

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

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.

Keywords: advanced driver assistance systems, detection and tracking of moving objects, joint integrated probabilistic data association, radar, stereo camera

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, 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],

*Corresponding author: Josip Ćesić

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

Email address: josip.cesic@fer.hr, ivan.markovic@fer.hr, igor.cvisic@fer.hr, ivan.petrovic@fer.hr (Josip Ćesić, Ivan Marković, Igor Cvišić, Ivan Petrović)

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 manoeuvring 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 Section 5 and Section 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 pro-

cessing. 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 on Lie groups (LG-EKF) [25]. This way we can track targets with the Kalman filter directly on the $SE(2)^2$. Considering mul-

target tracking, a lot of attention has been devoted to tractable random finite sets (RFS)-based approximations of the multitarget Bayes filter: probability hypothesis density (PHD) [26–28], cardinalized PHD [29, 30], and multitarget multi-Bernoulli filters [31–34]. On the other hand, data association-based algorithms, such as multiple hypothesis tracker (MHT) [35] and joint probabilistic data association (JPDA) filter [36], approach the problem by considering explicit measurement-to-target associations. In [37] the JPDA was extended to include the probability of target existence in order to alleviate the assumption of the constant and known number of targets in the scene. The two approaches are not orthogonal; filters very similar to the JPDA and MHT can be derived from the RFS theory [38, 39].

Detection results often serve as inputs to the tracking algorithm and the ADAS works most similar to the present paper are [8, 9]. In [8], the authors fuse the data from radar and image sensor to estimate the position, direction and width of objects in front of the vehicle. Therein, an ego-motion compensated tracking approach is presented which combines radar observations with the results of the contour-based image processing algorithm. The filtering aspect relies on the unscented Kalman filter and the constant turn rate and acceleration model. In [9] authors propose asynchronous independent processing of radar and vision data and use the interacting multiple model Kalman filter to cope with the changing dynamics, associating the observations via probability data association scheme. In particular, the combined motion models are the constant velocity and constant acceleration models.

Since both the stereo camera and the radar work at different frequencies, we use asynchronous filtering; in that respect our approach performs fusion at the state level. We propose to model radar and stereo measurements in polar coordinates within the LG-EKF scheme and we derive the required filter on the product of special Euclidean groups, $SE(2)^2$. We also provide an in-depth discussion on the behavior of the state uncertainty when fusing measurements from the used sensors. We believe that faithful uncertainty representation is an important aspect of ADAS, especially when safety applications are concerned. To handle varying dynamic behavior, our motion model will reside on $SE(2)^2$, since it can capture well a wide range of behavior [40]. To handle the multitarget scenario, we propose to use the JPDA filter, which, to the best of the authors' knowledge, is its first use within the Kalman filtering on Lie groups. The proposed approach is validated in real-life experiments, where the dataset was taken in urban scenarios with the sensor setup mounted on a moving vehicle (Fig. 1).

3. Mathematical preliminaries

3.1. Lie groups, Lie algebra and the concentrated Gaussian distribution

In this section, we provide notations and properties for matrix Lie groups and the associated Lie algebras which will be used for the $SE(2)^2$ filter. Lie group G is a group which has the structure of a smooth manifold (i.e. it is sufficiently often differentiable [41]) such that group composition and inversion are

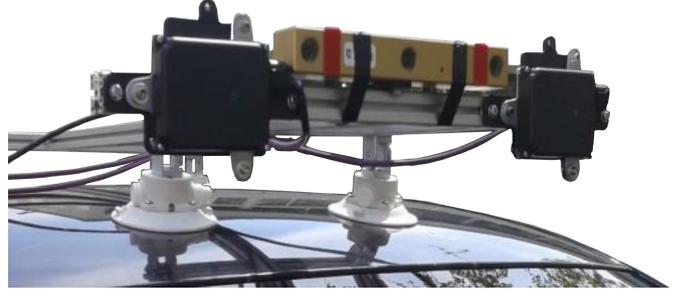


Figure 1: The experimental platform mounted on top of a vehicle, consisting of a stereo camera system and two radar units

smooth operations. Furthermore, for a matrix Lie group G , of which $SE(2)$ is an example, these operations are simply matrix multiplication and inversion, with the identity matrix $I^{n \times n}$ being the identity element [42].

Another important term is the Lie algebra \mathfrak{g} which is associated to a Lie group G . It is an open neighbourhood of $\mathbf{0}^{n \times n}$ in the tangent space of G at the identity $I^{n \times n}$. The matrix exponential \exp_G and matrix logarithm \log_G establish a local diffeomorphism

$$\exp_G : \mathfrak{g} \rightarrow G \text{ and } \log_G : G \rightarrow \mathfrak{g}. \quad (1)$$

The Lie algebra \mathfrak{g} associated to a p -dimensional matrix Lie group $G \subset \mathbb{R}^{n \times n}$ is a p -dimensional vector space defined by a basis consisting of p real matrices E_i , $i = 1, \dots, p$ [43]. A linear isomorphism between \mathfrak{g} and \mathbb{R}^p is given by

$$[\cdot]_G^\vee : \mathfrak{g} \rightarrow \mathbb{R}^p \text{ and } [\cdot]_G^\wedge : \mathbb{R}^p \rightarrow \mathfrak{g}. \quad (2)$$

Lie groups are generally non-commutative and require the use of two operators which enable the adjoint representation of (i) G on \mathbb{R}^p denoted as Ad_G and (ii) \mathbb{R}^p on \mathbb{R}^p denoted as ad_G [42, 44].

In order to define the concept of the concentrated Gaussian distribution on Lie groups, necessary for introduction of the LG-EKF, the considered Lie group needs to be a connected unimodular matrix Lie group [45], which is the case of the majority of Lie groups used in robotics.

Let the pdf of X be defined as [46]

$$p(X) = \beta \exp\left(-\frac{1}{2} [\log_G(X)]_G^\vee{}^T P^{-1} [\log_G(X)]_G^\vee\right), \quad (3)$$

where β is a normalizing constant. Let ϵ be defined as $\epsilon \triangleq [\log_G(X)]_G^\vee$. Under the assumption that the entire mass of probability is contained inside G , i.e., $\int_{\mathbb{R}^{n \times n} \setminus G} p(X) = 0$, ϵ can be described with $\epsilon \sim \mathcal{N}_{\mathbb{R}^p}(\mathbf{0}^{p \times 1}, P)$. This concept is called a concentrated Gaussian distribution (CGD) on G around the identity [25]. Furthermore, it is a unique parametrization space where the bijection between \exp_G and \log_G exists. Now, the distribution of X can be translated over G by using left action of the Lie group

$$X = \mu \exp_G\left([\epsilon]_G^\wedge\right), \text{ with } X \sim \mathcal{G}(\mu, P), \quad (4)$$

where \mathcal{G} denotes the CGD [46, 25]. By this, we have introduced the distribution forming the base for the LG-EKF.

3.2. The Special Euclidean group SE(2)

The group SE(2) describes rigid body motion in 2D and is formed as a semi-direct product of the plane \mathbb{R}^2 and the special orthogonal group SO(2) corresponding to translational and rotational elements. It is defined as

$$\text{SE}(2) = \left\{ \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}_{1 \times 2} & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \mid \{R, \mathbf{t}\} \in \text{SO}(2) \times \mathbb{R}^2 \right\}. \quad (5)$$

Now, we continue with providing the basic ingredients for working with SE(2), giving relations for operators presented in Section 3.1, needed for manipulations within the triplet: Lie group G, Lie algebra \mathfrak{g} , and Euclidean space \mathbb{R}^p .

For a Euclidean space vector $\mathbf{x} = [x \ y \ \theta]^T$, the most often associated element of the Lie algebra $\mathfrak{se}(2)$ is given as

$$[\mathbf{x}]_{\text{SE}(2)}^\wedge = \begin{bmatrix} [\mathbf{x}]_{\text{SO}(2)}^\wedge & x \\ \mathbf{0}_{1 \times 2} & y \\ & 0 \end{bmatrix} \in \mathfrak{se}(2) \quad (6)$$

$$[\mathbf{x}]_{\text{SO}(2)}^\wedge = \begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix} \in \mathfrak{so}(2). \quad (7)$$

Their inverses, $[\cdot]_{\text{SE}(2)}^\vee$ and $[\cdot]_{\text{SO}(2)}^\vee$, follow trivially from the relations (6) and (7), respectively.

