0.1 (0.21)	-0.13 (0.07)	0.08 (0.06)	0.24 (0.09)	0.21 (0.09)
Phase D	-0.1 (0.21)	-0.12 (0.07)	0.05 (0.15)	0.07 (0.03)	0.08 (0.03)

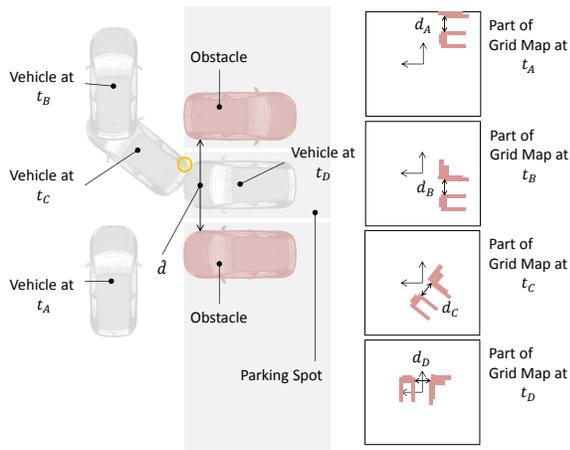TABLE II: Mean error and standard deviation for the parking space width estimated in the occupancy grid.

Fig. 6: Evaluation procedure for parking space width in the occupancy grid map. During various phases of a parking manoeuvre, the parking space width in the grid map is compared with manually measured ground truth data.

The V-Charge sensor setup does an excellent job of both perceiving the different obstacle classes encountered on a parking lot, as well as precisely measuring their metric extents. We have combined a set of precise sensors with limited FOV or range (stereo and ultrasonic) with the $360°$ surround view SfM. This gives us good metric performance in estimating the space between obstacles, and the extent of parking spaces. At the same time, the sensor setup completely covers the vehicle surroundings, enabling safe navigation on the parking lot.

VI. SEMANTIC MAPPING

The offline map from Sec. IV builds the metric layer of the map stack. It is further enhanced with a semantic layer (see Fig. 7) that comprises three distinct parts to support the planning modules: a road graph, detailing the positions of lanes, way directions and intersections; the locations of the parking spaces; and a speed profile at which it is sensible for the vehicle to move through the car park.
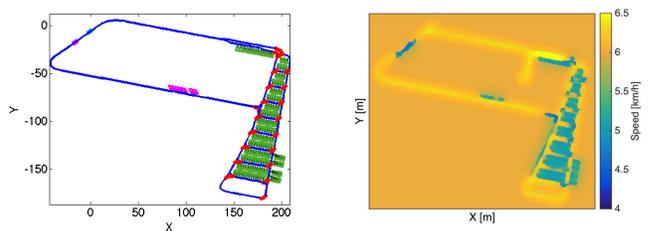

Fig. 7: Roadgraph (left) and speed map (right) created from the offline map depicted in Fig. 3. Numerous lanes, parking spaces (green rectangles), charging bays (magenta rectangles) and intersections (red curves) were automatically identified.

A. The Road Graph

The vehicle poses from the metric layer constitute the layout of the lanes in the car park, however trajectories will overlap imperfectly as the car passes over certain lanes more than once (Fig. 9a). To extract the underlying skeleton, a graph over the poses is imposed. Each pose forms a node, and edges are inserted between both subsequent and nearby nodes (Fig. 9a). By identifying cliques in the graph and iteratively replacing each clique (starting with the largest)

with its centre-point (or the pose nearest to its centre-point, to avoid drift over many iterations), clusters of poses are replaced by the underlying structure (Fig. 9b) yielding a clean graph (Fig. 9c). More details can be found in [16].

B. The Parking Labels

The vehicle poses in the metric map, together with fisheye images from the four calibrated cameras form the input of a registration process which outputs a synthetic overhead image as shown in Fig. 8. Labels such as parking spaces have greatly differing visual characteristics for different parking lots, thus requiring differently trained classifiers. To provide a transferable and low-maintenance mapping process, introspective classification [17], [18] is used to reason about uncertainty of labels, asking for human expertise where the software-based classifier is too uncertain. This enables the classifier to improve over time requiring less and less feedback from the human expert. Introspective classifiers make active learning more efficient [17], [18], enabling us to label all parking spaces within a few iterations (see Fig. 8).
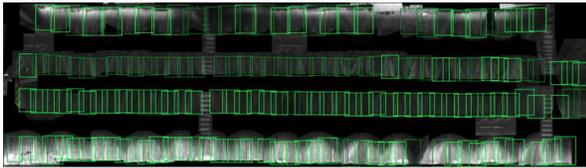

Fig. 8: Parking space classifications in the synthetic overhead image.

