Mechatronics 51 (2018) 46–58

Contents lists available at ScienceDirect

Mechatronics

journal homepage: www.elsevier.com/locate/mechatronics

Integration of vision and topological self-localization for intelligent vehicles☆

Yicheng Li[a,b], Zhaozheng Hu[a,*], Yuezhi Hu[a], Duanfeng Chu[a]

[a] Intelligent Transportation Systems (ITS) Research Center, Wuhan University of Technology, 1040 Heping Avenue, Wuchang District, Wuhan 430063, PR China
[b] National Engineering Research Center for Water Transport Safety, 1040 Heping Avenue, Wuchang District, Wuhan 430063, PR China

ARTICLE INFO

Keywords:
Intelligent vehicle
Self-localization
Topological localization
Visual localization
Vision-speed

ABSTRACT

Self-localization is a crucial task for intelligent vehicles. Conventional localization methods usually suffer from different limitations, such as low accuracy and blind areas for Global Positioning System (GPS), high cost for Inertial Navigation System (INS), and low robustness for vision Simultaneously Localization and Mapping (vSLAM). To overcome these problems, this study proposes a low-cost yet accurate localization method for intelligent vehicles, which only needs a monocular camera and a GPS receiver. First, the proposed method offers multiple feature spaces which are designed from GPS data, localization prediction model, image holistic features, and image local features. Each feature space, from which one candidate set is derived, can make qualitative localization achieved independently. Afterwards, we propose a novel method called K-Nearest Neighbors from Multiple Feature Spaces (KNN-MFS) to fuse these candidate sets. The closest node to the current vehicle position is drawn from the visual map to achieve image-level localization. Finally, the vehicle pose in the visual map, computed by metric localization, can further enhance the localization accuracy. The advantage is that when GPS signals are unavailable at times, the method can still achieve short-range localization by using other feature spaces. The proposed method has been validated with the actual data sets and public data sets. The actual data sets are collected along an industrial park and a rural urban fringe in Wuhan City, China, covering different times and weather conditions, and the total lengths of these routes have to be more than 8 km. The public data sets are Karlsruhe Institute of Technology and Toyota Technology Institute (KITTI). Experimental results show that the proposed method can adapt to different times and weather conditions with good robustness in varying environments, and the localization errors are less than 35 cm in all the tests in average. Experimental results on the same routes without GPS data are also reported, which demonstrate that the proposed method can achieve comparable localization accuracy.

1. Introduction

1.1. Background

In recent years, intelligent vehicles have attracted increasing attentions from a broad audience, both in industry and academic fields [1]. Self-localization of intelligent vehicle is to compute the current vehicle position, usually expressed in a rotation matrix and a translation vector. It is essential for intelligent vehicle navigation and driving decision making. Therefore, self-localization is a pre-requisite for autonomous driving of different levels (i.e., from Society of Automotive Engineers (SAE) Level 1 to SAE Level 4).

Over the past decades, various self-localization methods have been proposed. The easily mounted Global Positioning System (GPS) receiver is the most common sensor that localizes vehicle position. However, this sensor suffers from low localization accuracy (i.e., the error can be 10 m or more) and blind areas especially in urban environments [2]. To enhance the localization accuracy [3], an Inertial Navigation System (INS) is implemented in the vehicle, which also depends on GPS. The localization accuracy is significantly improved compared with only GPS receiver. However, this sensor is too expensive enough for people to apply its use on affordable vehicles and it has blind areas. Odometry is another choice as it is available in most in-vehicle systems. This kind of sensors is used for monitoring the rotation of wheels. The posese of vehicle can be estimated in a high frequency. However, the sensors only provide a relative localization. It means that odometry computes the current position relative the last pose. As a result, the estimation errors are accumulated over time [4]. Moreover, laser scanners are also used

☆ This paper was recommended for publication by Associate Editor Roger Dixon.

* Corresponding author.
E-mail address: zzhu@whut.edu.cn (Z. Hu).

https://doi.org/10.1016/j.mechatronics.2018.02.012
Received 31 December 2016; Received in revised form 31 December 2017; Accepted 25 February 2018

for vehicle localization [5,6]. Some researchers mount lasers on vehicles and use to perform continuous scan-matching. The collected three-dimensional (3D) Light Detection And Ranging (LIDAR) point clouds have localization capabilities [7], and localization accuracy is enhanced to a few centimeters. However, laser scanners are expensive and their localization procedure also has a high computational complexity. As discussed above, the intelligent vehicles need a low-cost and high-accuracy self-localization algorithm which can help them make a prompt and effective decision. With the development of in-vehicle cameras being used in large scale of vehicles, visual localization method is becoming a hot topic in recent years. Compared with lasers and INS, cameras are inexpensive. Vision localization method can either combine with GPS data or be independent of them. If GPS data are available, then this method can enhance the localization accuracy. Otherwise, intelligent vehicles can also use a camera to localize themselves in a short range.

1.2. Literature review

From the literature, one of the most popular vision localization methods is simultaneous localization and mapping (SLAM). This method solves the computational problem of constructing or updating a map of an unknown environment while simultaneously keeping track of an agent's location within it. Further details about SLAM are presented in [8–11]. Our work presented in this paper is strongly related to SLAM, such as place recognition. Place recognition sometimes refers to loop-closure detection in the SLAM literature [12]. The difference between the two kinds of methods is that the presented work divides localization into two steps, visual map creation and localization. This process can improve the efficiency of place recognition.