The exponential map for the SE(2) group is given as

$$\exp_{\text{SE}(2)}([\mathbf{x}]_{\text{SE}(2)}^\wedge) = \begin{bmatrix} \exp_{\text{SO}(2)}([\theta]_{\text{SO}(2)}^\wedge) & t_x \\ \mathbf{0}_{1 \times 2} & t_y \\ & 1 \end{bmatrix} \in \text{SE}(2) \quad (8)$$

$$\exp_{\text{SO}(2)}([\theta]_{\text{SO}(2)}^\wedge) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \in \text{SO}(2) \quad (9)$$

$$t_x = \frac{1}{\theta} [x \sin \theta + y(-1 + \cos \theta)] \quad (10)$$

$$t_y = \frac{1}{\theta} [x(1 - \cos \theta) + y \sin \theta]. \quad (11)$$

For $T = \{R, \mathbf{t}\} \in \text{SE}(2)$, the logarithmic map is

$$\log_{\text{SE}(2)}(T) = \begin{bmatrix} \mathbf{v} \\ \theta \end{bmatrix}_{\text{SE}(2)}^\wedge \in \mathfrak{se}(2) \quad (12)$$

$$\theta = \log_{\text{SO}(2)}(R) = \text{atan2}(R_{21}, R_{11}) \quad (13)$$

$$\mathbf{v} = \frac{\theta}{2(1 - \cos \theta)} \begin{bmatrix} \sin \theta & 1 - \cos \theta \\ \cos \theta - 1 & \sin \theta \end{bmatrix} \mathbf{t}. \quad (14)$$

The Adjoint operator Ad_G used for representing $T \in \text{SE}(2)$ on \mathbb{R}^3 is given as

$$\text{Ad}_{\text{SE}(2)}(T) = \begin{bmatrix} R & J\mathbf{t} \\ \mathbf{0}_{1 \times 2} & 1 \end{bmatrix} \text{ with } J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad (15)$$

while the adjoint operator ad_G for representing $\mathbf{x} \in \mathbb{R}^3$ on \mathbb{R}^3 is given by

$$\text{ad}_{\text{SE}(2)}(\mathbf{x}) = \begin{bmatrix} -\theta J & J\mathbf{v} \\ \mathbf{0}_{1 \times 2} & 1 \end{bmatrix}, \quad (16)$$

where $\mathbf{v} = [x \ y]^T \in \mathbb{R}^2$. Given the definitions above, we have all the needed ingredients for using the SE(2) motion group within the proposed approach.

4. Second order rigid body motion estimation

4.1. State space construction

As a rigid body, vehicle's state can be well described employing the rigid body motion group. Furthermore, when considering velocities of such an object, we can also represent these higher order moments by using the same motion group. Following the rigid body equivalent of the constant velocity motion model [47], here we model the vehicle by constructing the state space G as the Cartesian (direct) product of the two matrix Lie group SE(2) members [40]

$$\text{SE}(2) \times \text{SE}(2) = \text{SE}(2)^2. \quad (17)$$

The first SE(2) member is the position component, while the second one contributes the velocity components. This can be regarded as a white noise acceleration model [47] on the SE(2) group. Considering vehicle tracking applications, in contrast to other well established motion models—constant velocity, constant turn rate and velocity, constant curvature and velocity [48, 49]—the SE(2)² motion model provides more artificial flexibility. This flexibility is manifested through including the holonomic behavior over all three velocity components, i.e., the longitudinal, lateral, and rotational velocities, which have Wiener process characterization [47]. Such flexibility provides the ability to describe motion of objects appearing in ADAS, e.g., vehicles, motorcycles and pedestrians, and hence is appropriate for usage in our particular DATMO focused application.

Matrix Lie group composition and inversion are simple matrix multiplication and inversion, hence for all the calculations dealing with operations on G, we can use the symbolic representation constructed by placing the two SE(2) members of G block diagonally. The Lie algebra associated to the Lie group G is denoted as $\mathfrak{g} = \mathfrak{se}(2) \times \mathfrak{se}(2)$. The term $[\mathbf{x}]_G^\wedge$ is also constructed by placing both $\mathfrak{se}(2)$ members on the main diagonal, and correspondingly the exponential map on such G is as well formed block diagonally. For more details on the construction and symbolical representation of the groups of interest, please confer [40] where the state model was first proposed.

4.2. Motion model and prediction

The motion model satisfies the following equation

$$X_{k+1} = f(X_k, n_k) = X_k \exp_G([\hat{\Omega}_k + n_k]_G^\wedge), \quad (18)$$

where $X_k \in G$ is the state of the system at time k , G is a p -dimensional Lie group, $n_k \sim \mathcal{N}_{\mathbb{R}^p}(\mathbf{0}^{p \times 1}, Q_k)$ is white Gaussian noise and $\hat{\Omega}_k = \Omega(X_k) : G \rightarrow \mathbb{R}^p$ is a non-linear C^2 function. If the posterior distribution at step $k - 1$ follows the concentrated Gaussian distribution on matrix Lie Groups as $\mathcal{G}(\mu_{k-1}, P_{k-1})$. The predicted mean is given by [25]

$$\mu_{k+1|k} = \mu_k \exp_G([\hat{\Omega}_k]_G^\wedge). \quad (19)$$

We model the motion (18) by [40]

$$\Omega(X_k) = [T v_{x_k} \ T v_{y_k} \ T \omega_k \ 0 \ 0 \ 0]^T \in \mathbb{R}^6, \quad (20)$$

$$n_k = \left[\frac{T^2}{2} n_{x_k} \ \frac{T^2}{2} n_{y_k} \ \frac{T^2}{2} n_{\omega_k} \ T n_{x_k} \ T n_{y_k} \ T n_{\omega_k} \right]^T \in \mathbb{R}^6.$$

With this model, the system is corrupted with white noise over three components, i.e. n_x is the noise in the local x direction, n_y is the noise in local y direction and n_w is the noise in the rotational component.

Formula for propagating the covariance of $\epsilon_{k+1|k}$ through the general motion model (18) is given as in [25]

$$P_{k+1|k} = \mathcal{F}_k P_k \mathcal{F}_k^T + \Phi_G(\hat{\Omega}_k) Q_k \Phi_G(\hat{\Omega}_k)^T, \quad (21)$$

where the operator \mathcal{F}_k , a matrix Lie group equivalent to the Jacobian of $f(X_k, n_k)$, and Φ_G , are evaluated as

$$\begin{aligned} \mathcal{F}_k &= \text{Ad}_G \left(\exp_G \left([-\hat{\Omega}_k]_G^\wedge \right) \right) + \Phi_G(\hat{\Omega}_k) \mathcal{C}_k \\ \Phi_G(\mathbf{v}) &= \sum_{m=0}^{\infty} \frac{(-1)^m}{(m+1)!} \text{ad}_G(\mathbf{v})^m, \quad \mathbf{v} \in \mathbb{R}^p \\ \mathcal{C}_k &= \frac{\partial}{\partial \epsilon} \Omega \left(\mu_k \exp_G \left([\epsilon]_G^\wedge \right) \right) \Big|_{\epsilon=0}. \end{aligned} \quad (22)$$

The covariance propagation is challenging since it requires calculation of (22). The final expression for \mathcal{C}_k is thus given as

$$\mathcal{C}_k = \begin{bmatrix} \mathbf{0}^{3 \times 3} & T \cos \omega_k & -T \sin \omega_k & 0 \\ T \sin \omega_k & T \cos \omega_k & 0 & 0 \\ 0 & 0 & 0 & T \\ \mathbf{0}^{3 \times 3} & 0 & 0 & \mathbf{0}^{3 \times 3} \end{bmatrix}. \quad (23)$$

The complete derivation of \mathcal{C}_k is given in [40]. The adjoint operators Ad_G and ad_G are also formed block diagonally.

The last needed ingredient is the process noise covariance matrix Q_k . In the present paper, we perform sensor fusion in an asynchronous manner with the arrival of each measurement. Hence, we proceed by defining the process to follow continuous white noise acceleration model (CWNA) over the three components discussed previously. In the sequel, we derive the discrete time process noise by relating it to the continuous one [47]. Let \tilde{q}_x , \tilde{q}_y and \tilde{q}_ω denote the time-invariant continuous time process noise intensities reflecting power spectral density over all three components. Then, the process noise covariance matrix Q evaluates to

$$Q = \begin{bmatrix} \frac{T^3}{3} \tilde{q}_x & 0 & 0 & \frac{T^2}{2} \tilde{q}_x & 0 & 0 \\ 0 & \frac{T^3}{3} \tilde{q}_y & 0 & 0 & \frac{T^2}{2} \tilde{q}_y & 0 \\ 0 & 0 & \frac{T^3}{3} \tilde{q}_\omega & 0 & 0 & \frac{T^2}{2} \tilde{q}_\omega \\ \frac{T^2}{2} \tilde{q}_x & 0 & 0 & T \tilde{q}_x & 0 & 0 \\ 0 & \frac{T^2}{2} \tilde{q}_y & 0 & 0 & T \tilde{q}_y & 0 \\ 0 & 0 & \frac{T^2}{2} \tilde{q}_\omega & 0 & 0 & T \tilde{q}_\omega \end{bmatrix}, \quad (24)$$

At this point, we have defined all the necessary ingredients for the asynchronous prediction step of the LG-EKF filter.

4.3. Measurement model and correction

The discrete measurement model on the matrix Lie group is defined as

$$Z_{k+1} = h(X_{k+1}) \exp_{G'} \left([m_{k+1}]_{G'}^\wedge \right), \quad (25)$$

where $Z_{k+1} \in G'$, $h : G \rightarrow G'$ is a C^1 function and $m_{k+1} \sim \mathcal{N}_{\mathbb{R}^q}(\mathbf{0}^{q \times 1}, R_{k+1})$ is white Gaussian noise.

