Robotics and Autonomous Systems 83 (2016) 338–348

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Radar and stereo vision fusion for multitarget tracking on the special Euclidean group☆

Josip Ćesić*, Ivan Marković, Igor Cvišić, Ivan Petrović

University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Control and Computer Engineering, Unska 3, 10000 Zagreb, Croatia

HIGHLIGHTS

- Radar and stereo camera integration for tracking in ADAS.
- Detection and tracking of moving objects by filtering on matrix Lie groups.
- State space formed as a product of two special Euclidean groups.
- Employed banana-shaped uncertainties typical for range-bearing sensors and vehicles in motion.
- JIPDA filter for multitarget tracking on matrix Lie groups.

ARTICLE INFO

Article history:
Available online 12 May 2016

Keywords:
Advanced driver assistance systems
Detection and tracking of moving objects
Joint integrated probabilistic data association
Radar
Stereo camera

ABSTRACT

Reliable scene analysis, under varying conditions, is an essential task in nearly any assistance or autonomous system application, and advanced driver assistance systems (ADAS) are no exception. ADAS commonly involve adaptive cruise control, collision avoidance, lane change assistance, traffic sign recognition, and parking assistance—with the ultimate goal of producing a fully autonomous vehicle. The present paper addresses detection and tracking of moving objects within the context of ADAS. We use a multisensor setup consisting of a radar and a stereo camera mounted on top of a vehicle. We propose to model the sensors uncertainty in polar coordinates on Lie Groups and perform the objects state filtering on Lie groups, specifically, on the product of two special Euclidean groups, i.e., $SE(2)^2$. To this end, we derive the designed filter within the framework of the extended Kalman filter on Lie groups. We assert that the proposed approach results with more accurate uncertainty modeling, since used sensors exhibit contrasting measurement uncertainty characteristics and the predicted target motions result with *banana-shaped* uncertainty contours. We believe that accurate uncertainty modeling is an important ADAS topic, especially when safety applications are concerned. To solve the multitarget tracking problem, we use the joint integrated probabilistic data association filter and present necessary modifications in order to use it on Lie groups. The proposed approach is tested on a real-world dataset collected with the described multisensor setup in urban traffic scenarios.

1. Introduction

Reliable comprehension of the surrounding environment, under varying conditions, is an essential task in nearly any assistance or autonomous system application. Since the advent of autonomous vehicle research, scientific community has been actively engaged in developing advanced driver assistance systems (ADAS). ADAS commonly involve adaptive cruise control, collision avoidance, lane change assistance, traffic sign recognition, and parking assistance—with the final goal being a fully autonomous vehicle. ADAS have been in the focus of research for a few decades, intended to enhance the safety and reduce the possibility of a human error as a cause of road accidents [1]. An essential task in numerous ADAS applications is the detection and tracking of moving objects (DATMO), since it allows the vehicle to be aware of dynamic objects in its immanent surrounding and predict their future behavior. Since the robustness of such an application under varying environmental conditions represents a complex challenge,

☆ This work has been supported by the Unity Through Knowledge Fund under the project *Cooperative Cloud based Simultaneous Localization and Mapping in Dynamic Environments* and the European Regional Development Fund under the project *Advanced Technologies in Power Plants and Rail Vehicles*.

* Corresponding author.
E-mail addresses: josip.cesic@fer.hr (J. Ćesić), ivan.markovic@fer.hr (I. Marković), igor.cvisic@fer.hr (I. Cvišić), ivan.petrovic@fer.hr (I. Petrović).

http://dx.doi.org/10.1016/j.robot.2016.05.001
0921-8890/

it has become clear that there does not exist such a sensing system that could solely deliver full information required for adequate quality of ADAS applications [2].

Given that, ADAS commonly rely on using complementary sensing systems: vision, millimeter-wave radars, laser range finder (LRF) or combinations thereof. Radar units are able to produce accurate measurements of the relative speed and distance to the objects. LRF have higher lateral resolution than the radars and, besides accurate object distance, they can detect the occupancy area of an object and provide detailed scene representation [3]. Regarding the robustness, radar units are more robust to rain, fog, snow, and similar conditions that may cause inconveniences for LRF; but, they produce significant amount of clutter as a drawback. Vision-based sensing systems can also provide accurate lateral measurements and wealth of other information from images, thus provide an effective supplement to ranging-based sensor road scene analysis. As an example, a stereo vision sensor can provide target detection with high lateral resolution and less certain range, while usually bringing enough information for identification and classification of objects, whereas radar can provide accurate measurements of range and relative speed. Given the complementarity of radars and vision systems, this combination is commonly used in research for ADAS applications. For example, works based on a monocular camera use radar for finding regions of interest in the image [4–7], process separately image and radar data [8–10], use motion stereo to reconstruct object boundaries [11,12], while [13,14] use directly stereo cameras. Employing multiple sensors, and consequently exploiting their different modalities, requires fusion of the sensing systems at appropriate levels. Depending on the approach, fusion can roughly take place at three levels: before objects detection (low level) [13, 14], at the objects' detection level (fused list of objects) [12,10], or at the state level (updating the states of objects in the list for each sensor system) [9,8,15].

