3D Sensing-based Autonomous Parking

カメラによる空間検知技術を用いた自動駐車システム

Duong-Van Nguyen ドゥオンヴァン グエン
Petros Kapsalas ペトロス カプサラス
Shirish Khadilkar シリッシュ カディカール
Makoto Mochizuki 望月　誠

Abstract

This paper introduces the full 3D sensing camera-based autonomous parking system, which comprises a fast dense environmental sensing capability and an innovative fusion solution to accurately detect available parking slots as well as optimally park the ego-car to the planned target position, by only using surrounding-view-system cameras. This is also the one stroke autonomous parking system where the ego-vehicle can detect and park autonomously into difficult spots by one single maneuvering. With 500 test-drives containing varieties of challenging use-cases, the proposed parking system achieves a success rate of 94.1 % compared to 48.3 % of the selective advanced parking systems, meanwhile still being embedded friendly with 10 giga operations per second (GOPS). Also, the high-density map given by the system can be potentially used for many advanced functions in higher automation applications including autonomous driving.

1. Introduction

Recently autonomous driving has been a very hot topic in automotive, which attracts more than hundred billions of investment globally. However, due to the fact that automotive is strongly connected to safety and regulation, there are many problems still to be solved. For now, autonomous parking (AP), as part of autonomous driving technology, has become the utmost important and realistic intermediate step to pursuit, which is present in most of automotive R&D discussion as well as RFIs/RFQs from OEMs globally. Available automated parking system (APS) solutions [1][4][5][6][10][11][16][17][19] include the SONAR -based and Camera-based ones separately.

The SONAR-based has its own nature limitation of short range, sensitivity to temperature, wind, object surface, as well as limited depth information, etc. Therefore, SONAR-base system will need many strokes to let the car parking in. Also small thin obstacles and moving objects are not properly detected or lately detected, which limits the applicability and reliability of such system in a daily usage. Meanwhile the current camera-based solutions only utilize 2D image processing approach like park-marking detection, which has no 3D information to guide the vehicle navigation and thus being unsafe and sensitive to illumination changes.

This paper will introduce a novel and low-cost approach to tackle the autonomous parking problem. Indeed, the approach will only use the available camera-system from the car, for example surrounding view system (SVS) cameras, and thus no additional sensor cost. Our proposed solution scheme includes 3D environmental sensing by cameras only, Fusion Map, and Application Layers such as automated parking, garage parking, home-zone parking, and autonomous robot valet parking like in airport, park, or meta-office area.

2. Dense SLAM(1)

First, to acquire 3D information, this paper presents the novel fast dense 3D environmental reconstruction, which enables a semi-active sensing system from a passive sensing sensor like camera. Unlike a common 3D reconstruction approaches [12][13][20], we propose a direct matching between two consecutive images purely based on a mathematic model of geometric transformation. An initialization of the geometric transformation is taken by the vehicle-CAN signal (wheel ticks, steering wheel, etc.) whereby its Lie-group representation (6 degrees of freedom) is optimized iteratively [2][9].

Key equations for optimizing the transformation of camera's pose through direct image mapping are illustrated in Eq. (1), Eq. (2).

$$E(\xi) = \sum_{i=1}^{N} r_i(\xi)^2 \qquad (1)$$

Where ξ is the transformation of the camera pose, and r is the photometric difference between consecutive frames

$$r_i(\xi) = I_{ref}(p_i) - I\left(\Omega\left(p_i, D_{ref}(p_i), \xi\right)\right) \qquad (2)$$

I_{ref} is the reference image, D_{ref} is the depth of the corresponding world point, transformation Ω projects the world point into the current image. Step transformation update using Levenberg Marquardt (LM) is expressed in Eq. (3)

$$\Delta\xi_n = -\left[J^T J + \lambda diag\left(J^T J\right)\right]^{-1} J^T r(\xi_n) \qquad (3)$$

(1) Simultaneous Localization and Mapping

Where J is Jacobian of r(ξ) over ξ, and λ is a damping factor. In order to be robust to outliers arising from occlusions or reflections, different weighting schemes should be used, resulting in an iteratively re-weighted least-squares problem, see more in [8].