C. The Speed Map

To increase safety for pedestrians and other road users, lower speeds in crowded areas or areas with poor visibility are recommended. We encode this speed information in a 2D grid. First, we create a probabilistic graphical model using a prior based on the positions of lanes and parking spaces. Locations of observed pedestrians mark positive observations in a Markov random field imposed over the prior map. A smoothing procedure over the map allows us to update the prior to include observed pedestrians and slow down the vehicle as it approaches areas with expected high activity (see Fig. 7, right). See [16] for the details of this process.
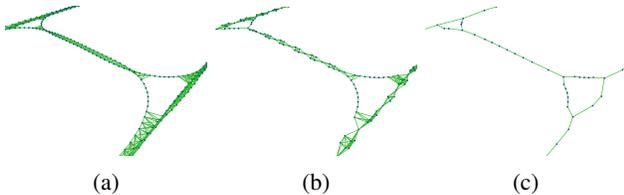

(a)   (b)   (c)

Fig. 9: a) Densely connected graph of vehicle poses. Edges are created between both nearby and subsequent poses. b) The graph after one round of pruning. The largest cliques are replaced by the nearest pose to their centre-point. c) The finished graph after the iterative pruning procedure.

VII. Communication and Scheduling

A typical V-Charge mission starts with the vehicle's drop-off by the customer in a designated drop-off area and the required clearance confirmation via his or her smartphone.

For this purpose the project developed an Android application (see Fig. 10) involving a user design study to optimise usability. This includes providing "one-glance" visual feedback of the valet parking status (pick-up, drop-off, etc.) and avoiding user feedback for non-safety-relevant actions.
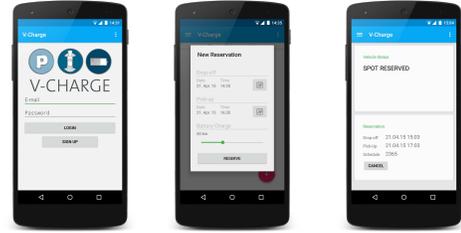

Fig. 10: Android user interface

Secure data transmission concepts [19] between vehicle(s) and a remote parking management server [20] accessed via local wireless links or mobile data connections were developed. The server is responsible for park/charge resource management, monitoring, and distribution of mission information to the vehicles.

Handling of large amounts of parking spaces and charging stations by the parking management server were demonstrated in simulation [20]. In contrast to regular (human) or mixed-mode (human + robots) parking scenarios, there is a significant potential for space (and, consequently, cost) savings in fully-automated valet parking. A coordinated parking management [21] based on Vehicle-to-Infrastructure communication may increase parking density by allowing vehicles to block other cars. This way, a parking lot needs fewer roads for more parking spaces and the maximum parking density increases, albeit at the expense of *shunting operations* of parked cars. We showed that very good pick-up times of about $1\,\mathrm{min}$ are possible with very little overhead in terms of shunting distance and time, while significantly improving parking density as compared to conventional parking lots.

VIII. Visual Localisation

After the user's clearance, the vehicle localises itself in the offline map at the drop-off location. Localisation uses solely the monocular cameras and natural landmarks to strive for cost efficiency for both vehicle owners and parking lot operators. Its basis constitutes the offline map (see Sec. IV) containing sparse landmarks together with feature descriptors in a 3D geometric space. To localise online, 3D landmarks around the current vehicle pose are queried from the offline map and projected into the camera images. 2D features are extracted from the current images and feature matching is performed using both a threshold on distance in image-space and on descriptor distance. Given these correspondences, the initial pose guess is refined through nonlinear least-squares optimisation. Outliers are managed through via a robust cost function through a Huber M-Estimator [22]. In between consecutive optimisation cycles, the last solution is propagated using wheel odometry. Our approach to *Multi-Session Mapping* addresses the challenge of robust visual

localisation in presence of lighting-, weather-, structural and seasonal changes. As a single dataset map will quickly lead to reduced localisation performance, we augment this map with additional datasets from different times and days. The core procedure for this is identical to the localisation algorithm above. Pose estimates are used to incorporate the trajectory of a new dataset in a geometrically consistent way. Correspondences from the matching step allow offline bundle adjustment, while tracking features along the new dataset allows to incorporate landmarks from a new environmental condition. This procedure is repeated until the map contains enough datasets to guarantee successful localisation in all required conditions. To fight the unbounded increase of map size, multiple summarisation strategies to prune redundant landmarks have been investigated, including well-known strategies proposed from literature [23], [24]. A combined score of the number of datasets a landmark was observed in and the total number of observations gave best results.

