Collaborative Localization Based on Traffic Landmarks for Autonomous Driving

Chen, Siheng; Zhang, Ningxiao; Sun, Huifang

TR2020-064 June 03, 2020

Abstract

Localizing an autonomous vehicle in real-time is critical for robust autonomous driving. As a standard approach, the mapbased localization is robust and fast; however, it is expensive to create and maintain a large-scale high-definition map. In this paper, we propose an online localization technique based on the vehicle-to-vehicle communication and traffic landmark detection; called collaborative localization. This can potentially serve as a new complement to the standard localization solutions. We theoretically show that multiple vehicles with multiple traffic landmarks would significantly improve the localization performance. We then propose a practical algorithm, which leverages graph matching to handle practical issues, such as traffic landmark association. The experimental results validate the potential of the proposed methods.

IEEE International Symposium on Circuits and Systems (ISCAS)

COLLABORATIVE LOCALIZATION BASED ON TRAFFIC LANDMARKS FOR AUTONOMOUS DRIVING

Siheng Chen [1] *Ningxiao Zhang* [1,2] *Huifang Sun* [1]

[1] Mitsubishi Electric Research Laboratories, Cambridge, USA
[2] The Pennsylvania State University, State College, PA, USA

ABSTRACT

Localizing an autonomous vehicle in real-time is critical for robust autonomous driving. As a standard approach, the map-based localization is robust and fast; however, it is expensive to create and maintain a large-scale high-definition map. In this paper, we propose an online localization technique based on the vehicle-to-vehicle communication and traffic landmark detection; called *collaborative localization*. This can potentially serve as a new complement to the standard localization solutions. We theoretically show that multiple vehicles with multiple traffic landmarks would significantly improve the localization performance. We then propose a practical algorithm, which leverages graph matching to handle practical issues, such as traffic landmark association. The experimental results validate the potential of the proposed methods.

***Index Terms*—** Autonomous driving, localization, vehicle-to-vehicle communication, graph matching, traffic landmarks

1. INTRODUCTION

Autonomous driving might fundamentally reduce traffic accidents caused by driver errors and save parking spaces, especially in the urban areas [1, 2, 3]. Autonomous software is a complicated system that includes multiple modules, such as localization, perception, planning and control [4]. A robust autonomous vehicle requires each module to be reliable; however, the current autonomous software is usually based on machine learning techniques and can hardly achieve reliable performances for those cases that are not included in the training dataset. It is thus believed that the communication techniques, such as the vehicle-to-vehicle communication, can be used to ease the designs of many autonomous modules and make autonomous vehicles more reliable. In this paper, we consider using communication techniques to improve the localization of an autonomous vehicle.

Localization is the task of finding the ego-position of an autonomous vehicle relative to a reference position. This module is crucial because an autonomous vehicle must localize itself to follow the correct lane and provide priors for the perception, prediction, and planning modules. Since GPS usually has huge variances in urban area, it cannot be used to provide precise localization. A standard localization solution for autonomous driving is the map-based localization; that is, we aim to find the location of the best match in an predesigned high-definition (HD) map [5]. Compared to simultaneous localization and mapping (SLAM) [6, 7], the map-based localization takes advantages of an offline HD map, providing both fast and precise localization; however, it is expensive to create and update a large-scale fully-annotated HD map.

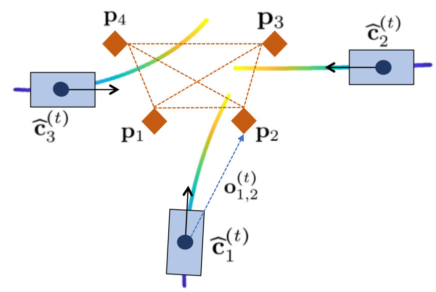


Fig. 1: Collaborative localization. We consider multiple vehicles are driving in the scene. Each vehicle is carrying sensors to detect the ranges of the nearby traffic landmarks. Even the locations of traffic landmarks are unknown, their pairwise distances are fixed. We thus use graph matching to take advantage of this property and achieve localization. We also allow vehicles to communication and share observation with each other, further improving the localization performance. In the plot, the $\mathbf{p}_n$ denotes the locations of traffic landmarks, the $\widehat{c}_m^{(t)}$s denote the real-time locations of vehicles, and $\mathbf{o}_{n,m}^{(t)}$s denote the real-time range observation from vehicles to traffic landmarks. We use a graph structure to represent the pairwise distances between traffic landmarks. This unknown, yet fixed graph structure is the key to localization.