The predicted system state is described with $X_{k+1|k} \sim \mathcal{G}(\mu_{k+1|k}, P_{k+1|k})$ and now we proceed to updating the state by incorporating the newly arrived measurement $Z_{k+1} \in G'$. In this case, we propose the measurements to arise in the space of a Lie group constructed as $G' = \text{SO}(2) \times \mathbb{R}^1$, measuring the current position of the tracked object in 2D in polar coordinates. The radar and the stereo camera, as well as many other widely spread on-board sensing systems, perceive the surrounding from a single point, and hence perform the measurement in polar coordinates. Thus the uncertainty of such measurements, i.e. the measurement likelihood, resembles *banana-shaped* contours rather than the elliptical ones. In order to integrate such sensing modalities into the LG-EKF, we now introduce necessary ingredients for the update step of the filter.

The measurement function is mapping $h : \text{SE}(2) \times \text{SE}(2) \rightarrow \text{SO}(2) \times \mathbb{R}^1$. It is given as

$$h(X_{k+1}) = \begin{bmatrix} \exp_{\text{SO}(2)} \left(\left[\arctan \frac{y_{k+1}}{x_{k+1}} \right]_{\text{SO}(2)}^\wedge \right) \\ \exp_{\mathbb{R}^1} \left(\left[\sqrt{x_{k+1}^2 + y_{k+1}^2} \right]_{\mathbb{R}^1}^\wedge \right) \end{bmatrix}. \quad (26)$$

The exponential and logarithm on \mathbb{R}^p follows a mapping procedure and is only a matter of representation. Hence we introduce $\exp_{\mathbb{R}}$ for implementation purposes only, to follow the matrix representation of the procedure, hence each composition and inversion follow matrix multiplication and inversion procedures, even when working with Euclidean space. In particular, the Euclidean space is a trivial example of a matrix Lie group, so the representation of $\mathbf{v} \in \mathbb{R}^p$ in the form of a Lie algebra $[\mathbf{v}]_{\mathbb{R}^p}^\wedge \subset \mathbb{R}^{p+1 \times p+1}$ and matrix Lie group $\exp_{\mathbb{R}^p}([\mathbf{v}]_{\mathbb{R}^p}^\wedge) \subset \mathbb{R}^{p+1 \times p+1}$ is given as

$$[\mathbf{v}]_{\mathbb{R}^p}^\wedge = \begin{bmatrix} \mathbf{0}^{p \times p} & \mathbf{v} \\ \mathbf{0}^{1 \times p} & 0 \end{bmatrix} \text{ and } \exp_{\mathbb{R}^p}([\mathbf{v}]_{\mathbb{R}^p}^\wedge) = \begin{bmatrix} \mathbf{I}^{p \times p} & \mathbf{v} \\ \mathbf{0}^{1 \times p} & 1 \end{bmatrix}. \quad (27)$$

One should note that there exists a trivial mapping between the members of the triplet \mathbf{v} , $[\mathbf{v}]_{\mathbb{R}^p}^\wedge$ and $\exp_{\mathbb{R}^p}([\mathbf{v}]_{\mathbb{R}^p}^\wedge)$, hence the formal inverses of the terms from (27) are omitted here.

Lets now define the following innovation term

$$\begin{aligned} \tilde{Z}_{k+1} &= \left[\log_{G'} \left(h(\mu_{k+1|k})^{-1} Z_{k+1} \right) \right]_{G'}^\vee \\ &= \mathcal{H}_{k+1} \epsilon_{k+1|k} + m_{k+1} + \mathcal{O} \left(\|\epsilon_{k+1|k}\|^2, \|m_{k+1}\|^2 \right) \end{aligned} \quad (28)$$

which is linear in the lie algebraic error $\epsilon_{k+1|k} \sim \mathcal{N}_{\mathbb{R}^p}(\mathbf{0}^{p \times 1}, P_{k+1|k})$. Now, we can apply the classical update equations employing the measurement model to update the Lie algebraic mean and covariance, such that $\epsilon_{k+1}^- \sim \mathcal{N}_{\mathbb{R}^p}(\mu_{k+1}^-, P_{k+1}^-)$. The update step of the filter, based on the measurement model (25), strongly resembles the standard EKF update procedure [48], relying on the Kalman gain K_{k+1} and innovation vector \mathbf{v}_{k+1} , and is calculated as

$$\begin{aligned} K_{k+1} &= P_{k+1|k} \mathcal{H}_{k+1}^T \left(\mathcal{H}_{k+1} P_{k+1|k} \mathcal{H}_{k+1}^T + R_{k+1} \right)^{-1} \\ \mathbf{v}_{k+1} &= \left[\log_{G'} \left(h(\mu_{k+1|k})^{-1} Z_{k+1} \right) \right]_{G'}^\vee. \end{aligned} \quad (29)$$

Hence the updated Lie algebraic error ϵ_{k+1}^- is given as

$$\begin{aligned} \mu_{k+1}^- &= K_{k+1} \mathbf{v}_{k+1} \\ P_{k+1}^- &= (\mathbf{I}^{p \times p} - K_{k+1} \mathcal{H}_{k+1}) P_{k+1|k}. \end{aligned} \quad (30)$$

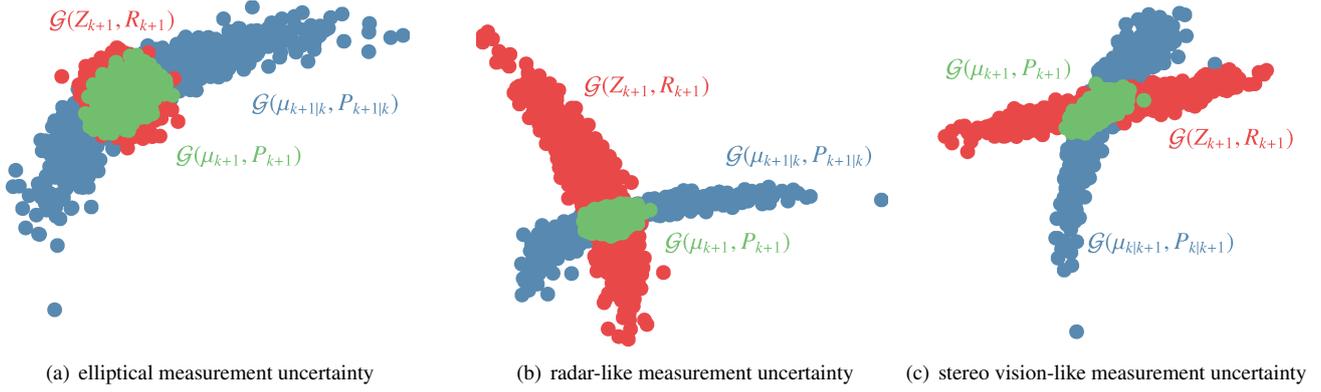


Figure 2: Examples of LG-EKF state uncertainties when updated with sensors having different measurement characteristics. The filter prediction in blue follows the $SE(2)^2$ motion model, measurement is depicted in red, and the updated state is depicted in green. We can notice that the LG-EKF filter can capture a wide range of uncertainty contours; from Gaussian elliptic uncertainties to *banana-shaped* uncertainties typical for range-bearing sensors and vehicles in motion with non-zero turn rate deviation.

The matrix \mathcal{H}_k can be seen as the measurement matrix of the system, i.e. a matrix Lie group equivalent to the Jacobian of $h(X_k)$, and is given as

$$\mathcal{H}_{k+1} = \frac{\partial}{\partial \epsilon} \left[\log_{G'} \left(h(\mu_{k+1|k})^{-1} h(\mu_{k+1|k} \exp_G([\epsilon]_G^\wedge)) \right) \right]_{G|\epsilon=0}^\vee. \quad (31)$$

The final expression of the measurement matrix \mathcal{H}_{k+1} is given as follows

$$\mathcal{H}_{k+1} = \begin{bmatrix} \frac{-y \cos \theta + x \sin \theta}{x^2 + y^2} & \frac{x \cos \theta + y \sin \theta}{x^2 + y^2} & & \\ \frac{x \cos \theta + y \sin \theta}{\sqrt{x^2 + y^2}} & \frac{y \cos \theta - x \sin \theta}{\sqrt{x^2 + y^2}} & \mathbf{0}^{2 \times 4} & \end{bmatrix}. \quad (32)$$

Note that the subscript indices determining the time step of the filter have been omitted in the previous expression due to clarity, i.e. $\theta_{k+1|k} \triangleq \theta$, $x_{k+1|k} \triangleq x$ and $y_{k+1|k} \triangleq y$. Detailed derivation of the matrix \mathcal{H}_{k+1} is given in the Appendix.

The update procedure is expected to deliver the concentrated Gaussian distribution such that $X = \mu \exp_G([\epsilon]_G^\wedge)$, with expectation $\mathbb{E}[\epsilon] = 0^{p \times 1}$. However, since operating with generally non-Euclidean spaces, we have $\mathbb{E}[\epsilon_{k+1}^-] = \mu_{k+1}^- \neq 0^{p \times 1}$ which is resolved by state reparametrization [25]. The mean and the covariance are updated as

$$\begin{aligned} \mu_{k+1} &= \mu_{k+1|k} \exp_G([\mu_{k+1|k}^-]_G^\wedge) \\ P_{k+1} &= \Phi_G(\mu_{k+1}^-) P_{k+1|k}^- \Phi_G(\mu_{k+1}^-)^T. \end{aligned} \quad (33)$$

As in the case of the prediction step, the state $X_{k+1} \sim \mathcal{G}(\mu_{k+1}, P_{k+1})$ has remained \mathcal{G} -distributed after correction. Now we have all the means for updating the filter by calculating the Kalman gain K_{k+1} and the innovation vector ν_{k+1} (29), and finally correcting the mean μ_{k+1} and the covariance matrix P_{k+1} (33).