From the works of Cremers.et.al. [7][8][9], the photometric constraint is facing various issues when dealing with highly distorted images from fisheye lenses. Meanwhile the purely scale-change issue at the "Focus of Expansion" area is challenging for a monocular pinhole camera. In our findings, the instability of a direct approach comes mainly from the deformation effect of rectifying images and the bias in optimization process due to improper inputs. First, for simplifying the problem, most of existing direct-approaches rectify input-images as a pre-processing step, which is not recommended in case of highly distorted images. For example, applying [9] with our wide-range camera (190°×180°), the degradation of the performance is more than 30 % compared to the report in [9] which used a pin-hole camera.

Second, camera-pose optimization is a probabilistic process, which relies on the belief of having more inliers than outliers in order to get it converged. For a monocular-pinhole camera, the field of view is narrow, and thus almost just having a scale change when the ego-vehicle moves straight forward. Such tiny changes between consecutive images causes ambiguous for the optimization, which was already mentioned in [8], and thus a back-end optimization was proposed in [9] to improve the scale issue, but still not fully solved.

Therefore, we propose to directly map raw images, and having an outlier removal scheme based on spatial property to select proper points for the optimization. This means that our solution builds a sophisticated geometric model for the lens, up to 11 or high order polynomials [3]. By doing so, we can directly compute the Jacobian of the residual in Eq. (1), and thus predicting its next movement instead of computing such Jacobian for every new image, as described in [9]. This makes our solution is 2xtimes faster and 3xtimes denser than previous arts [8][9] while the accuracy remains similar. The Block-diagram of Dense SLAM is illustrated in Fig. 1.

Eventually, 4000 to 6000 of 3D points per camera are reconstructed per 21 ms by ARM57(2) profiling. This is by far faster compared to the traditional SLAM approach which might take up to 300 ms by the same configuration and expectation [12][13][20].

(2) ARM is a registered trademark or trademark of Arm Ltd..

Using Velodyne HDL-64 as the golden standard sensor for 100 captured sequences with fisheye cameras, the accuracy of the system is slightly better than the state of the art with spatial error lesser than 5 % of object distance, see Fig. 2. In this evaluation, only the closest depth per ray-tracking is taken into account to avoid the complication of either multi-to-one or one-to-multi mapping between the Velodyne and camera systems.

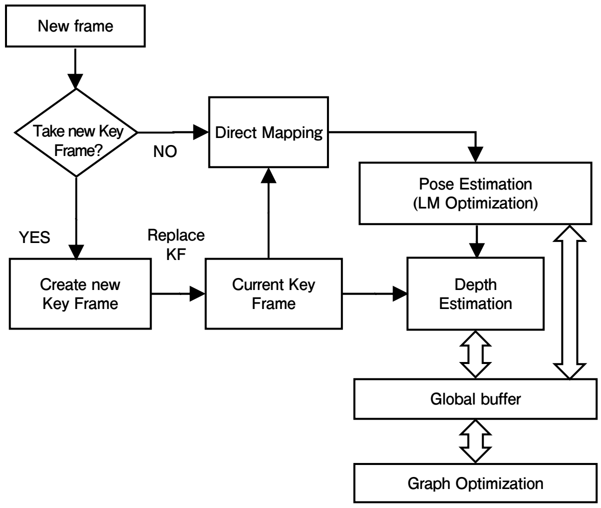


Fig. 1 Block Diagram of Dense SLAM

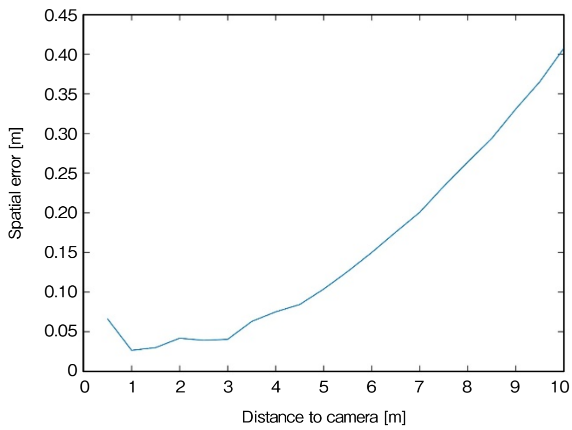


Fig. 2 Spatial error against object-distance by 1.3 M fisheye camera 190° × 180°, 1280 × 960 pixels

Since our approach directly maps consecutive images instead of feature matching, it is much more invariant against detection of homogeneous surface. Indeed, texture-less objects, if having a shading effect due to a reflection to the sunlight, will be well captured by our solution.