Long-term experiments in challenging outdoor scenarios proved reliable localisation across different conditions with an accuracy better than $10\,\mathrm{cm}$ (groundtruth by DGPS) [25].

IX. OBJECT DETECTION AND CLASSIFICATION

While navigating along the route, the vehicle perceives its surrounding and separates static obstacles from other mobile traffic participants to enable smooth navigation in mixed traffic. We have developed two dynamic object detection and classification pipelines, namely *Stereo Obstacle Classification* and *Obstacle Detection by Classification with* $360°$ *Tracking*. In *Stereo Obstacle Classification* we make use of precise object hypotheses obtained from a clustering step on the stereo output. In the latter one, we solely rely on the monocular surround view system to detect and track other traffic participants $360°$ around the vehicle.

A. Stereo Obstacle Classification

To classify 3D stereo object hypotheses we exploit the large FOV of the front fisheye camera. Hypotheses are first passed through a coarse pre-filter checking for e.g. reasonable geometric extents. Each surviving candidate is then projected into the unwarped fisheye image and a region of interest (ROI) is generated. A state-of-the-art feature extraction step based on aggregate channel features (ACF) is performed on the ROIs, followed by a soft-cascade classifier. We used one classifier for vehicles and two classifiers for near/far pedestrians. More details can be found in [26].

The *accuracy* is evaluated on publicly available automotive datasets using the precision/recall and miss-rate metric (see Fig. 11). It constitutes a lower bound on the system performance as neither stereo clustering nor the downstream object tracking are separately evaluated and is in line with the state of the art.

B. Obstacle Detection by Classification with $360°$ Tracking

The *Stereo Obstacle Classification* can only operate on objects inside the stereo camera's FOV. To exploit the potential of the monocular surround view system, we have developed a system that is able to detect and track pedestrians
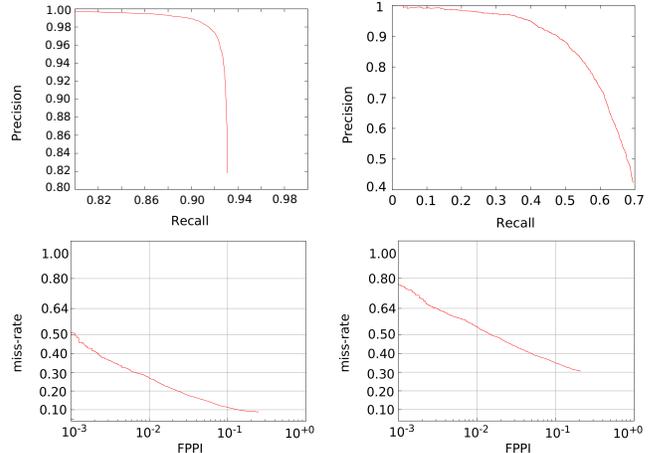

Fig. 11: Evaluation of vehicle (left) and pedestrian classifiers.

Classifier	# Samples	Model	Correct Detection
Pedestrian 32x64	130017	Pinhole Cylindrical	96.7127% 96.7150%
Pedestrian 48x96	67158	Pinhole Cylindrical	98.1148% 98.1044%
Vehicle 38x38	58120	Pinhole Cylindrical	82.4260% 82.3623%

TABLE III: Comparison between pinhole and cylindrical model.

and vehicles all around the car [27]. A Soft-Cascade+ACF classifier is run on each camera image to detect vehicles and pedestrians. Here we use the cylindrical camera model described in [27] to minimize distortions outside the image center while retaining the large FOV. For multi-camera tracking first an inter-camera association algorithm is applied, followed by an unscented Kalman filter to track obstacles.

As no dataset with annotated fisheye images is publicly available, we study the impact of post-warped images on the classifier's performance. The classifier is compared on pinhole images warped with the cylindrical model and pinhole images from public datasets. In Tab. III a comparison of correct detections between the two models is given. Three different classifiers have been evaluated: two pedestrian classifiers (32x64 and 48x96 pixels) and one vehicle classifier (38x38 pixels). The cylindrical model has a negligible impact on the classifier performance, hence the results presented in the previous paragraph can be still considered valid. Exploiting the overlapping FOV between cameras improves detection performance (see Fig. 12), especially for crossing or temporarily occluded objects.