Figure 2 shows examples of LG-EKF filter state uncertainties updated with three different sensors types. In all the examples

the filter prediction follows the $SE(2)^2$ motion model and yields *banana-shaped* state uncertainties. In Fig. 2(a) we show an example of updating the filter with a sensor having elliptical measurement uncertainty; this resembles ‘classical’ Gaussian like uncertainty. In Fig. 2(b) we depict update with a sensor that has larger uncertainty in the bearing than in the range and the update of the filter acts as ‘intersecting’ the two *banana-shaped* distributions. This example resembles update performed with a radar unit. Finally, Fig. 2(c) shows the example with sensor having larger uncertainty in the range than in the bearing. Notice how the prediction uncertainty skews to the right indicating that the vehicle had higher probability of turning right than left. This example resembles update performed with the stereo vision sensor. From the above examples we can see how the filter can handle diverse measurement uncertainties and efficiently fuse them with the information from the prediction step. Having finished with the single target filtering, what is left is to resolve the LG-EKF tracking with multiple targets in the scene.

5. Joint Integrated Probabilistic Data Association

Assume that we are tracking multiple targets, $\{\mathcal{T}_1, \dots, \mathcal{T}_{l_k}\}$, with the number of targets, l_k , varying with time, i.e., targets can appear and disappear from the sensors’ field-of-view. Let Z_k denote the set of all measurements at time step k

$$Z_k = \{Z_k^j : j = 1, \dots, m_k\},$$

and $Z_{1:k} = \{Z_1, \dots, Z_k\}$ the history of all the measurements. The vector Z_k , besides target originating measurements, also contains clutter which is a Poisson distributed random variable. The main issue at hand is how to appropriately assign the received measurement set to the targets in track, and how to manage the target appearance and disappearance.

The JIPDA [37] approaches this problem by estimating the following a posteriori density for each \mathcal{T}_i

$$p(X_k^i, \chi_k^i | Z_{1:k}) = p(X_k^i | \chi_k^i, Z_{1:k}) p(\chi_k^i | Z_{1:k}), \quad (34)$$

i.e., the density of the target's state X_k^i and its existence χ_k^i given all the measurements up to and including k . Note that in the present paper, X_k^i is distributed according to $\mathcal{G}(\mu_k^i, P_k^i)$ as in the case of (4). For the probability of target existence, we adopt the Markov Chain One model [37]

$$p(\chi_k^i | Z_{1:k-1}) = p_S p(\chi_{k-1}^i | Z_{1:k-1}), \quad (35)$$

where p_S denotes the probability that target will continue to exist at step k given that it existed at step $k-1$.

In order to alleviate computational complexity, at each scan tracks are separated into clusters which share selected measurements. As a criteria for measurement-to-track validation, the gating principle is used where based on the innovation uncertainty (29) a gating volume is defined, and measurements falling within are accepted as cluster members. For notation clarity we will not differentiate measurements belonging to the cluster from those outside of the clusters. The former will participate in the data association operations, while the latter will be treated as candidates for new tracks initialization. For filtering on LG, validation gate is defined in the algebra where measurements are associated to targets, and if multiple targets share the same measurements they are formed into a cluster. The ensuing formulae will pertain to a single cluster and all the measurements and targets are assumed to belong to the cluster.

Upon availability of a set of new measurements $Z_k = \{Z_k^j : j = 1, \dots, m_k\}$, the following set of hypotheses is built:

$$\begin{aligned} \theta_k^{ij} &= \{Z_k^j \text{ is caused by } \mathcal{T}_i\}, \quad j = 1, \dots, m_k, \text{ and} \\ \theta_k^{i0} &= \{\text{none of the measurements is caused by } \mathcal{T}_i\}. \end{aligned}$$

The total probability formula implies that the posterior density for object \mathcal{T}_i at scan k is given by [37]

$$\begin{aligned} p(X_k^i, \chi_k^i | Z_{1:k}) &= p(\chi_k^i | Z_{1:k}) \sum_{j=0}^{m_k} p(X_k^i | \theta_k^{ij}, \chi_k^i, Z_{1:k}) p(\theta_k^{ij} | \chi_k^i, Z_{1:k}) \\ &= p(\chi_k^i | Z_{1:k}) \sum_{j=0}^{m_k} \beta_k^{ij} p(X_k^i | \theta_k^{ij}, \chi_k^i, Z_{1:k}), \end{aligned} \quad (36)$$

where $\beta_k^{ij} = p(\theta_k^{ij} | \chi_k^i, Z_{1:k})$ represent a posteriori *data association probabilities* conditioned on object existence. Explicitly, β_k^{ij} is the probability that measurement z_k^j is caused by \mathcal{T}_i and β_k^{i0} is the probability that none of the measurements is caused by \mathcal{T}_i . The densities $p(X_k^i | \theta_k^{ij}, \chi_k^i, Z_{1:k})$ represent 'classically' updated LG-EKF (30) for $j = 1, \dots, m_k$, while for $j = 0$ the density is just the prediction calculated via (18) and (21). Parameters of the mixture components are denoted by $\mu_{k+1}^{ij,-}$ and $P_{k+1}^{ij,-}$, specifically, when $j = 0$, $\mu_{k+1}^{i0,-} = \mu_{k+1|k}^i$ and $P_{k+1}^{i0,-} = P_{k+1|k}^i$.

In order to calculate β_k^{ij} we need to take into account measurement-to-object associations events jointly across the set of objects in the cluster. This means that hypothesis θ_k^{ij} consists of all *feasible joint events* \mathcal{E} where each track is assigned zero or one measurement and each measurement is allocated to zero

or one track; thus, they partition the hypothesis θ_k^{ij} and

$$p(\theta_k^{ij}, \chi_k^i | Z^{1:k}) = \sum_{\mathcal{E} \in \theta_k^{ij}} P(\mathcal{E} | Z^{1:k}), \quad j > 1, \quad (37)$$

$$p(\theta_k^{i0} | Z^{1:k}) = 1 - p(\theta_k^{ij}, \chi_k^i | Z^{1:k}). \quad (38)$$

Furthermore, probability that \mathcal{T}_i exists and that no measurement in the cluster is its detection, is given by [37]

$$p(\theta_k^{i0}, \chi_k^i | Z^{1:k}) = \frac{(1 - P_D^i P_G^i) p(\chi_k^i | Z^{1:k})}{1 - P_D^i P_G^i p(\chi_k^i | Z^{1:k})} p(\theta_k^{i0} | Z^{1:k}). \quad (39)$$

To calculate $P(\mathcal{E} | Z^{1:k})$, for each joint event \mathcal{E} we define: set of targets allocated no measurement, $T_0(\mathcal{E})$, and set of tracks allocated one measurement, $T_1(\mathcal{E})$. Following [37, 38] we obtain

$$\begin{aligned} P(\mathcal{E} | Z^{1:k}) &= C_k^{-1} \prod_{i \in T_0(\mathcal{E})} (1 - P_D^i P_G^i P(\chi_k^i | Z^{1:k-1})) \\ &\quad \cdot \prod_{i \in T_1(\mathcal{E})} P_D^i P_G^i P(\chi_k^i | Z^{1:k-1}) \frac{p_k^i(\tau(\mathcal{E}, i))}{\rho_k(\tau(\mathcal{E}, i))}, \end{aligned} \quad (40)$$

where P_D^i is the probability of \mathcal{T}_i being detected, P_G^i is the probability that the correct measurement will be inside the validation gate of \mathcal{T}_i , $\tau(\mathcal{E}, i)$ is the index of measurement allocated to \mathcal{T}_i under joint event \mathcal{E} , $\rho_k(\tau(\mathcal{E}, i))$ denotes a priori clutter measurement density at $z_k^{\tau(\mathcal{E}, i)}$, and C_k^{-1} is the normalization constant calculated from the fact that \mathcal{E} are mutually exclusive and form an exhaustive set, i.e., $\sum_{\mathcal{E}} P(\mathcal{E} | Z^{1:k}) = 1$. The innovation is calculated by using results from (29)

$$p_k^i(\tau(\mathcal{E}, i)) = \frac{1}{P_G^i} p_k^i(v_k^{\tau(\mathcal{E}, i)}; 0, \mathcal{H}_{k+1} P_{k+1|k} \mathcal{H}_{k+1}^T + R_{k+1}). \quad (41)$$

The innovation in (41) is normalized by P_G^i in order to account for the validation gating, i.e., since it is truncated to integrate to unity. Finally, we have all the elements to determine the probability of target existence

$$p(\chi_k^i | Z^{1:k}) = \sum_{j=0}^{m_k} p(\theta_k^{ij}, \chi_k^i | Z^{1:k}), \quad (42)$$

and to calculate the data association probabilities

$$\beta_k^{ij} = \frac{p(\theta_k^{ij}, \chi_k^i | Z^{1:k})}{p(\chi_k^i | Z^{1:k})}, \quad j = 0, \dots, m_k. \quad (43)$$