In this paper, we consider a hypothetical setting: multiple autonomous vehicles are driving around an urban area; each vehicle has the ability to accurately detect traffic landmarks, such as traffic lights, traffic signs and land markings; and each vehicle has the ability to communicate a small amount information with its neighboring vehicle in real-time. Based on this hypothetical setting, we propose collaborative localization, which allows autonomous vehicles to localize each other collaboratively; see Figure 1. We first theoretically show that using multiple traffic landmarks with multiple autonomous vehicles significantly benefit localization. We next take advantages of the fixed pairwise relationships between traffic landmarks, and use graph matching to fix the practical issues, such as the association of traffic landmarks and the alignment of actors' headings across time. We finally present a practical algorithm for the collaborative localization of autonomous vehicles. In the experiments, we generate a simulation dataset based on the bicycle model. Based on the simulation dataset, we validate the proposed theory and the practical algorithm.

2. RELATED WORK

Simultaneous localization and mapping(SLAM). SLAM is a technique to generate the map of a robot's surroundings and locates the robot in that map at the same time [6, 7]. The related techniques have been widely used in autonomous driving and many other robotic systems. Some popular approaches include the Bayes-filter-based algorithms, and graph-based algorithms. The Bayes-filter-based algorithms solve the SLAM problem in an online fashion. They employ Bayes filters to predict and optimize the map and LiDAR poses iteratively based on the online sensor measurements. The graph-based algorithms optimize all the poses together by using all sensor measurements across time. They construct a pose graph that models the relations among poses. They thus convert the SLAM problem is thus to the minimization of the total sum of the edge weights of a pose graph.

Map-based localization. The basic idea of the map-based localization is to estimate the 6 degree-of-freedom pose of a vehicle by matching the real-time sensing data to the predesigned HD map. For autonomous driving, a map-based localization system usually consists of two components [5, 2]. The first component is the LiDAR-to-map registration, which computes the vehicle pose by registering real-time LiDAR data to the HD map. The second component is the multisensor fusion, which estimates a confident pose from the multisensor readings, including IMU, odometer, and GPS, as well as the pose output from the LiDAR-to-map registration. Compared to SLAM, the map-based localization provides much more reliable and faster localization; however, it is expensive to create and update a city-scale HD map. Extracting the detailed traffic-related semantic features, such as traffic lights and lane boundaries, heavily relies on human supervision, which is both expensive and time-consuming.

Vehicle-to-everything (V2X). V2X is a series of communication techniques that enable information transmission from a vehicle to other entities that may affect the vehicle. It incorporates multiple types of communication, such as vehicle-to-infrastructure (V2I), vehicle-to-network (V2N) and vehicle-to-vehicle (V2V) [8, 9]. It is believed to be a powerful complement to machine learning techniques to achieve autonomous driving [10]. The hypothetical setting considered in this paper is based on V2V.

3. METHODOLOGY

In this section, we formulate the task of collaborative localization. We then introduce four hypothetical strategies to achieve the localization. We finally use graph matching to alleviate the constraints and propose a practical algorithm.

3.1. Problem Formulation

The overall goal of collaborative localization is to allow multiple actors to localize themselves through sharing a small amount of observation information to each other. Consider M actors in a scene and $\mathbf{c}_m^{(t)} \in \mathbb{R}^2, \mathbf{h}_m^{(t)} \in [-\pi, \pi)$ be the ground-truth position and heading of the ith actor at time t, respectively. Consider each actor is carrying sensors, such as LiDAR and camera, and has the ability to detect all fixed traffic landmarks from predesigned categories in a scene in real-time, such as traffic lights and traffic signs [11, 12]. We can then convert the detections to the corresponding range observations.