For example, a homogeneous white-sitting baby (35cm height) is well detected from a quite far distance (up to 8m, depending on camera resolution) as illustrated in Fig. 3. This

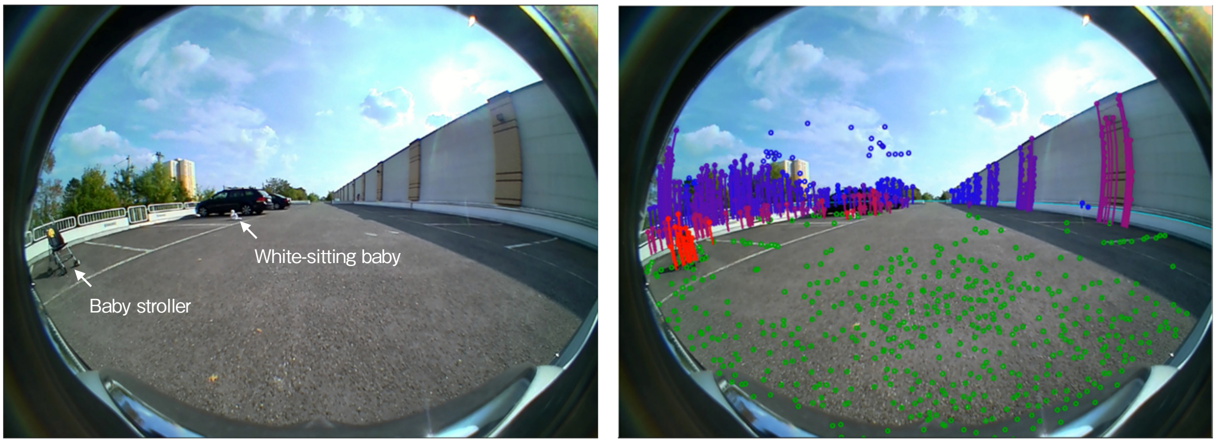


Fig. 3　Left: raw image; Right: reprojection image (red-near obstacles, green-ground, blue-far obstacles)

would be infeasible if using traditional feature-based approaches. Still, our approach is fundamentally based on “bright-ness constancy” assumption at 30 fps (frames/s) or 60 fps, and thus a degradation in performance is expected on cameras with heavy flickering effect or other “ghost” effects caused by ISP.

Putting the point cloud given by the dense-SLAM into an occupancy grid map (OGM), a relatively dense map is obtained with very detailed environmental information, which is comparable to a map obtained by Velodyne HDL-32.

Fig. 4 is an example of our system’s outputs for 3D reconstruction and OGM representation, which was captured in Frankfurt Motor Show (IAA) September 2019.

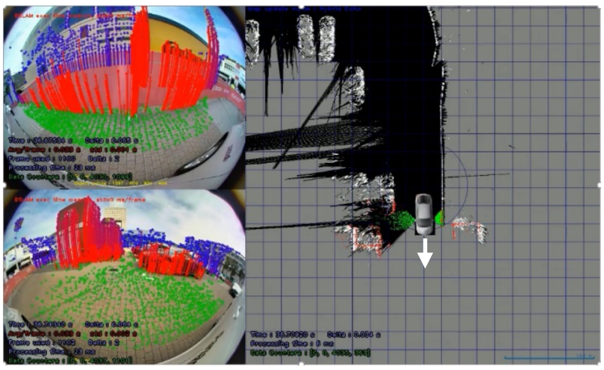


Fig. 4　Left: reprojection of point cloud on the perspective views of the left and right mirror camera; color notation includes red (near obstacle), green (ground), blue (far obstacles). Right: Occupancy Grid Map s.t. white (obstacle), black (free-space), grey (unknown).

3. Sensor Fusion Map

Updating an occupancy grid map from a given point cloud is well-known in robotics with the use of Lidar sensors [14][15][18]. However a direct applying of such techniques is not efficient and results in a noise map. This is due to the fact that the point cloud provided by the monocular camera-based system has much larger uncertainty or error for far-objects. Thus, there has to be a mechanism to improve the accuracy of objects’ locations when approaching closer. We, therefore, invent an innovative fusion solution on asynchronous map-updating, which leads to obtain a stable and highly accurate environmental map. This includes 1) active-free space fusion to update and correct obstacle points and clear free-space using ground points; 2) sensor free-space to handle obstacle decaying as well as overall cleaning the map over time. The Block-Diagram of Fusion Map is illustrated in Fig. 5. Whereby the fineness of the generated fusion map is comparable to the map given by a Velodyne HDL-32 which costs up to 40 000 USD; meanwhile our SVS system costs lesser than 60 USD.