Note that all the operations concerning a specific target \mathcal{T}_i , described so far in the section, are carried out in the pertaining algebra of $\mu_{k+1|k}^i$, since, we are still at the update stage of the LG-EKF. To calculate the final a posteriori state estimate the JIPDA logic dictates reducing the mixture in (36) to a single density with the following parameters [50, 38]

$$\mu_{k+1}^{i,-} = \sum_{j=0}^{m_k} \beta_k^{ij} \mu_{k+1}^{ij,-}, \quad (44)$$

$$P_{k+1}^{i,-} = \sum_{j=0}^{m_k} (P_{k+1}^{ij,-} + \mu_{k+1}^{ij,-} (\mu_{k+1}^{ij,-})^T) - \mu_{k+1}^{i,-} (\mu_{k+1}^{i,-})^T. \quad (45)$$

As in the case of the LG-EKF update, $\mathbb{E}[\epsilon_{k+1}^-] = \mu_{k+1}^{i,-} \neq 0^{p \times 1}$; thus, before mapping the updated state and covariance to G we have to perform reparametrization [40]

$$\begin{aligned} \mu_{k+1}^i &= \mu_{k+1|k}^i \exp_G([\mu_{k+1}^{i,-}]_G^\wedge) \\ P_{k+1}^i &= \Phi_G(\mu_{k+1}^{i,-}) P_{k+1}^{i,-} \Phi_G(\mu_{k+1}^{i,-})^T. \end{aligned} \quad (46)$$

6. Experimental results

6.1. System overview

The experiments were carried out using two radar units and a stereo camera system, mounted on a sensor platform on top of a vehicle. The sensor platform was constructed so that the stereo camera is placed in-between the two radar units as shown in Fig. 1.

In the present paper we used the Continental Short Range Radar 209 – 2 units (measurement range of 50 m) configured to operate in the cluster mode, at a rate of 15 Hz. The field of view is 150° horizontally and 12° vertically, with the resolution in the horizontal direction of 1° , while there is no discrimination of the angle in vertical direction, and hence the radar cluster data can be considered as 2D measurements. After each scanning cycle the radar can deliver a cluster consisting of up to 128 detections. In the prefiltering stage we dismissed all the cluster measurements whose radar cross section, i.e., the measure of the reflective strength, did not exceed -5 dBm.

The stereo images were recorded with the monochrome Point Grey Bumblebee XB3 camera system. This system is a 3-sensor multi-baseline stereo camera with 1.3 mega-pixel global shutter sensors. The image resolution is 1280×960 pixels, with horizontal field of view of 66° . The experiments were carried out at the maximum frame rate of 16 Hz, and by using the largest, 24 cm long baseline, since the expected target measurement range is up to 50 m. The stereo image synchronization was executed internally, while the experiment was recorded in the auto-exposure mode of the camera.

Given that the sensors are closely spaced, mechanically aligned using custom-made plates on the same rail, and since we perform sensor fusion at the state level, the inter-sensor calibration was done by measuring the mounting position displacements by hand. Moreover, due to the coarse nature of radar measurements we find the current rail-mounting sufficiently precise to assert that differences in the orientation of the sensor coordinate frames can be neglected for case of the present sensor setup. However, for arbitrary radar and stereo vision setups a closer inspection of the calibration problem might be required [7]. Furthermore, special attention was taken to assure the clock synchronization, since our approach relies on state estimation performed in an asynchronous manner. Although both sensors work at close frequencies, generally this might not be the case, and the approach of asynchronous filtering is kept for the sake of generality. The prediction step directly depends on the time period T , i.e., the time passed between the two consecutive steps k and $k + 1$. Therefore, a clock drift or large delay in data acquisition could significantly affect the performance of the algorithm.

6.2. Stereo detection procedure

The main goal of the stereo image processing part of the algorithm is to detect moving objects in the scene, while the motion of the observer makes this task especially challenging. However, this work focuses on the estimation procedure and the fusion of two sensor modalities, hence the sole stereo based detection of moving objects is only briefly described.

The first part of the algorithm works on the ego-motion estimation, which results in transformation matrix between the previous and the current camera frame. Regarding this issue, we employed our SOFT algorithm [24], which has proven to be very robust on the appearance of moving objects in the scene, illumination changes, various specularities, sensor overexposure etc. However, SOFT uses very sparse set of salient feature points, which are not sufficient to reliably detect objects in the scene. Therefore, we employ the corner detector from [51] for detection of semi-dense set of feature points. Position and velocity of each detected feature is estimated in 3D Euclidean space. Now we need to determine the correspondences between features in the left and the right image of the current and previous frame. For this purpose we have used the optical flow procedure presented in [52], and have computed the correspondences by using the stereo block matching algorithm from [53].

Since the images are rectified, all the feature points from the previous frame are projected into 3D world frame through a standard pinhole camera model, and then are back-projected into the current camera frame by compounding the position with the motion matrix obtained from the ego-motion algorithm. Such transformed 3D points from the previous frame connected to corresponding 3D points from the current frame form a vector field, with each vector representing a motion of corresponding 3D point relative to the world frame. Since the measurement uncertainties are highly anisotropic in 3D space, it is difficult to accurately determine the motion intensity along the optical axis direction. Hence, we project the vectors into the image plane where the uncertainties are more evenly distributed, and apply the threshold on the magnitude of motion of each point. The remaining vectors are then connected into clusters by respecting both translational and rotational parameters. Finally, we consider each clusters corresponds to a moving object if at least 3 vectors appear within it, and describe it with the centroid point of all the corresponding points. The positions of the moving objects detected with the stereo camera system are finally projected into the radar plane and passed to the multitarget tracking algorithm presented in Sec. 5.

The projection of raw detections of the stereo vision based detection (red circles) and radar readings (green circles) onto the image plane along with filter tracks (yellow circles) is shown in Fig. 3. The images represent four snapshots of the experiment which illustrates the drawbacks of using just a single sensing technology. For example, in the top-most snapshot within Fig. 3 the radar did not capture the two motorcycles, while the stereo camera managed to detect their motion. The second snapshot gives an example of a busy intersection, while the third snapshot shows an example where the vehicle right in front of the ego-vehicle was not detected by the stereo camera

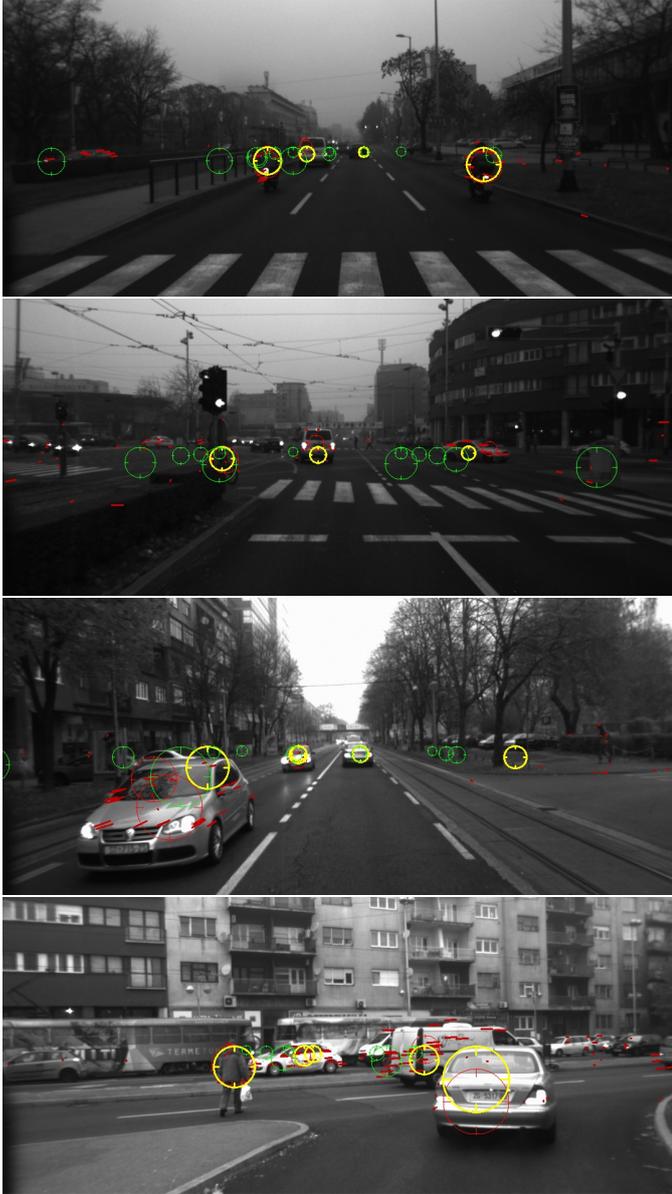


Figure 3: Four snapshots of experiments illustrating detections of the stereo camera (red circles) and radar readings (green circles), which serve as the input for the tracking algorithm (yellow circles). The red lines depict optical flow vectors of the detected motion. An accompanying video is available at [youtube.be/Rpf87bxZwYk](https://www.youtube.com/watch?v=Rpf87bxZwYk).

due to moving along the camera's optical axis whereas the radar provided consistent detections and the vehicle was tracked by the filter. The final snapshot shows an example where the radar did not detect a vehicle and a pedestrian, while the stereo camera managed to consistently detect their motion and respective filter tracks were obtained.