X. MOTION PLANNING

The V-Charge motion planning stack is comprised of three modules, coordinated through the *mission executive* module responsible for high-level task assignment. The *Global Planning* module computes topological plans via a graph-search on the static road network from Sec. VI-A from e.g. the drop-off zone to a designated parking space or vice versa. The *Local Planning* module is designed to provide system-compliant online motion planning along a sequence of lanes, considering advanced vehicle dynamics and dynamic objects

Fig. 12: The multiple viewpoints available in the surround view tracking of dynamic objects allow to overcome occlusions (the far pedestrian is visible only in the left image).

in the environment. The *Parking Planner* is especially suited for planning potentially more complex manoeuvres in and out of narrow parking spaces.

A. The Local Planner

A fast local motion planner [26], [28] computes motion commands, mitigating static obstacles and pedestrians as well as vehicles. It operates in a sampling-based manner, generating numerous system-compliant candidate motions along a reference path in a tree-like fashion. This trajectory roll-out scheme is widely used for automotive applications [29], [30], however differs from these related works in the way candidate motion primitives are constructed. Instead of using geometric primitives or parametrized functions that might not conform with the non-holonomic system model of a car and have to be pruned at a later stage, the candidate motions are constructed via a forward simulation of a detailed vehicle model. In conjunction with a simulated controller regulating the system towards samples of a lane-aligned manifold, these candidate motions are inherently drivable and challenging system characteristics such as dead times and actuator limits can be modelled easily. The forward simulation time is still low compared to the one of collision detection.

Other traffic participants (see Sec. IX) are predicted over a finite planning horizon of $10\,\text{s}$. For vehicles, road network information from the semantic layer is used to stabilise these mid-term predictions. Higher-level behaviour modifications were implemented to support smooth convoying/platooning and handling of intersections. Candidate ego motions are tested for collisions against other agents' predictions with a fast, time-aware bounding volume hierarchy data structure in workspace-time space [31]. To account for the uncertainty of the movement predictions of other decision making agents over time, predicted collisions with those are incorporated as an exponentially decaying cost term in the local planner's cost function. This reduces erratic behaviour of the ego vehicle to potentially irrelevant collisions in the far future and adds resilience to noise in the object detection chain.

The reactive planning approach shows excellent navigation performance in static environments including narrow sections. Navigation in dynamic scenarios shows promising results for convoying/platooning and avoidance of pedestrians. Statistics for the local motion planner were automatically generated from real-world test data, recorded over the course of 6:47 hours ($24.63\,\text{km}$) of automated operation. In $99.3\,\%$ of the time the planner operated in nominal mode; in $0.2\,\%$, no valid collision-free candidate motion could be found. Fig. 13 depicts two situations mostly responsible for the

latter case: pedestrians moving in very close proximity to the vehicle (left); and passing through narrow gates with only a few tens of centimetres margin on each side (right). In the first situation the purely reactive planning approach adopted is over-conservative. A cooperative one would remedy this shortcoming and is among our current research goals.
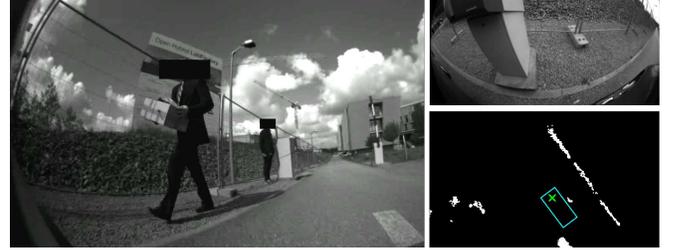

Fig. 13: Difficult situations for the local motion planner. Pedestrians moving in close proximity to the car (left) and passing through narrow gates (right).