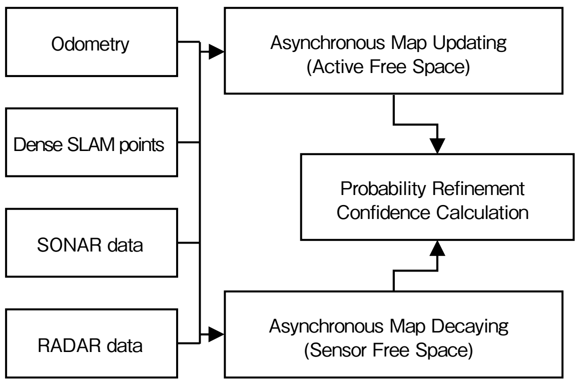


Fig. 5　Block Diagram of Sensor Fusion map

Additionally our sensor fusion map also allow a common logOdd mechanism [14][15] to add sonar detection information into the grid map if available. Concretely, fusing Cameras’ and sonars' outputs results in a relatively dense environmental map where detailed measures on spatial distances (to obstacles by their "red" polygon) and spaces (yellow quadrilateral to find empty slots) can be exploited, see Fig. 6.

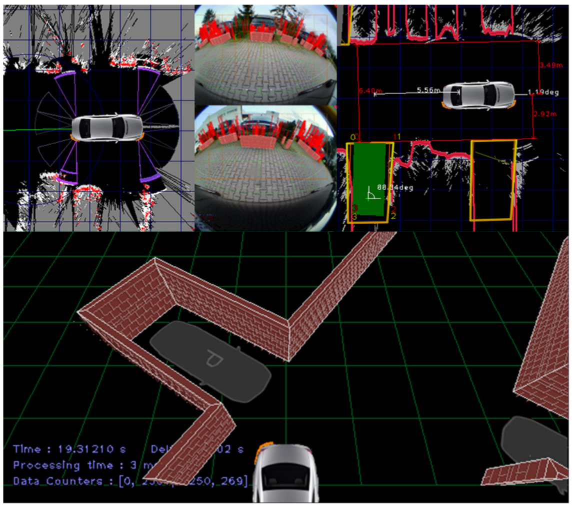


Fig. 6 Top-left: fusion map between Cameras' point cloud and Sonar (purple beam). Top-right:object boundaries are obtained in form of polygons based on the fusion map, and possible empty slots are analyzed by yellow quadrilaterals. Bottom:example of 3D visualization achieved for parking application using our detailed

Apparently, one might observe that the map looks very similar with or without Sonar information. Still having a fusion between sonar and camera's information will increase the confidence level of objects' existence or free space, which is very useful for the estimation of potential automatic emergency braking. Consequently, polygon-generation and parking slot finder, which are derived from only 3D environmental sensing or together with park-marking detection will help to automatically and precisely detect possible empty parking targets in both sides of the ego-vehicle, see Fig. 6 and Fig. 7. Whereby either the vehicle system by its training or the driver by mobile-apps can decide or confirm the final target position.

Fig. 7 Example of park-marking detection projected in both perspective view and top-view on the sides. White dashed park-marking is detected and displayed with "P" mark.

Remarkably, the spatial error of parking slot detection is less than 5cm in the near range (< 7 m, depending on camera resolution) and less than 5 % of object distance in overall. This is by far much more accurate than the current state-of-the art solution, and very close to the result of Velodyne HDL-32 with more advantages of having 2D image information.

4. Autonomous Parking System

Even though having a detailed environmental map, highly accurate parking slot finder is not trivial, especially without park-marking or with a huge diversity of parking marking shapes and qualities across countries. In our solution, we use a graphical model to analyze possible parking spaces, which takes into account history of detected slots, boundaries of surrounding obstacles, and referenced initiative parking slot models. This helps to obtain a more generic parking solution with high confidence. The detailed understanding of the environment, as in Fig. 6, and the high confident and accurate detection of the target slot enable our system to detect both-sides parking slots as well as park in one stroke or maneuvering, which is infeasible from others. An example of our one-stroke parking capability is shown in Fig. 8 where the vehicle will directly park in the target position by one maneuvering without gear-shift change. Every available market-APS system needs a minimum of 3 strokes for box parking and 2 strokes for parallel parking for the common use-case with an available slot between two obstacles. As market-demand on a much shorter time-to-park due to the average of 5 strokes-parking for the current state-of-the art, our system has a PLUS with respect to this aspect.