Let $\{\mathbf{p}_n \in \mathbb{R}^2\}_{n=1}^N$ be a set of traffic landmarks, where $\mathbf{p}_n$ is the position of the nth traffic landmark. Let $\mathbf{o}_{m,n}^{(t)} \in \mathbb{R}^2$ be the observation from the mth actor to the nth traffic landmark at time t. We thus have

$$\mathbf{o}_{m,n}^{(t)} = R(\mathbf{h}_m^{(t)}) \left(\mathbf{p}_n - \mathbf{c}_m^{(t)} + \epsilon_{m,n}^{(t)} \right) \in \mathbb{R}^2, \quad (1)$$

where $R(\cdot) \in SO(2)$ is the rotation matrix and $\epsilon_{m,n}^{(t)}$ is the observation noise. The observation noise comes from the detection error, which could be related to the size of the traffic landmarks and the relative range between an actor and a traffic landmark. In this paper, we simply consider the Gaussian noise model; that is, $\epsilon_{m,n}^{(t)} \sim \mathcal{N}(0, \sigma^2)$. We consider that the actors can share the information about their initial positions and headings, $\mathbf{c}_m^{(0)}$ and $\mathbf{h}_m^{(0)}$, as well as the observations, $\mathbf{o}_{m,n}^{(t)}$ in real-time. We aim to estimate all the actors' positions and headings, $\mathbf{c}_m^{(t)}$ and $\mathbf{h}_m^{(t)}$, in real-time.

3.2. Hypothetical localization strategies

To ease the problem, we first make two assumptions to achieve some theoretical insights. Suppose that (1) the associations of traffic landmarks across frames are known; and (2) the headings of all the actors are known. Based on these, we consider the following four localization strategies.

Single-actor-single-landmark (SASL)-based strategy. We first consider that each actor uses its own observation without any communication and there is only a single landmark; that is, $M = 1$ and $N = 1$.

Theorem 1. Let the estimated position of the actor at time t be

$$\widehat{\mathbf{c}}_1^{(t)} = \mathbf{c}_1^{(0)} + R(-\mathbf{h}_1^{(0)})\mathbf{o}_{1,1}^{(0)} - R(-\mathbf{h}_1^{(t)})\mathbf{o}_{1,1}^{(t)}. \quad (2)$$

The localization error is $\mathbb{E} \left\| \widehat{\mathbf{c}}_1^{(t)} - \mathbf{c}_1^{(t)} \right\|_2^2 = 2\sigma^2$.

Proofs are omitted to conserve space.

Single-actor-multiple-landmark (SAML)-based strategy. We next consider that there are multiple landmarks; that is, $M = 1$ and $N > 1$.

Theorem 2. Let the estimated position of the actor at time t be

$$\widehat{\mathbf{c}}_1^{(t)} = \mathbf{c}_1^{(0)} + \frac{1}{N}\sum_{n=1}^{N}\left(R(-\mathbf{h}_1^{(0)})\mathbf{o}_{1,n}^{(0)} - R(-\mathbf{h}_1^{(t)})\mathbf{o}_{1,n}^{(t)}\right). \quad (3)$$

The localization error is $\mathbb{E}\left\|\widehat{\mathbf{c}}_1^{(t)} - \mathbf{c}_1^{(t)}\right\|_2^2 = 2\sigma^2/N$.

Comparing Theorems 1 and 2, we see that leveraging multiple landmarks significantly reduce the localization error.

Multiple-actor-single-landmark (MASL)-based strategy. We next allow multiple actors to communicate with each other and there is a single landmark; that is, $M > 1$ and $N = 1$.