B. The Parking Planner

In order to obtain "natural" parking manoeuvres, i.e. manoeuvres that contain only a minimal number of cusps and use little space, a three-stage planning cascade with increasing complexity is used (see Tab. IV). The first two stages are comprised of deterministic approaches as humans tend to pull in with comparable behaviour, i.e. with either a single move or with two changes of direction. A hybrid A* planner [32] is applied in the third stage providing the highest flexibility. The parking planner proceeds through these three stages until a collision-free manoeuvre is found. Its robustness was confirmed in 267 documented trials (backward and forward) performed over a time-period of about two months with an average success rate of $97\,\%$.
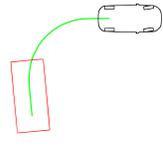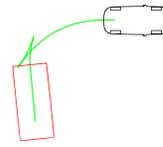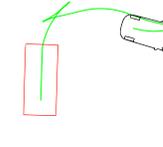
Stage 1: *Reeds & Shepp* [33] *planner*	Stage 2: *Deterministic 3-motion planner*	Stage 3: *Hybrid A* planner*
Yields one single move with two straight lines that are connected by a circular arc. No change of direction allowed.	Yields moves consisting of two cusps, i.e. two straight lines are connected by three circles. Circles are placed with a deterministic strategy.	A modified state-lattice search that provides the highest flexibility. An arbitrary number of cusps is allowed in the moves.

TABLE IV: The three planning stages applied in the parking planner.

Parking onto a charging plate requires an *accuracy* of $\pm 15\,\text{cm}$ in longitudinal direction and $\pm 10\,\text{cm}$ in lateral direction. We evaluate the mean absolute deviation of the parking planner's final pose in extensive real-world experiments wrt. these requirements. This deviation is $3.3\,\text{cm} \pm 2.1\,\text{cm}$ in longitudinal and $2.1\,\text{cm} \pm 1.0\,\text{cm}$ in lateral direction for forward parking manoeuvres. For backward parking manoeuvres, due to more complex manoeuvres with several cusps involved, we

obtained $6.8\,\mathrm{cm}\pm6.5\,\mathrm{cm}$ in longitudinal and $6.4\,\mathrm{cm}\pm4.6\,\mathrm{cm}$ in lateral direction.

The required *parking spot width* for the parking manoeuvre is evaluated in simulation to remove disruptive factors like perception- and actuation inaccuracies. In Tab. V the success rate and average number of parking moves is given for different simulated widths. Note that the non-deterministic outcome of the parking manoeuvre stems from varying initial poses used. The success rate decreases the tighter the parking space is while the average number of moves increases. Adding the uncertainty of $0.12\,\mathrm{m}\pm0.05\,\mathrm{m}$ of the perception system after passing the parking space (phase B, see Tab. II), we arrive at a required clearance of approximately $0.6\,\mathrm{m}$ for a success rate of $100\,\%$. These results are in line with our full system field tests.

parking space width	clearance	# parking moves	success rate
2.64 m	0.44 m	1.10	100 %
2.44 m	0.24 m	1.50	90 %
2.36 m	0.16 m	1.53	80 %
2.28 m	0.08 m	2.25	75 %

TABLE V: Parking success rate and average number of parking moves for an ego vehicle width of $2.2\,\mathrm{m}$.

XI. CONCLUSION

The V-Charge project showed reliable automated driving in designated areas using close-to-market sensors only. The vehicle was successfully demonstrated at various indoor and outdoor parking lots and garages. We see two main research areas that should be further investigated in future projects: lifelong mapping and calibration concepts to minimise map- and vehicle maintenance work; and interaction-aware motion planning and object prediction approaches potentially involving machine learning techniques to cope with other traffic participants in close proximity.

REFERENCES

[1] P. Furgale *et al.*, "Toward Automated Driving in Cities using Close-to-Market Sensors, an Overview of the V-Charge Project," in *IEEE Intelligent Vehicles Symposium (IV)*, pp. 809–816, 23–26 June 2013.

[2] L. Heng *et al.*, "Camodocal." http://www.github.com/hengli/camodocal.

[3] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1793–1800, Nov 2013.

[4] L. Heng, M. Bürki, G. H. Lee, P. Furgale, R. Siegwart, and M. Pollefeys, "Infrastructure-based calibration of a multi-camera rig," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4912–4919, May 2014.

[5] L. Heng, P. Furgale, and M. Pollefeys, "Leveraging image-based localization for infrastructure-based calibration of a multi-camera rig," *Journal of Field Robotics*, vol. 32, no. 5, pp. 775–802, 2015.

[6] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd ed., 2004.

[7] R. Pless, "Using many cameras as one," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[8] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Motion estimation for a self-driving car with a generalized camera," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[9] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Structureless pose-graph loop-closure with a multi-camera system on a self-driving car," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[10] G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Robust pose-graph loop-closures with expectation-maximization," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2013.

[11] C. Häne, T. Sattler, and M. Pollefeys, "Obstacle detection for self-driving cars using only monocular cameras and wheel odometry," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2015.