6.3. Real-world experiments

The experiments were conducted with the sensor platform equipped vehicle driving through an urban environment. The algorithm was tested in several highly dynamic scenarios, involving cars, trams and pedestrians. The process noise inten-

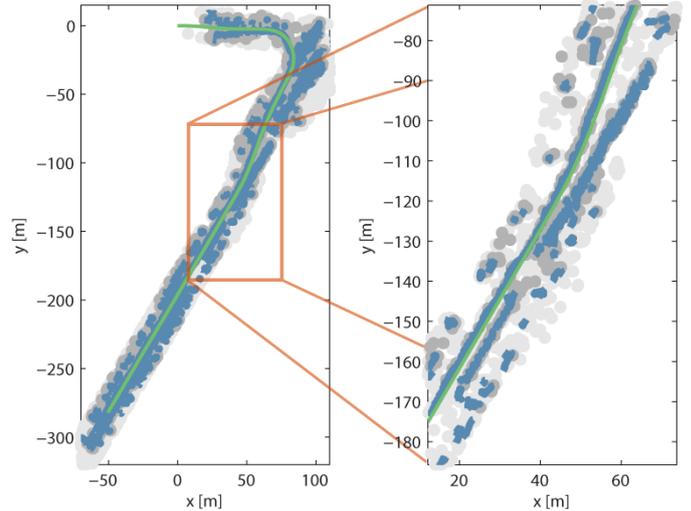


Figure 4: The experimental scenario in which the platform vehicle turned right and kept driving down an avenue. The left part shows the entire 2D projection of the experiment where light and dark gray dots correspond to stereo and radar measurements, blue lines correspond to existing moving objects in the surrounding, and green line represents the ego motion of the vehicle (starting from (0, 0)).

sities for the asynchronous filter were set to $\tilde{q}_x = \tilde{q}_y = 1$ and $\tilde{q}_\omega = (2 \frac{\pi}{180})^2$. The clutter size and the probability of detection were set to $c^{radar} = 10$ and $P_D^{radar} = 0.7$, respectively. The radar unit likelihood was configured such that the measurement uncertainty in the bearing component was $m_{k+1}^{\phi, radar} \sim \mathcal{N}_{\mathbb{R}^1}(0, 2^\circ)^2$, while the measurement uncertainty in the range component was $m_{k+1}^{r, radar} \sim \mathcal{N}_{\mathbb{R}^1}(0, 0.25^2)$. The clutter size related to the stereo vision system was set to $c^{stereo} = 2$, while the detection probability was $P_D^{stereo} = 0.75$. The stereo vision likelihood was configured so that the measurement uncertainty in the bearing component was $m_{k+1}^{\phi, stereo} \sim \mathcal{N}_{\mathbb{R}^1}(0, 0.5^\circ)^2$, while the measurement uncertainty in the range component was $m_{k+1}^{r, stereo} \sim \mathcal{N}_{\mathbb{R}^1}(0, 1^2)$. The JIPDA filter gating probability was $P_G = 0.9$, and the survival probability was $p_S = 0.95$. We have implemented an approach where the tracks are confirmed to be truly existing objects once the probability of object existence exceeded the value of $p(\chi_k^i | Z_{1:k}) = 0.9$. The tracks were deleted once the probability of existence fell below $p(\chi_k^i | Z_{1:k}) = 0.1$.

The first experiment, lasting about 60 s, involved a scenario in which the vehicle turned right and kept driving down an avenue. The results of this experiment are shown in Fig. 4. In this example it is important to note the very dense traffic on the left-hand side of the vehicle during the turn, which represents a very busy intersection (see the most bottom image in Fig. 3). However, due to high radar clutter, it occasionally happened that the clutter caused false tracks. Such an example can be seen on the right-hand image of Fig. 4. Even though the algorithm manages to track the vehicles on the road (in both directions), some objects, like the roadside hedges next to the road and the pertaining radar clutter, have caused the algorithm to detect them too as true targets. In this experiment, after raw sensor data preprocessing, on average there were 6.46 radar detections and

1.69 stereo camera detections per frame which yielded 3560 filter initializations and 228 confirmed tracks.

In the second experiment, lasting about 85 s, the vehicle drove in one direction along a three lane avenue, performed a u-turn (at the same busy intersection as in Fig. 4) and kept driving forward. The results of this experiment are shown in Fig. 5. The dataset was collected on a three lane road, where the vehicle drove in the middle lane, and detected vehicles in both the left and right lane. It can be noticed that again some radar measurements have caused the algorithm to believe that roadside objects corresponds to true targets. By analyzing the results we have noticed the occasional appearance of false positive trajectories, i.e. the ones that correspond to roadside hedges. In this experiment, after raw sensor data preprocessing, on average there were 12.19 radar detections and 3.0 stereo camera detections per frame which yielded 6935 filter initializations and 450 confirmed tracks. It is also important to mention that we have conducted the experiments during a foggy day, which presented challenging conditions for the stereo image processing.

6.4. Discussion

The presented experimental results illustrate the ability of the proposed approach to track moving objects in the context of ADAS with sensing systems of different modalities, i.e., the radar unit and the stereo camera system—a combination of sensing technologies that has recently been adopted by many car manufacturers. However, to the best of the authors’ knowledge, none of the available datasets using these sensors contain ground truth data, hence it is difficult to ensure a quantitative real-world experimental evaluation of the proposed approach. Still, in our previous work [40] we have performed an in-depth evaluation of filtering on Lie Groups in simulations, and proven the advantages of using $SE(2)^2$ state space for tracking whenever the characteristics of the system are such that the Euclidean space can not fully account for the geometry of the state space, while in this work we have applied the mentioned results for multitarget tracking in an ADAS application, and particularly for the sensors whose measurements arise in polar coordinates. Hence, in the present paper we omit an in-depth simulation based evaluation of the LG-EKF procedure.

From the viewpoint of estimation, the advantages of the proposed approach lie in the flexibility of modeling the sensors’ and the tracked object’s uncertainty and motion. This can prove advantageous in projecting the object’s future motion and uncertainty thereof for applications such as collision avoidance or motion planning of autonomous vehicles. The detection procedure of the stereo camera does not rely on a specific appearance of objects and can detect arbitrary motion, including that of cars, vans, motorcycles, and pedestrians as shown in Fig. 3. However, therein lies also the disadvantage of being able to detect only objects exhibiting relative motion with respect to the ego-motion. Objects moving in parallel to the car with the exact same velocity, thus in the image appearing as static, and objects moving along the optical axis can be difficult to detect with the stereo camera. This necessitates then the need for fusing data with other sensors, such as the radar, which can then comple-

ment these situations and yield better range measurements for objects further away from the ego-vehicle.

Also, as mentioned in Sec. 2, the JIPDA filter in its basic Kalman filter-like form represents a well-established approach for multitarget tracking problems. By performing the presented experiments, we have verified the approach of joining the two fundamental multitarget tracking building blocks: the state estimation and probabilistic data association scheme, both based on the geometry of Lie Groups. Given the above, we believe this work will not only serve as a DATMO reference, but also as a guideline for using the LG-EKF in various ADAS aspects.

7. Conclusion

In this paper we have addressed the detection and tracking problem, within the context of advanced driver assistance systems, with a multisensor setup consisting of a radar unit and a stereo camera. The stereo camera estimated relative displacement of the vehicle, using stereo visual odometry, generating measurements as cluster centers of optical flow vectors not conforming to the estimated motion. The radar directly reported its measurements to the filter, thus complementing the stereo camera measurements. Since the two sensors worked at different frequencies, sensor measurements were fused using an asynchronous Kalman filter on Lie groups.

This particular representation was proposed so as to most faithfully model the uncertainties of both the sensor measurements and the vehicle’s state. Concretely, the radar and the stereo camera were modeled as polar sensors, while the vehicle’s state resided on the Lie group $SE(2)^2$. This enabled us to reliably model the uncertainties as having *banana-shaped* contours, when such a situation arises, in contrast to elliptical uncertainty contours given by the ‘classical’ Gaussian distribution. To solve the multitarget tracking problem we adapted the JIPDA filter to work with the Kalman filter on Lie groups. In the end, the proposed filter performance was presented on a real-world dataset recorded in urban traffic scenarios.