Theorem 3. Let the estimated center position of the mth vehicle at time t is

$$\widehat{\mathbf{c}}_m^{(t)} = \frac{1}{M}\sum_{i=1}^{M}\left(\widehat{\mathbf{c}}_i^{(0)} + R(-\mathbf{h}_i^{(0)})\mathbf{o}_{i,1}^{(0)}\right) - R(-\mathbf{h}_m^{(t)})\mathbf{o}_{m,1}^{(t)}. \quad (4)$$

The localization error is $\mathbb{E}\left\|\widehat{\mathbf{c}}_m^{(t)} - \mathbf{c}_m^{(t)}\right\|_2^2 = \left(1+\frac{1}{M}\right)\sigma^2$.

Comparing Theorems 1 and 3, we see that allowing multiple actors to communicate also reduces the localization error.

Multiple-actor-multiple-landmark (MAML)-based strategy. We finally allow multiple actors to communicate with each other and there are multiple landmarks; that is, $M > 1$ and $N > 1$. The multiple-actor-multiple-landmark-based localization strategy is as follows.

Theorem 4. Let the estimated center position of the mth vehicle at time t is

$$\widehat{\mathbf{c}}_m^{(t)} = \frac{1}{M}\sum_{i=1}^{M}\left(\mathbf{c}_i^{(0)} + \frac{1}{N}\sum_{n=1}^{N}\left(R(-\mathbf{h}_i^{(0)})\mathbf{o}_{i,n}^{(0)} - R(-\mathbf{h}_m^{(t)})\mathbf{o}_{m,n}^{(t)}\right)\right) \quad (5)$$

The localization error is $\mathbb{E}\left\|\widehat{\mathbf{c}}_m^{(t)} - \mathbf{c}_m^{(t)}\right\|_2^2 = \left(1+\frac{1}{M}\right)\frac{\sigma^2}{N}$.

Theorem 4 shows that leveraging multiple actors and landmarks can significantly reduce the localization error.

3.3. Practical algorithm

Theorems 1, 2, 3, and 4 show the potential benefit of using collaborative localization via traffic landmarks; however, the unknown association of traffic landmarks across frames and the unknown headings in real-time prevent those strategies from being practical. Here we present a graph-matching-based technique to associate the traffic landmarks across frames and estimate actors' headings.

Traffic landmark association. At each frame, an actor can detect several traffic landmarks; however, the index for the same traffic landmark could be changed across frames. We need to associate the same traffic landmarks. Here we use graph matching to achieve the traffic landmark association. The intuition behind this idea is that the relative positions between traffic landmarks are fixed. A graph can be constructed to reflect the pairwise distances between traffic landmarks at each frame. We can permute the order of nodes at one frame to match that at another frame, thus converting the association task into graph matching.

For the mth actor at time t, we can construct a graph $\mathrm{A}_m^{(t)} \in \mathbb{R}^{N\times N}$ to model the pairwise relationships between traffic landmarks, whose nodes are traffic landmarks and edges are pairwise distances. The (n,n')th element is

$$(\mathrm{A}_m^{(t)})_{n,n'} = \left\|\mathbf{o}_{m,n'}^{(t)} - \mathbf{o}_{m,n}^{(t)}\right\|_2^2. \quad (6)$$

To find an appropriate permutation to match graphs constructed at different frames, we need to solve the following optimization problem about the permutation matrix J.

$$\widehat{\mathrm{J}}_m^{(t)} = \arg\min_{\mathrm{J}\in\{0,1\}^{N\times N}} \left\|\mathrm{A}_m^{(0)} - \mathrm{J}\,\mathrm{A}_m^{(t)}\,\mathrm{J}^T\right\|_2^2, \quad (7)$$
$$\text{subject to } \mathrm{J}\,\mathbf{1}_N = \mathbf{1}_N, \mathrm{J}^T\,\mathbf{1}_N = \mathbf{1}_N.$$

The solution permutes the order of the traffic landmarks at time t to match the order at time 0 for the mth actor. The optimization problem (7) is a standard graph matching problem, which can be solved by many existing packages [13].