In this method, the core task of vision-based localization is to find the closest node in the visual map. Visual place recognition is the main method to accomplish this goal [13]. When an image of a place is input, a robot or vehicle need to decide this place whether or not has already been seen. This is based on whether image features can be stably extracted. Scale-invariant Feature Transform (SIFT) is perhaps one of the most popular feature extraction methods [14]. This method has strong extraction stability and accurate detection results. However, its complex computational processes give rise to lower detection efficiency. Based on this situation, Speeded-Up Robust Features (SURF) is proposed [15]. This method simplifies the computation and has a similar result to SIFT. In [16], Valgren and Lilienthal evaluate the features of SIFT and SURF and use both methods for long-term topological localization. However, the matching efficiency of SURF features cannot meet the requirements of intelligent vehicle localization, and using these descriptors delays the localization speed. To improve the matching efficiency, Rublee et al. propose a fast binary descriptor, called Oriented FAST and Rotated BRIEF (ORB) [17]. ORB is at two orders of magnitude faster than SIFT and one order of magnitude faster than SURF. Mur-Artal et al. [18] present an ORB-based monocular SLAM system in real time. They use ORB descriptor for loop closing, relocalization, mapping and tracking. The system can be evaluated in outdoor environments. Therefore, we also utilize ORB feature descriptor to realize intelligent vehicle localization in our approach.

However, the only application of feature matching could result in large computation quantity and low localization accuracy because of the huge data source in a visual map. Hence, more advanced methods should be proposed to enhance the accuracy. For example, Son et al. propose a key frame selection method to reduce the matching complexity [19]. Images of a visual map are divided into key and non-key frames by checking the matching number of feature points. Query image is first matched with key frames, and the nearest key frame is found from the visual map. This approach can reduce the complexity of matching. Topological localization is another model that enhances accuracy and reduces computation complexity. For example, Lategahn et al. [20] propose a two-step approach. First, a holistic feature is matched with the features from the visual map through topological localization; then, a dynamic programming process is used to find the node closest to the current position. The node is in the map. However, this procedure needs to use the 30 positions as prior information. As a result, the dynamic programming has high computational complexity. Similarly, Andreasson et al. use a topological approach by SIFT feature matching in an indoor environment [21]. In addition, Murillo et al. [22] propose a two-step approach. First, they obtain the topological localization results, which are used for SIFT feature matching. Thereafter, they refine the localization results by computing the camera pose. Meanwhile, Badino et al. [23,24] design the localization system with Bayesian filter and topological localization taken into consideration. One previous localization result is used as prior information. They compute the condition probability of each possible position, and then localize the result. This model is simple and leads to some outliers. However, only using vision-based localization methods will lead to accumulative errors. The above mentioned methods are generally used in short-range routes.

There are some problems for visual localization, such as changes of illumination and long-range localization [13]. To address these problems, Furgal and Barfoot describe a visual teach and repeat system by a stereo camera [25]. Visual teach is similar with visual map creation. Stereo features are extracted by stereo images and then used to build a map database. In the step of repeat, which is similar with localization, features are matched with map database for localization. To improve the accuracy, they also capture stereo images by stereo camera for visual odometry. 3D poses obtained from visual odometry are fused with localization results. This system is able to localize a long-range route. The system brings us inspiration: we can utilize stereo camera for visual map creation and monocular camera for localization. Consequencely, the 3D poses are obtained from visual map. It can reduce the cost of localization. Moreover, Churchill and Newman [26] set up a learning model to address changes of illumination. They first build a world model over many runs. The runs accumulate distinct visual experiences in concert. When the robot navigates itself, these previous experiences provide prior information for localization. Overall, this method requires huge data storage for world model.

There are also some studies using additional sensors, such as Global Navigation Satellite System (GNSS), speedometers or odometry, as aid to intelligent vehicle localization. The advantage of using additional sensors is that they can eliminate accumulated errors caused by vision-based localization. For example, Li et al. [27] propose a localization method by fusing images, GPS data, and map data. GPS data are used to determine a possible position range of a vehicle. Vision based lane detection provides lateral positioning. Vision based traffic sign detection provides longitudinal localization. Simulation-based experiments show that the accuracies of longitudinal and lateral localization are 0.51 and 0.09 m, respectively. Experiments on actual data, Rehder et al. describe the sensor fusion of GPS and INS in long-range odometry. Therefore, a high accuracy of visual odometry can be achieved even if only distant feature points are available [28]. Schreiber et al. [29] integrate an onboard stereo camera system with a low-cost GNSS receiver. In their approach, when GNSS data are available, these data are used for localization. Otherwise, they use vision-based localization as visual odometry. Their research does not use data fusion method, and this approach enables vehicles to achieve a 10 m-level localization accuracy. For multiple data fusion, Hojun et al. [30] propose an Extended Kalman Filter (EKF) to fuse GPS data, vehicle velocity and yaw rate and image data. The velocity and yaw rate are collected by in-vehicle sensors. The research achieves a sub-meter-level localization accuracy. Shunsuke et al. [31] develop a localization system that integrates GNSS, INS, and camera. To achieve lane-level localization, Bayesian filter is used to fuse GNSS data, INS data and images. This system can achieve meter-level localization. Krajnik et al. fuse odometry and monocular vision to accomplish navigation [32]. In their system, the odometry provides a relative localization result. The monocular vision is to