Appendix A. Derivation of \mathcal{H}

As part of the update step we need to derive the matrix \mathcal{H}_{k+1} denoting the LG-EKF equivalent to the Kalman filter measurement Jacobian. Before proceeding with explicit derivation, we define the measurement function $h(X_{k+1})$ as

$$h(X_{k+1}) = \begin{bmatrix} \exp_{SO(2)} \left(\left[\arctan \frac{y_{k+1}}{x_{k+1}} \right]_{SO(2)}^\wedge \right) \\ \exp_{\mathbb{R}^1} \left(\left[\sqrt{x_{k+1}^2 + y_{k+1}^2} \right]_{\mathbb{R}^1}^\wedge \right) \end{bmatrix} \quad (\text{A.1})$$

For this purpose we start with the definition of the Lie algebraic error $\epsilon = [\epsilon_x \ \epsilon_y \ \epsilon_\theta \ \epsilon_{v_x} \ \epsilon_{v_y} \ \epsilon_\omega]$. We further provide the prerequisites for deriving \mathcal{H} . We firstly give the expression which is an

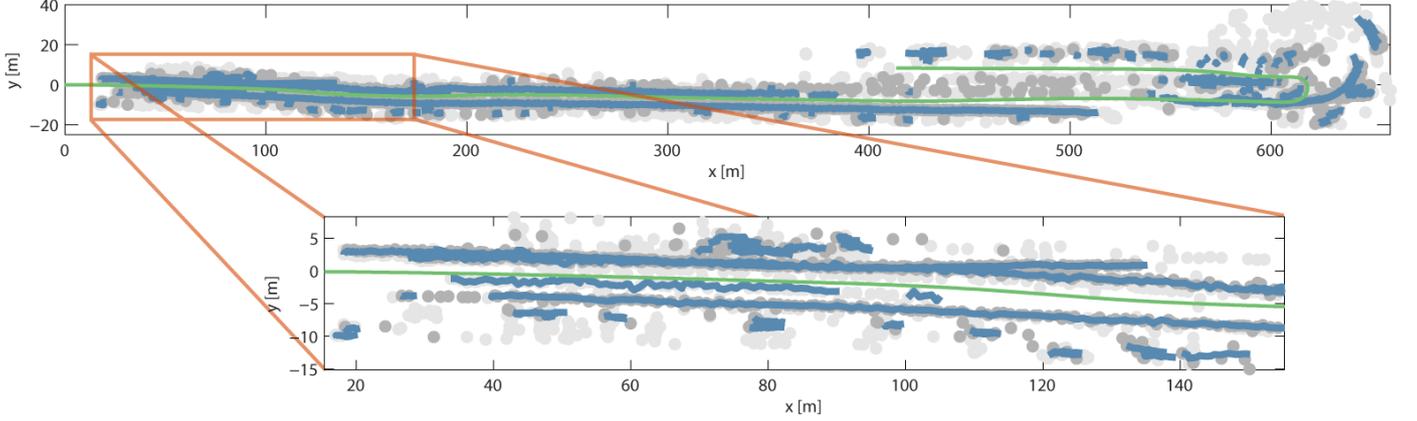


Figure 5: The experimental scenario in which the vehicle drove in direction, performed a u-turn, and kept driving forward. The upper part shows the entire 2D projection of the experiment where light and dark gray dots correspond to stereo and radar measurements, blue lines correspond to existing moving objects in the environment, and green line represents the ego motion of the vehicle starting from (0, 0).

argument for evaluating \mathcal{H}

$$h(\mu_{k+1|k})^{-1} h(\mu_{k+1|k} \exp_G([\epsilon]_G^\wedge)) = \begin{bmatrix} h_c^1 \\ h_c^2 \end{bmatrix} = \quad (\text{A.2})$$

$$\begin{bmatrix} \exp_{\text{SO}(2)}^{-1} \left(\left[\text{atan2} \frac{y_{k+1}}{x_{k+1}} \right]_{\text{SO}(2)}^\wedge \right) \exp_{\text{SO}(2)} \left(\left[\text{atan2} \frac{y_{k+1}^\epsilon}{x_{k+1}^\epsilon} \right]_{\text{SO}(2)}^\wedge \right) \\ \exp_{\mathbb{R}^1}^{-1} \left(\left[\sqrt{x_{k+1}^2 + y_{k+1}^2} \right]_{\mathbb{R}^1}^\wedge \right) \exp_{\mathbb{R}^1} \left(\left[\sqrt{x_{k+1}^{\epsilon 2} + y_{k+1}^{\epsilon 2}} \right]_{\mathbb{R}^1}^\wedge \right) \end{bmatrix}.$$

where x_{k+1}^ϵ and y_{k+1}^ϵ denote variables extracted from the current matrix Lie group system state X_{k+1} , compound with the Lie algebraic error mapped via the \exp_G . These two variables are hence given as

$$\begin{aligned} x_{k+1}^\epsilon &= x_{k+1} + \cos \theta_{k+1} f - \sin \theta_{k+1} g \\ y_{k+1}^\epsilon &= y_{k+1} + \sin \theta_{k+1} f + \cos \theta_{k+1} g. \end{aligned} \quad (\text{A.3})$$

where the terms f and g follow terms

$$\begin{aligned} f &= [\epsilon_x \sin \epsilon_\theta + \epsilon_y (-1 + \cos \epsilon_\theta)] \epsilon_\theta^{-1} \\ g &= [\epsilon_x (1 - \cos \epsilon_\theta) + \epsilon_y \sin \epsilon_\theta] \epsilon_\theta^{-1}. \end{aligned} \quad (\text{A.4})$$

The function to be partially derived is obtained by taking the logarithm on G' as follows

$$\left[\log_{G'} \left(\begin{bmatrix} h_c^1 \\ h_c^2 \end{bmatrix} \right) \right]_{G'}^\vee = \begin{bmatrix} \left[\log_{\text{SO}(2)} (h_c^1) \right]_{\text{SO}(2)}^\vee \\ \left[\log_{\mathbb{R}^1} (h_c^2) \right]_{\mathbb{R}^1}^\vee \end{bmatrix}. \quad (\text{A.5})$$

Let \mathcal{H}_{k+1}^1 and \mathcal{H}_{k+1}^2 denote the two rows of (A.5). In order to derive (31), we need to determine partial derivatives and multivariate limits over all directions of the Lie algebraic error vec-

tor. This result is given as

$$\begin{aligned} \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_x} \Big|_0 &= \frac{-y_{k+1|k} \cos \theta_{k+1|k} + x_{k+1|k} \sin \theta_{k+1|k}}{x_{k+1|k}^2 + y_{k+1|k}^2} \\ \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_y} \Big|_0 &= \frac{x_{k+1|k} \cos \theta_{k+1|k} + y_{k+1|k} \sin \theta_{k+1|k}}{x_{k+1|k}^2 + y_{k+1|k}^2} \\ \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_x} \Big|_0 &= \frac{x_{k+1|k} \cos \theta_{k+1|k} + y_{k+1|k} \sin \theta_{k+1|k}}{\sqrt{x_{k+1|k}^2 + y_{k+1|k}^2}} \\ \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_y} \Big|_0 &= \frac{y_{k+1|k} \cos \theta_{k+1|k} - x_{k+1|k} \sin \theta_{k+1|k}}{\sqrt{x_{k+1|k}^2 + y_{k+1|k}^2}} \\ \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_\theta} \Big|_0 &= 0 \quad \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_\theta} \Big|_0 = 0 \\ \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_{v_x}} \Big|_0 &= 0 \quad \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_{v_y}} \Big|_0 = 0 \quad \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_\omega} \Big|_0 = 0 \\ \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_{v_x}} \Big|_0 &= 0 \quad \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_{v_y}} \Big|_0 = 0 \quad \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_\omega} \Big|_0 = 0 \end{aligned} \quad (\text{A.6})$$

The final measurement matrix \mathcal{H}_{k+1} is given as

$$\mathcal{H}_{k+1} = \begin{bmatrix} \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_x} \Big|_0 & \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_y} \Big|_0 & \frac{\partial \mathcal{H}_{k+1}^1}{\partial \epsilon_\theta} \Big|_0 & \mathbf{0}^{2 \times 3} \\ \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_x} \Big|_0 & \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_y} \Big|_0 & \frac{\partial \mathcal{H}_{k+1}^2}{\partial \epsilon_\theta} \Big|_0 & \end{bmatrix}. \quad (\text{A.7})$$

Even though the term (A.5) appears involved, the relations (A.6) are actually obtained by patient algebraic manipulations and hence the detailed derivation is not shown here.

References

- [1] E. Dickmanns, The development of machine vision for road vehicles in the last decade, in: Intelligent Vehicle Symposium (IV), IEEE, 2002, pp. 268 – 281.
- [2] S. Sivaraman, M. M. Trivedi, Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis, IEEE Transactions on Intelligent Transportation Systems 14 (4) (2013) 1773–1795.