Heading alignment. The heading of an actor can be rotated all the time, which changes the coordinate system of observations. We can leverage the pairwise relationships between traffic landmarks to find the relative rotation from a given frame to the reference frame for the same actor. To find an appropriate rotation to synchronize the headings across frames, we need to solve the following optimization problem.

$$\widehat{\mathrm{R}}_m^{(t)} = \arg\min_{\mathrm{R}\in SO(2)} \sum_{n,n'} \left\|\mathrm{R}(\mathbf{o}_{m,n}^{(t)} - \mathbf{o}_{m,n'}^{(t)}) - (\mathbf{o}_{m,n}^{(0)} - \mathbf{o}_{m,n'}^{(0)})\right\|_2^2 \quad (8)$$

The detailed algorithm of solving (8) sees [14]. The solution $\widehat{\mathrm{R}}_m^{(t)}$ estimates the relative rotation from the frame t to the reference frame, $R(\mathbf{h}_m^{(t)} - \mathbf{h}_m^{(0)})$. We thus can obtain the estimation of the heading at time t; that is,

$$\widehat{\mathbf{h}}_m^{(t)} = R^{-1}\left(\widehat{\mathrm{R}}_m^{(t)} R(\mathbf{h}_m^{(0)})\right). \quad (9)$$

Note that (1) we need no less than three traffic landmarks to solve (8); and (2) since we are dealing with multiple traffic landmarks, we need to do traffic landmark association first.

Localization algorithms. Based on the traffic landmark association and the heading alignment, we are able to make the single-actor-multiple-landmark-based strategy and the multiple-actor-multiple-landmark-based strategy practical. Note that two single-landmark-based strategies are not feasible because we cannot align the headings of an actor across time. see Algorithm 1 for the complete procedures for the multiple-actor-multiple-landmark (MAML)-based algorithm. The single-actor-multiple-landmark (SAML)-based algorithm is a special case of the multiple-actor-multiple-landmark-based algorithm when we set the number of actors to one.

4. EXPERIMENTAL RESULTS

In this section, we generate a simulation dataset. Based on the simulation dataset, we first validate our hypothetical strategies and then test two practical algorithms.

Algorithm 1 Multiple-actor-multiple-landmark (MAML)-based localization algorithm

Input t frame ID
$\mathbf{c}_m^{(0)}$ initial position of the mth actor
$\mathbf{h}_m^{(0)}$ initial heading of the mth actor
$\mathbf{o}_{m,n}^{(t)}$ observation of the mth actor to the nth landmark at time t

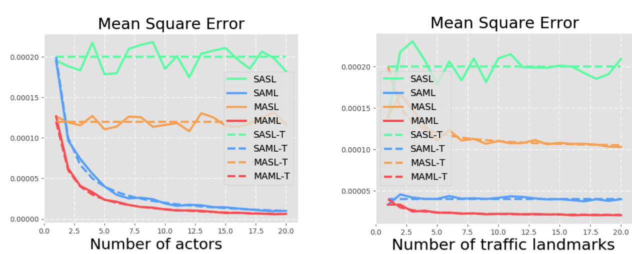
Output $\widehat{\mathbf{c}}_m^{(t)}$ real-time position of the mth actor
$\widehat{\mathbf{h}}_m^{(t)}$ real-time heading of the mth actor

Function for m = 1 to M
construct a graph $\mathrm{A}_m^{(t)}$ via (6)
obtain the permutation matrix $\widehat{\mathrm{J}}_m^{(t)}$ via (7)
associate traffic landmarks, $\widehat{\mathrm{J}}_m^{(t)}\,\mathrm{A}_m^{(t)}\left(\widehat{\mathrm{J}}_m^{(t)}\right)^T$
obtain the relative rotation, $\widehat{\mathrm{R}}_m^{(t)}$, via (8)
estimate the heading at time t, $\widehat{\mathbf{h}}_m^{(t)}$, via (9)
estimate the position at time t, $\widehat{\mathbf{c}}_m^{(t)}$, via (5)
return $\widehat{\mathbf{c}}_m^{(t)}, \widehat{\mathbf{h}}_m^{(t)}$