Since in ADAS applications sensors with very different characteristics are used; e.g. radar with higher lateral uncertainty, but precise range estimation, and stereo camera with low lateral uncertainty but higher range imprecision, question arises on how to faithfully model the uncertainty of the state, estimated asynchronously with such sensors. Moreover, since in urban scenarios targets can exhibit varying dynamic behavior, a flexible motion model, capable of capturing the maneuvering diversity, should be used.

In the present paper, which is a continuation of our previous work presented in [16], we use a combination of a radar and a stereo vision system to perform the target tracking task. Our previous work focused on developing an appearance-based detection approach, while this paper deals with the tracking part of the DATMO procedure and uses a motion-based detection technique. Given the previous discussion, our first contribution is in modeling radar and stereo measurements arising in polar coordinates as members of Lie Groups $SO(2) \times \mathbb{R}^1$, and in estimating the target state as the product of two special Euclidean motion groups $SE(2) \times SE(2) = SE(2)^2$. This is performed within the framework of the extended Kalman filter on Lie groups, which we derive for the proposed system design. Furthermore, the target motion model also resides on the same group product and as such will yield the required model flexibility. This will not only enable us to correctly model sensor uncertainties, but also to have higher diversity in the uncertainty representation of the state estimates. For example, besides the standard Gaussian elliptically shaped uncertainty, proposed representation also supports the so called *banana-shaped* uncertainties. The second contribution of the paper is the adaptation of the joint integrated probabilistic data association (JIPDA) filter for multitarget on the $SE(2)^2$. To the best of the author's knowledge, this is the first use of a filtering on Lie Groups for a multitarget tracking application.

The rest of the paper is organized as follows. Section 2 presents related work and the present paper's contributions. Section 3 presents mathematical background of the LG-EKF, while Section 4 derives the proposed asynchronous LG-EKF on $SE(2)^2$ with polar measurements. The multitarget tracking with JIPDA filter on $SE(2)^2$ is described in Sections 5 and 6 presents the real-world experimental results. In the end, Section 7 concludes the paper.

2. Related work and progress beyond

Several distinct research fields relate to the study presented in this paper. These include the state estimation on Lie groups, multitarget tracking, stereo vision- and radar-based signal processing. We focus our overview of related work in the pertinent fields by considering results relevant to the present application.

To detect objects of interest, vision algorithms can resort to (i) appearances at a single time step, and (ii) motion over several frames [2]. In [17] authors employ detection procedure based on appearances in the disparity space, where clustering and extraction of moving objects are performed. The work in [18] focuses on ego-motion estimation, while moving objects stem from clustering the estimated motions in the filtered point cloud. Scene flow, i.e., the motion in 3D from stereo sequences, was used in [19,20], where adjacent points describing similar flow are considered to belong to a single rigid object. In [21] objects are also extracted from the scene flow, after which clustering is performed, and the iterative closest point algorithm is used to determine the vehicles' pose. Approach in [22] combines depth and optical flow-based clustering with an active learning-based method. In [23] pedestrians were isolated from the stereo point cloud and their pose estimated using a visibility-based 3D model, which is capable of predicting occlusions and using them in the detection process.

Concerning radar and stereo vision integration, in [14] approach based on fitting the model of a vehicle contour to both stereo depth image and radar readings was presented. First, the algorithm fits the contour from stereo depth information and finds the closest point of the contour with respect to the vision sensor. Second, it determines the closest point of the radar observation and fuses radar's and vision's closest points. By translating the initially fitted contour to the fused closest point, the resulting contour is obtained and located. Another low level integration approach was presented in [13]. In particular, the edge map of the stereo image is split into layers corresponding to different target depths so that the layers contain edge pixels of targets at different depth ranges. Hence, the original multitarget segmentation task is decomposed into several single target segmentation tasks on each depth-based layer, thus lowering the computational costs of the segmentation.

In the present paper each sensor reports its detections independently. To estimate the interim vehicle displacement, we use our visual stereo odometry algorithm (named SOFT) presented in [24]. Features not conforming to the computed displacement are considered as moving objects and are grouped together to yield measurements which are then fed to the tracking algorithm. In that respect our approach would fall within the motion-based detection approaches. The radar sensor complements detections from the stereo camera, and reports to the tracking algorithm a list of possible obstacle detections.

Irrespective of the used sensor setup, in traffic scenarios one must address the problem of multitarget tracking. This entails estimation (tracking) of each target's state and dealing with the problem of associating correct measurements to the tracked targets in cluttered environments, i.e. solving the data association problem. Commonly, for state estimation the Kalman filter and its non-linear variants are used. However, in order to achieve the proposed state uncertainty representation and motion model flexibility, in the present paper we use the extended Kalman filter