- [3] M. Oliveira, V. Santos, A. Sappa, P. Dias, Scene representations for autonomous driving: An approach based on polygonal primitives, in: *Robot 2015: Second Iberian Robotics Conference*, Vol. 417, Springer, 2016, pp. 503–515.
- [4] B. Steux, C. Lurgeau, L. Salesse, D. Wautier, Fade: a vehicle detection and tracking system featuring monocular color vision and radar data fusion, in: *Intelligent Vehicle Symposium (IV)*, 2002, pp. 632–639.
- [5] G. Alessandretti, A. Broggi, P. Cerri, Vehicle and guard rail detection using radar and vision data fusion, *IEEE Trans. on Intelligent Transp. Systems* 8 (1) (2007) 95–105.
- [6] Z. Ji, D. Prokhorov, Radar-vision fusion for object classification, in: *International Conference on Information Fusion (FUSION)*, IEEE, 2008, pp. 1–6.
- [7] T. Wang, N. Zheng, J. Xin, Z. Ma, Integrating millimeter wave radar with a monocular vision sensor for on-road obstacle detection applications, *Sensors* 11 (9) (2011) 8992–9008.
- [8] E. Richter, R. Schubert, G. Wanielik, Radar and Vision-based Data Fusion - Advanced Filtering Techniques for a Multi-object Vehicle Tracking System, in: *Intelligent Vehicles Symposium (IV)*, IEEE, 2008, pp. 120–125.
- [9] F. Liu, J. Sparbert, C. Stiller, IMM-PDA Vehicle Tracking System using Asynchronous Sensor Fusion of Radar and Vision, in: *Intelligent Vehicles Symposium (IV)*, IEEE, 2008, pp. 168–173.
- [10] R. O. Chavez-Garcia, T.-D. Vu, J. Bulet, O. Aycard, Frontal object perception using radar and mono-vision, in: *Intelligent Vehicles Symposium (IV)*, IEEE, 2012, pp. 159–164.
- [11] T. Kato, Y. Ninomiya, I. Masaki, An Obstacle Detection Method by Fusion of Radar and Motion Stereo, *IEEE Transactions on Intelligent Transportation Systems* 3 (3) (2002) 182–187.
- [12] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. C. Antonello, M. Miglietta, Obstacle detection and classification fusing radar and vision, in: *Intelligent Vehicles Symposium (IV)*, IEEE, 2008, pp. 608–613.
- [13] Y. Fang, I. Masaki, B. Horn, Depth-Based Target Segmentation for Intelligent Vehicles: Fusion of Radar and Binocular Stereo, *IEEE Transactions on Intelligent Transportation Systems* 3 (3) (2002) 196–202.
- [14] S. Wu, S. Decker, P. Chang, T. Camus, J. Eledath, Collision sensing by stereo vision and radar sensor fusion, *IEEE Transactions on Intelligent Transportation Systems* 10 (4) (2009) 606–614.
- [15] I. Marković, I. Petrović, Bayesian Sensor Fusion Methods for Dynamic Object Tracking — a Comparative Study, *Automatika* 55 (4) (2014) 386–398.
- [16] M. Obrvan, J. Česić, I. Petrović, Appearance based vehicle detection by radar-stereo vision integration, in: *Robot 2015: Second Iberian Robotics Conference*, Vol. 417, Springer, 2016, pp. 437–449.
- [17] A. Broggi, A. Cappelunga, C. Caraffi, S. Cattani, S. Ghidoni, P. Grisleri, P. P. Porta, M. Posterli, P. Zani, TerraMax vision at the Urban challenge 2007, *IEEE Trans. on Intell. Transp. Systems* 11 (1) (2010) 194–205.
- [18] B. Lefaudeux, F. Nashashibi, Real-time visual perception: Detection and localisation of static and moving objects from a moving stereo rig, in: *Intelligent Transportation Systems Conference (ITSC)*, 2012, pp. 522–527.
- [19] P. Lenz, J. Ziegler, A. Geiger, M. Roser, Sparse scene flow segmentation for moving object detection in urban environments, in: *Intelligent Vehicles Symposium (IV)*, 2011, pp. 926–932.
- [20] A. Wedel, T. Brox, T. Vaudrey, C. Rabe, U. Franke, D. Cremers, Stereoscopic scene flow computation for 3D motion understanding, *International Journal of Computer Vision* 95 (1) (2011) 29–51.
- [21] B. Barrois, S. Hristova, C. Wohler, F. Kummert, C. Hermes, 3D Pose Estimation of Vehicles Using a Stereo Camera, in: *Intelligent Vehicles Symposium (IV)*, 2009, pp. 267–272.
- [22] E. Ohn-bar, S. Sivaraman, M. Trivedi, Partially Occluded Vehicle Recognition and Tracking in 3D, in: *Intell. Veh. Symposium (IV)*, 2013, pp. 1350–1355.
- [23] J. Almeida, V. Santos, Pedestrian pose estimation using stereo perception, in: *Robot 2015: Second Iberian Robotics Conference*, Vol. 417, Springer, 2016, pp. 491–502.
- [24] I. Cvišić, I. Petrović, Stereo odometry based on careful feature selection and tracking, in: *European Conference on Mobile Robots (ECMR)*, 2015.
- [25] G. Bourmaud, R. Mégret, M. Arnaudon, A. Giremus, Continuous-Discrete Extended Kalman Filter on Matrix Lie Groups Using Concentrated Gaussian Distributions, *Journal of Mathematical Imaging and Vision* 51 (1) (2014) 209–228.
- [26] R. P. S. Mahler, Multitarget Bayes filtering via first-order multitarget moments, *IEEE Transactions on Aerospace and Electronic Systems* 39 (4) (2003) 1152–1178.
- [27] B.-N. Vo, S. Singh, A. Doucet, Sequential monte carlo methods for multi-target filtering with random finite sets, *IEEE Transactions on Aerospace and Electronic Systems* 41 (4) (2005) 1224–1245.
- [28] B.-N. Vo, W.-K. Ma, The Gaussian mixture probability hypothesis density filter, *IEEE Transactions on Signal Processing* 54 (11) (2006) 4091–4104.
- [29] R. Mahler, PHD filters of higher order in target number, *IEEE Transactions on Aerospace and Electronic Systems* 43 (4) (2007) 1523–1543.
- [30] B.-T. Vo, B.-N. Vo, A. Cantoni, Analytic implementations of the cardinalized probability hypothesis density filter, *IEEE Transactions on Signal Processing* 55 (7) (2007) 3553–3567.
- [31] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*, Artech House, 2007.
- [32] B. T. Vo, B. N. Vo, A. Cantoni, The cardinality balanced multi-target multi-Bernoulli filter and its implementations, *IEEE Transactions on Signal Processing* 57 (2) (2009) 409–423.
- [33] S. Reuter, B. T. Vo, B. N. Vo, K. Dietmayer, The labeled multi-Bernoulli filter, *IEEE Transactions on Signal Processing* 62 (12) (2014) 3246–3260.
- [34] H. Deusch, S. Reuter, K. Dietmayer, The labeled multi-Bernoulli SLAM filter, *IEEE Signal Processing Letters* 22 (10) (2015) 1561–1565.
- [35] D. Reid, An algorithm for tracking multiple targets, *IEEE Transactions on Automatic Control* 24 (6) (1979) 843–854.
- [36] Y. Bar-Shalom, E. Tse, Sonar tracking of multiple targets using joint probabilistic data association filter, *Automatica* 11 (1975) 451–460.
- [37] D. Mušicki, R. Evans, Joint Integrated Probabilistic Data Association – JIPDA, *IEEE Transactions on Aerospace and Electronic Systems* 40 (3) (2004) 1093 – 1099.
- [38] S. Challa, M. R. Morelande, D. Mušicki, R. J. Evans, *Fundamentals of Object Tracking*, 2011.
- [39] J. L. Williams, Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA and association-based MeMber, *IEEE Transactions on Aerospace and Electronic Systems* 51 (3) (2015) 1664–1687.
- [40] J. Česić, I. Marković, I. Petrović, Moving object tracking employing rigid body dynamics on matrix Lie groups, in: *Submitted to International Conference on Information Fusion (FUSION)*, 2016.
- [41] C. Hertzberg, R. Wagner, U. Frese, L. Schröder, Integrating Generic Sensor Fusion Algorithms with Sound State Representations through Encapsulation of Manifolds, *Information Fusion* 14 (1) (2013) 57–77.
- [42] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*, Springer, 2012.
- [43] W. Park, Y. Wang, G. S. Chirikjian, The Path-of-Probability Algorithm for Steering and Feedback Control of Flexible Needles, *The International Journal of Robotics Research* 29 (7) (2010) 813–830.
- [44] T. D. Barfoot, P. T. Furgale, Associating Uncertainty With Three-Dimensional Poses for Use in Estimation Problems, *IEEE Transactions on Robotics* 30 (3) (2014) 679–693.
- [45] Y. Wang, G. S. Chirikjian, Nonparametric Second-Order Theory of Error Propagation on Motion Groups, *International Journal on Robotic Research* 27 (11) (2008) 1258–1273.
- [46] K. C. Wolfe, M. Mashner, G. S. Chirikjian, Bayesian Fusion on Lie Groups, *Journal of Algebraic Statistics* 2 (1) (2011) 75–97.
- [47] Y. Bar-Shalom, T. Kirubarajan, X.-R. Li, *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, Inc., 2002.
- [48] S. Thrun, W. Burgard, D. Fox, *Probabilistic Robotics*, The MIT Press, 2006.
- [49] R. Schubert, C. Adam, M. Obst, N. Mattern, V. Leonhardt, G. Wanielik, Empirical evaluation of vehicular models for ego motion estimation, in: *Intelligent Vehicles Symposium, IEEE, 2011*, pp. 534–539.
- [50] D. Salmund, Mixture reduction algorithms for point and extended object tracking in clutter, *IEEE Transactions on Aerospace and Electronic Systems* 45 (2) (2009) 667–686.
- [51] J. Shi, C. Tomasi, Good features to track, in: *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94.*, 1994 IEEE Computer Society Conference on, 1994, pp. 593–600.
- [52] J.-Y. Bouguet, Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm, *Tech. Rep. 2*, Intel Corporation, Microsoft Research Labs (2000).
- [53] K. Konolige, *Small vision systems: Hardware and implementation*, in: *Robotics Research*, Springer London, 1998, pp. 203–212.