Simulation dataset. We randomly sample N positions from the unit square $[0,1]^2$ to generate N traffic landmarks. To generate the real-time trajectory for each one of N actors, we consider a dynamic model as follows [15].

$$\begin{cases} \text{center}: & \mathbf{c}_m^{(t+\tau)} = \mathbf{c}_m^{(t)} + \int_{\eta=t}^{t+\tau} \mathbf{v}_i^{(\eta)} \mathbf{u}_m^{(\eta)} d\eta \in \mathbb{R}^2, \\ \text{speed}: & \mathbf{v}_m^{(t+\eta)} = \mathbf{v}_m^{(t)} + \int_{\eta=t}^{t+\tau} \mathbf{a}_m^{(\eta)} d\eta \in \mathbb{R}, \\ \text{heading vector}: & \mathbf{u}_m^{(t)} = \left[\cos(\mathbf{h}_m^{(t)}) \quad \sin(\mathbf{h}_m^{(t)})\right]^T \in \mathbb{R}^2, \\ \text{heading}: & \mathbf{h}_m^{(t+\tau)} = \mathbf{h}_m^{(t)} + \int_{\eta=t}^{t+\tau} \mathbf{w}_m^{(\eta)} d\eta \in \mathbb{R}. \end{cases}$$

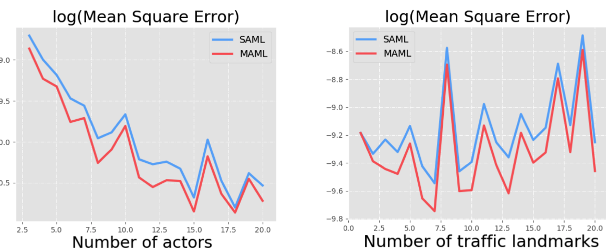
We randomly generate the initial position, speed and heading of each actor, $\mathbf{c}_m^{(0)}, \mathbf{v}_m^{(0)}, \mathbf{h}_m^{(0)}$, as well as the acceleration profile and angular speed profile, $\mathbf{a}_m^{(t)}$ and $\mathbf{w}_m^{(t)}$. The variance of the observation noise is $\sigma^2 = 0.1$.



(a) $M = 5$. (b) $N = 5$.

Fig. 2: Validation of hypothetical strategies. Plots (a) shows the localization errors as a function of the number of traffic landmarks. Plots (b) shows the localization errors as a function of the number of traffic actions. We see that the simulation matches the theory well.

Validation of hypothetical strategies. Here we assume the association of traffic landmarks and the real-time headings are known. We aim to validate the effectiveness of using multiple actors and traffic landmarks. Figure 2 shows the localization performances with various pairs of M, N under four strategies. The localization performance is evaluated based on the mean square error averaged across all the actors and all the frames. We see that (1) given a fixed number of actors ($M = 5$), the MASL and MAML-based strategies significantly outperforms when the number of traffic landmarks increases; and (2) given a fixed number of traffic landmarks ($N = 5$), the SAML and MAML-based strategies significantly outperforms when the number of actors increases; (3) the MAML-based strategy outperforms the other three strategies across all the scenarios. These three observations all match the theoretical analysis in Theorems 1, 2, 3, and 4. To summarize, using multiple actors and traffic landmarks can improve the real-time localization performance in theory.



(a) $M = 5$. (b) $N = 5$.

Fig. 3: Test of practical algorithms. Enabling multiple actors to communicate with each other makes the localization error decrease by around 20% in average.

