

# Journal of Advances in Information Fusion

A semi-annual archival publication of the International Society of Information Fusion

Regular Papers	Page
<b>Extended Object Tracking: Introduction, Overview, and Applications</b> ..... 139 <i>Karl Granström, Chalmers University of Technology, Sweden</i> <i>Marcus Baum, University of Goettingen, Germany</i> <i>Stephan Reuter, Ulm University, Germany</i>	139
<b>Multi Detection Joint Integrated Probabilistic Data Association Using Random Matrices with Applications to Radar-Based Multi Object Tracking</b> ..... 175 <i>Michael Schuster, Konstanz University of Applied Sciences, Germany</i> <i>Johannes Reuter, Konstanz University of Applied Sciences, Germany</i> <i>Gerd Wanielik, Chemnitz University of Technology, Germany</i>	175
<b>Converted Measurements Bayesian Extended Target Tracking Applied to X-band Marine Radar Data</b> ..... 189 <i>Gemine Vivone, Atlantic Treaty Organization (NATO) Science and Technology Organization (STO), Italy</i> <i>Paolo Braca, Atlantic Treaty Organization (NATO) Science and Technology Organization (STO), Italy</i> <i>Karl Granström, Chalmers University of Technology, Sweden</i> <i>Antonio Natale, IREA-CNR, Italy</i> <i>Jocelyn Chanussot, University of Iceland, Iceland</i>	189
<b>Priority-Based Tracking of Extended Objects</b> ..... 211 <i>Kevin Wyffels, Cornell University, USA</i> <i>Mark Campbell, Cornell University, USA</i>	211
<b>Extended Object Tracking with Exploitation of Range Rate Measurements</b> ..... 228 <i>Steven Bordonaro, Naval Undersea Warfare Center, USA</i> <i>Peter Willett, University of Connecticut, USA</i> <i>Yaakov Bar-Shalom, University of Connecticut, USA</i> <i>Marcus Baum, University of Connecticut, USA</i> <i>Tod Luginbuhl, Naval Undersea Warfare Center, USA</i>	228

*From the  
Editor-In-Chief*

*Guest Editorial:  
Foreword to the  
Special Issue on  
Extended Object  
Tracking*

# INTERNATIONAL SOCIETY OF INFORMATION FUSION

The International Society of Information Fusion (ISIF) is the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied INFORMATION FUSION technologies. Technical areas of interest include target tracking, detection theory, applications for information fusion methods, image fusion, fusion systems architectures and management issues, classification, learning, data mining, Bayesian and reasoning methods.

## JOURNAL OF ADVANCES IN INFORMATION FUSION: December 2017

---

<b>Editor-In-Chief</b>	Uwe D. Hanebeck	Karlsruhe Institute of Technology (KIT), Germany; +49-721-608-43909; uwe.hanebeck@ieee.org
Associate	Stefano Coraluppi	Systems & Technology Research, USA; +1 781-305-4055; stefano.coraluppi@ieee.org
<b>Administrative Editor</b>	David W. Krout	University of Washington, USA; +1 206-616-2589; dkrou@apl.washington.edu
Associate	Ruixin Niu	Virginia Commonwealth University, Richmond, Virginia, USA; +1 804-828-0030; rniu@vcu.edu
Associate	Marcus Baum	Karlsruhe Institute of Technology (KIT), Germany; +49-721-608-46797; marcus.baum@kit.edu

## EDITORS FOR TECHNICAL AREAS

---

<b>Tracking</b>	Stefano Coraluppi	Systems & Technology Research, USA; +1 781-305-4055; stefano.coraluppi@ieee.org
Associate	Paolo Braca	NATO Science & Technology Organization, Centre for Maritime Research and Experimentation, Italy; +39 0187 527 461; paolo.braca@cmre.nato.int
<b>Detection</b>	Pramod Varshney	Syracuse University, Syracuse, New York, USA; +1 315-443-1060; varshney@syr.edu
<b>Fusion Applications</b>	Ben Slocumb	Numerica Corporation; Loveland, Colorado, USA; +1 970-461-2000; bjslocumb@numerica.us
Associate	Ramona Georgescu	United Technologies Research Center, East Hartford, Connecticut, USA; 860-610-7890; georgera@utrc.utc.com
<b>Image Fusion</b>	Lex Toet	TNO, Soesterberg, 3769de, Netherlands; +31 346356237; lex.toet@tno.nl
Associate	Ting Yuan	Mercedes Benz R&D North America, USA; +1 669-224-0443; dr.ting.yuan@ieee.org
<b>High-Level Fusion</b>	Lauro Snidaro	Università degli Studi di Udine, Udine
<b>Fusion Architectures and Management Issues</b>	Chee Chong	BAE Systems, Los Altos, California, USA; +1 650-210-8822; chee.chong@baesystems.com
<b>Classification, Learning, Bayesian and Other Reasoning Methods</b>	Nageswara S. V. Rao	Oak Ridge National Laboratory, USA; +1 865-574-7517;
Associate	Claude Jauffret	Université de Toulon, La Garde, France; + 33 (0) 4 94 14 24 14; jauffret@univ-tln.fr
Associate	Jean Dezert	ONERA, Chatillon, 92320, France; +33 146734990; jean.dezert@onera.fr