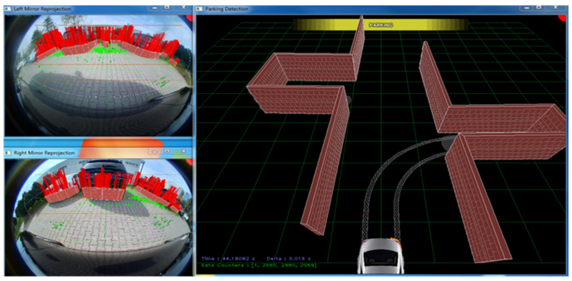


Fig. 8 Left: perspective views of the left and right mirror camera with reprojection of point cloud.
Right: 3D view for APS with planned trajectory in backward parking.

5. Experiments and Results

Our system includes 4 fisheye-cameras mounted at the mirrors, back and front of the ego-car following a standard configuration of a surrounding view system in automotive. Additionally, 12 sonars are installed on the bumper and distributed by 4 at the front, 4 at the back and 2 at each side. Our testing system is illustrated in the Fig. 9.

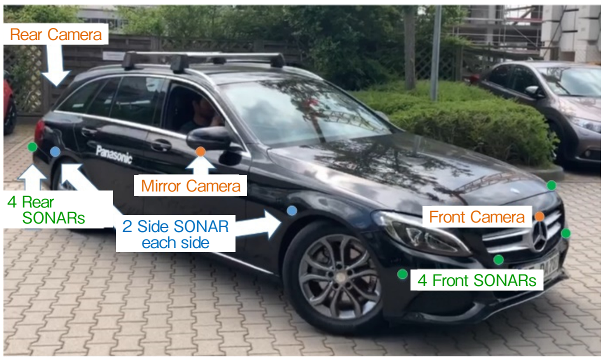


Fig. 9 Vehicle platform for testing our solution

500 test-drives on 20 different use-cases of parking scenarios have been evaluated with our system against other 6 available top-ranked market systems. The configuration and setup of scenarios are illustrated in Fig. 10. Additional scenarios, including baby/pedestrian dolls, bushes, cones, chains, etc. are also included in our tests. Preliminary qualitative experiments against similar solutions from major OEMs, offered in commercial vehicles, proved that our system performs superior to the competition. Concretely the success rate of automated parking is 94.1 % from our system and [51.2 %, 55.3 %, 47.2 %,

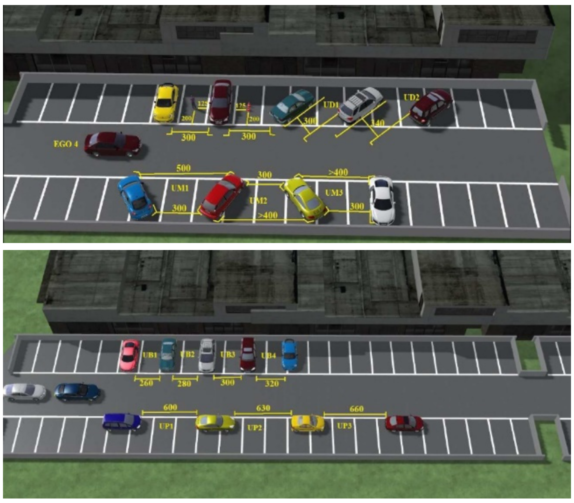


Fig. 10 Multiple scenarios set-up for benchmarking. The testing area is the parking zone of Panasonic Europe (Langen, Germany). Due to General Data Protection Regulation (GDPR) law, no real picture can be shown, but the simulated picture is very close to the real one.

58.7 %, 41.7 %, 40.7 %] respectively from the others. All challenging scenarios, including mis-aligned, narrow, diagonal, and between arbitrary objects (bushes, cones, chains, poles) are captured by our system but NOT by the market systems.