Test of practical algorithms. Now we consider that the association of traffic landmarks and the real-time headings are unknown as in a practical scenario and use graph matching to find the association and align the heading. We test the practical localization performances of using the SAML-based and MAML-based localization algorithms. Figure 3 shows the localization performances with various pairs of M, N under the SAML and MAML algorithms. The localization performance is evaluated based on the mean square error averaged across all the actors and frames. We see that (1) given a fixed number of actors ($M = 5$), the SAML and MAML-based algorithms significantly outperforms as the number of traffic landmarks increases; and (2) given a fixed number of traffic landmarks ($N = 5$), the SAML and MAML-based algorithms stay similarly as the number of actors increases; and (3) the MAML-based algorithm outperforms the SAML-based algorithm by 20% in average.

5. CONCLUSIONS

In this paper, we propose collaborative localization that allows autonomous vehicles to localize themselves by detecting traffic landmarks and communicating with each other. We theoretically show that multiple vehicles with multiple traffic landmarks would significantly improve the localization performance. We then propose practical algorithms. The experimental results validate the potential of the proposed methods. This technique could potentially serve as a complementary approach to the current solution.

6. REFERENCES

[1] A. Taeihagh and H. Si Min Lim, "Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks," *Transport Reviews*, vol. 39, no. 1, pp. 103–128, Jan. 2019.

[2] C. Badue, R. Guidolini, R. Vivacqua Carneiro, P. Azevedo, V. Brito Cardoso, A. Forechi, L. Ferreira Reis Jesus, R. Ferreira Berriel, T. Meireles Paixão, F. Mutz, T. Oliveira-Santos, and A. Ferreira De Souza, "Self-driving cars: A survey," arXiv:1901.04407 [cs.RO], Jan. 2019.

[3] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Transactions on Intelligent Vehicles*, Jan. 2019, Submitted.

[4] C. Urmson, J. Anhalt, D. Bagnell, C. R. Baker, R. Bittner, M. N. Clark, J. M. Dolan, D. Duggins, T. Galatali, C. Geyer, M. Gittleman, S. Harbaugh, M. Hebert, T. M. Howard, S. Kolski, A. Kelly, M. Likhachev, M. McNaughton, N. Miller, K. M. Peterson, B. Pilnick, R. Rajkumar, P. E. Rybski, B. Salesky, Y-W. Seo, S. Singh, J. M. Snider, A. Stentz, W. Whittaker, Z. Wolkowicki, J. Ziglar, H. Bae, T. Brown, D. Demitrish, B. Litkouhi, J. Nickolaou, V. Sadekar, W. Zhang, J. Struble, M. Taylor, M. Darms, and D. Ferguson, "Autonomous driving in urban environments: Boss and the urban challenge," in *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic, George Air Force Base, Victorville, California, USA*, 2009, pp. 1–59.

[5] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, 2018.

[6] H. F. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robot. Automat. Mag.*, vol. 13, no. 2, pp. 99–110, 2006.

[7] T. Bailey and H. F. Durrant-Whyte, "Simultaneous localization and mapping: part II," *IEEE Robot. Automat. Mag.*, vol. 13, no. 3, pp. 108–117, 2006.

[8] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, 2014.

[9] H. Ye, G. Y. Li, and B-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Vehicular Technology*, vol. 68, no. 4, pp. 3163–3173, 2019.

[10] S. Zhang, H. Zhang, B. Di, and L. Song, "Cellular cooperative unmanned aerial vehicle networks with sense-and-send protocol," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1754–1767, 2019.

[11] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision, 2017, Venice, Italy, October 22-29, 2017*, 2017, pp. 2980–2988.

[12] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn.*, Salt Lake City, UT, USA, June 2018, pp. 4490–4499.

[13] Si Zhang and Hanghang Tong, "Attributed network alignment: Problem definitions and fast solutions," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 9, pp. 1680–1692, 2019.

[14] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 698–700, 1987.

[15] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, "Kinematic and dynamic vehicle models for autonomous driving control design," in *2015 IEEE Intelligent Vehicles Symposium, IV 2015, Seoul, South Korea, June 28 - July 1, 2015*, 2015, pp. 1094–1099.