Manuscripts are submitted at <http://jaif.msubmit.net>. If in doubt about the proper editorial area of a contribution, submit it under the unknown area.

## INTERNATIONAL SOCIETY OF INFORMATION FUSION

---

Mila Mihaylova, *President*

Lyudmila Mihaylova, *President-elect*

Stefano Coraluppi, *Secretary*

Chee Chong, *Treasurer*

Dale Blair, *Vice President Publications*

David W. Krout, *Vice President Communications*

Lance Kaplan, *Vice President Conferences*

Anne-Laure Jousselme, *Vice President Membership*

Paulo Costa, *Vice President Working Groups*

Uwe Hanebeck, *JAIF EIC*

Roy Streit, *Perspectives EIC*

Journal of Advances in Information Fusion (ISSN 1557-6418) is published semi-annually by the International Society of Information Fusion. The responsibility for the contents rests upon the authors and not upon ISIF, the Society, or its members. ISIF is a California Nonprofit Public Benefit Corporation at P.O. Box 4631, Mountain View, California 94040. **Copyright and Reprint Permissions:** Abstracting is permitted with credit to the source. For all other copying, reprint, or republication permissions, contact the Administrative Editor. Copyright© 2017 ISIF, Inc.

# From the Editor-in-Chief:

December 2017



Karl Granström



Marcus Baum

## Guest Editorial: Foreword to the Special Issue on *Extended Object Tracking*

Multi-Object tracking, i.e., the successive determination of the number and states of objects based on sensor measurements, is an enabling technology in many areas such as surveillance, (mobile) robotics, autonomous driving and automation. Traditional multi-object tracking algorithms are based on the “small target” assumptions, i.e., (i) objects evolve independently, (ii) objects can be modelled as a point, and (iii) objects gives rise to at most one measurement per scan. These assumptions are usually valid for target objects that are far away from the sensor or for a low sensor resolution, e.g., in air surveillance.

However, due to recent advances in sensor technology, as well as novel applications involving objects in the near-field of sensors, it is becoming increasingly common that the objects occupy several sensor resolution cells. Hence, there is an increasing demand for so-called extended object tracking algorithms that can deal with objects that gives rise to multiple measurements from different spatially distributed measurement sources.

Extended object tracking comes with many new challenges: For example, in contrast to a point target, it is necessary to determine the shape of the object, which is a highly nonlinear problem. Furthermore, data association becomes much more challenging for multiple extended objects because a huge amount of association events is feasible. Extended object tracking is applicable to many different sensors, such as radar, lidar, and camera.

The special issue consists of five articles, some of which are significantly extended versions of papers that have been presented at the International Conference on Information Fusion (FUSION) within the special session “*Extended Object and Group Tracking.*”

The first paper by Granström, Baum and Reuter, titled “*Extended Object Tracking: Introduction, Overview and Applications,*” gives a tutorial introduction to the extended object tracking problem. The most common methods are discussed and overviewed. Furthermore, several real-world applications of extended object tracking are illustrated.

The second paper by Schuster, Reuter and Wanielik, titled “*Multi Detection Joint Integrated Probabilistic Data Association Using Random Matrices with Applications to Radar-Based Multi Object Tracking*,” is about an extended object version of the Joint Probabilistic Data Association (JPDAF) filter that allows for multiple detections per object and employs the random matrix approach for shape estimation. The method is experimentally evaluated by means of vessel tracking using a high-resolution radar sensor.

The third paper by Vivone, Braca, Granström, Natale, and Chanussot, titled “*Converted Measurements Bayesian Extended Target Tracking Applied to X-band Marine Radar Data*,” deals with the tracking of marine vessels using X-band marine radar and the random matrix extended target model. X-band radars are flexible and low-cost tools that provide high resolution measurements in polar coordinates. To track targets modelled in Cartesian coordinates, it is necessary to convert the measurements from polar to Cartesian coordinates before they are input into the tracking algorithm. The paper presents a method for modelling detections affected by polar noise, and derives an efficient estimator. The estimator is evaluated using 10 real-world datasets.