6. Conclusion

In conclusion, the proposed approach is the full camera-based autonomous parking solution, which offers the one-stroke parking for all common use-cases and challenging use-cases, meanwhile being accurate and embedded friendly for automotive production. Relatively dense environmental map provided by this approach is highly beneficial for reliable higher-automation functions and autonomous driving. The technology is easily adaptable for varieties of camera sensors at minimum cost. Since our mission is to develop products that will enrich the lives of our customers, we are very proud to introduce this cutting-edge technology enabling high performance at low-cost, and working world-wide under diverse conditions.

References

[1] A. Giesekeet al., Vehicular automated parking system, US20190308609A1 US patent, Magna Electronics Inc, 2019.

[2] Brian C. Hall, "Lie Groups, Lie Algebras, and Representations," ISBN 978-3-319-13467-3. Springer Verlag, Berlin, second edition, 2015.

[3] D. Scaramuzza et al., "A exible technique for accurate omnidirectional camera calibration and structure from motion," IEEE International Conference of Vision Systems, New York, USA, 2006.

[4] D.-V. Nguyen, "3D reconstruction for vehicle," EP3293700B1 European Patent, Panasonic Industry Europe GmbH, 2019.

[5] D.-V. Nguyen et al., "Method for determining a respective boundary of at least one object, sensor device, driver assistance device and motor vehicle," DE102014007565A1 German Patent, Connaught Electronics Ltd, 2015.

[6] J. Bammert et al., "Method and device for supporting a driver of a vehicle during the process of leaving a parking space," EP2257458B1 European Patent, Valeo Schalter und Sensoren GmbH, 2012.

[7] J. Engel et al., "Large-scale direct slam with stereo cameras," International Conference on Intelligent Robots and Systems (IROS), Hamburg, Sept.- Oct. 2015.

[8] J. Engel et al., "LSD-SLAM: Large-Scale Direct Monocular SLAM," ECCV 2014, Zurich, Sept. 2014.

[9] J. Engel et al., "Direct sparse odometry," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[10] K. Hoffsommer et al., "Method for supporting a driver of a motor vehicle when parking in a diagonal parking space, driver assistance

device and motor vehicle with a driver assistance device,” DE102013015349A1 European Patent, Valeo, 2013.

[11] L.-D. Elie et al., “System and method for autonomous valet parking using plenoptic cameras,” US10046760B2 US patent, Ford Global Technologies LLC, 2018.

[12] R. I. Hartley et al., Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, ISBN: 0521540518, second edition, 2004.

[13] R. Mur-Artal et al., “ORB-SLAM: a Versatile and Accurate Monocular SLAM System,” IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, 2015.

[14] S. Thrun et al., “The graph-slam algorithm with application to large-scale mapping of urban structures,”, International Journal of Robotics Research, vol. 25, issue 5-6, pp. 403-429, 2006.

[15] S. Thrun, D.et al., “A probabilistic approach to concurrent mapping and localization for mobile robots,” Machine Learning, vol. 31, pp. 29-53, Apr. 1998.

[16] S. Toledo et al., “Vehicle park assist system and method for parking a vehicle using such system,” US8378850B2 US patent, Ford Global Technologies LLC, 2013.

[17] Seong Soo Lee, “Parking assist device and parking assistance method for a vehicle,” DE102014116140B4 German patent, Hyundai Mobis Co Ltd, 2018.

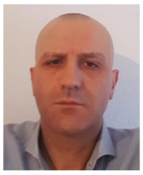
[18] T.-D. Barfoot, State Estimation for Robotics, Cambridge University Press, Cambridge, ISBN 9781107159396, first edition, 2017.

[19] Wellen Sham, “Automated vehicle parking,” US10407059B2 US patent, 2019.

[20] Y. Ma, et al., “An Invitation to 3-D Vision”, Springer Verlag, Berlin, ISBN: 0387008934, first edition, 2003.

Information About the Authors

Duong-Van Nguyen ドゥオンヴァン グエン
PhD in Robotics
Panasonic Automotive Systems Europe GmbH
パナソニックASヨーロッパ有限会社



Petros Kapsalas ペトロス カプサラス
PhD in Computer Vision
Panasonic Automotive Systems Europe GmbH
パナソニックASヨーロッパ有限会社

Shirish Khadilkar シリッシュ カディカール
Panasonic Automotive Systems Europe GmbH
パナソニックASヨーロッパ有限会社

Makoto Mochizuki 望月 誠
PhD in Engineering
Panasonic Automotive Systems Europe GmbH
パナソニックASヨーロッパ有限会社