[12] R. Yang and M. Pollefeys, "Multi-resolution real-time stereo on commodity graphics hardware," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2003.

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

[14] C. Häne, L. Heng, G. H. Lee, A. Sizov, and M. Pollefeys, "Real-time direct dense matching on fisheye images using plane-sweeping stereo," in *Int. Conf. on 3D Vision (3DV)*, 2014.

[15] C. Häne, C. Zach, J. Lim, A. Ranganathan, and M. Pollefeys, "Stereo depth map fusion for robot navigation," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2011.

[16] H. Grimmett, M. Bürki, L. Paz, P. Piniés, P. Furgale, I. Posner, and P. Newman, "Integrating Metric and Semantic Maps for Vision-Only Automated Parking," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, May 2015.

[17] H. Grimmett, R. Paul, R. Triebel, and I. Posner, "Knowing When We Don't Know: Introspective Classification for Mission-Critical Decision Making," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2013.

[18] R. Triebel, H. Grimmett, R. Paul, and I. Posner, "Driven Learning for Driving: How Introspection Improves Semantic Mapping," in *Int. Symposium on Robotics Research (ISRR)*, 2013.

[19] J. Timpner, D. Schürmann, and L. Wolf, "Secure Smartphone-based Registration and Key Deployment for Vehicle-to-Cloud Communications," in *Proceedings of the ACM Workshop on Security, Privacy and Dependability for Cyber Vehicles*, pp. 31–36, Nov. 2013.

[20] J. Timpner and L. Wolf, "Design and Evaluation of Charging Station Scheduling Strategies for Electric Vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 579–588, 2014.

[21] J. Timpner, S. Friedrichs, J. van Balen, and L. Wolf, "k-Stacks: High-Density Valet Parking for Automated Vehicles," in *IEEE Intelligent Vehicles Symposium (IV)*, pp. 895–900, IEEE, June 2015.

[22] P. J. Huber, "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, vol. 35, pp. 73–101, Mar. 1964.

[23] K. Konolige and J. Bowman, "Towards lifelong visual maps," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1156–1163, Oct 2009.

[24] F. Dayoub *et al.*, "Long-term experiments with an adaptive spherical view representation for navigation in changing environments," *Robotics and Autonomous Systems*, vol. 59, pp. 285–295, may 2011.

[25] P. Mühlfellner, M. Bürki, M. Bosse, W. Derendarz, R. Philippsen, and P. Furgale, "Summary maps for lifelong visual localization," *Journal of Field Robotics*, 2015.

[26] U. Schwesinger, P. Versari, A. Broggi, and R. Siegwart, "Vision-only fully automated driving in dynamic mixed-traffic scenarios," *it - Information Technology*, vol. 57, no. 4, pp. 231–242, 2015.

[27] M. Bertozzi, L. Castangia, S. Cattani, A. Prioletti, and P. Versari, "360° detection and tracking algorithm of both pedestrian and vehicle using fisheye images," in *IEEE Intelligent Vehicles Symposium (IV)*, pp. 132–137, IEEE, 2015.

[28] U. Schwesinger, M. Rufli, P. Furgale, and R. Siegwart, "A Sampling-Based Partial Motion Planning Framework for System-Compliant Navigation along a Reference Path," in *IEEE Intelligent Vehicles Symposium (IV)*, pp. 391–396, June 2013.

[29] M. Werling *et al.*, "Optimal trajectories for time-critical street scenarios using discretized terminal manifolds," *The Int. Journal of Robotics Research*, vol. 31, pp. 346–359, Dec. 2011.

[30] F. von Hundelshausen *et al.*, "Driving with tentacles: Integral structures for sensing and motion," *Journal of Field Robotics*, vol. 25, no. 9, pp. 640–673, 2008.

[31] U. Schwesinger, R. Siegwart, and P. Furgale, "Fast Collision Detection Through Bounding Volume Hierarchies in Workspace-Time Space for Sampling-Based Motion Planners," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 63–68, May 2015.

[32] D. Dolgov *et al.*, "Practical search techniques in path planning for autonomous driving," *1st Int. Symp. on Search Techniques in Artificial Intelligence and Robotics*, June 2008.

[33] J. Reeds and L. Shepp, "Optimal paths for a car that goes both forwards and backwards," *Pacific Journal of Mathematics*, vol. 145, no. 2, pp. 367–393, 1990.