The fourth paper by Wyffels and Campbell, titled “*Priority Based Tracking of Extended Objects*,” presents a framework for allocating computational resources to achieve tracking with variable precision for different objects. Inspired by human perception, higher relevance is given to objects that are closer to the sensor, i.e., closer to, e.g., the robot or ego-vehicle. Three levels of relevance are used, with more accurate, and computationally expensive, models used for objects with higher relevance, and cheaper models used for objects with lower relevance.

The fifth paper by Bordonaro, Willett, Bar-Shalom, Luginbuhl, and Baum, titled “*Extended Object Tracking with Exploitation of Range Rate Measurements*,” considers the raw measurements (range, bearing and range rate) from a single scan and develops an expectation maximization (EM) algorithm that provides the target position, velocity, heading and turn rate. The obtained single scan estimate is subsequently used in a Kalman filter for recursive estimation.

The special issue covers several different aspects of extended object tracking, and represents an excellent sample of current research trends in the area, and of state-of-the-art methods and results. We sincerely hope that you will find reading the papers as interesting and enjoyable as we have. We would like to extend our sincere gratitude to the JAIF Editorial Board for the possibility to prepare this special issue, and to the Editor-in-Chief for the support and the encouragement. Finally, we wish to thank the authors for their contributions to the special issue and to the research area in general.

Karl Granström  
Department of Signals and Systems  
Chalmers University of Technology,  
Gothenburg, Sweden  
E-mail: karl.granstrom@chalmers.se

Marcus Baum  
Institute of Computer Science  
University of Goettingen, Goettingen, Germany  
E-mail: marcus.baum@cs.uni-goettingen.de

Guest Associate Editors

# Extended Object Tracking: Introduction, Overview, and Applications

KARL GRANSTRÖM  
MARCUS BAUM  
STEPHAN REUTER

This article provides an elaborate overview of current research in extended object tracking. We provide a clear definition of the extended object tracking problem and discuss its delimitation to other types of object tracking. Next, different aspects of extended object modelling are extensively discussed. Subsequently, we give a tutorial introduction to two basic and well used extended object tracking approaches—the random matrix approach and the Kalman filter-based approach for star-convex shapes. The next part treats the tracking of multiple extended objects and elaborates how the large number of feasible association hypotheses can be tackled using both Random Finite Set (RFS) and Non-RFS multi-object trackers. The article concludes with a summary of current applications, where four example applications involving camera, X-band radar, light detection and ranging (LIDAR), and red-green-blue-depth (RGB-D) sensors are highlighted.

Manuscript received January 29, 2016; released for publication February 8, 2017.

Refereeing of this contribution was handled by Stefano Coraluppi.

Authors' addresses: K. Granström, Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden (E-mail: karl.granstrom@chalmers.se). M. Baum, Institute of Computer Science, University of Goettingen, Goettingen, Germany (E-mail: marcus.baum@cs.uni-goettingen.de). S. Reuter, Institute of Measurement, Control, and Microtechnology, Ulm University, Ulm, Germany (E-mail: stephan.reuter@uni-ulm.de).

1557-6418/17/\$17.00 © 2017 JAIF

## I. INTRODUCTION

Multiple Target Tracking (MTT) denotes the process of successively determining the number and states of multiple dynamic objects based on noisy sensor measurements. Tracking is a key technology in many areas such as robotics, surveillance, autonomous driving, automation, medicine, and sensor networks.

Traditionally, MTT algorithms have been tailored for scenarios with multiple remote objects that are far away from the sensor, e.g., as in radar-based air surveillance. In such scenarios, an object is not always detected by the sensor, and if it is detected, at most one sensor resolution cell is occupied by the object. From traditional scenarios, specific assumptions on the mathematical model of MTT problems have evolved including the so-called “small object” assumptions:

- The objects evolve independently,
- each object can be modelled as a point without any spatial extent, and
- each object gives rise to at most a single measurement per time frame/scan.

MTT based on the “small object” assumptions is a highly complex problem due to sensor noise, missed detections, clutter detections, measurement origin uncertainty, and an unknown and time-varying number of targets. The most common approaches to MTT are:

- Multiple Hypothesis Tracking (MHT) [23], [106], [154],
- Joint Probabilistic Data Association (JPDA) [4], [6], [61],
- Probabilistic Multiple Hypothesis Tracking (PMHT) [177], [203], and
- Random Finite Sets (RFS) approaches [125], [127].

In the hypothesis-oriented MHT [154] and track-oriented MHT [106], the probability and log-likelihood ratio of a track, respectively, are calculated recursively. The JPDA type approaches blend data association probabilities on a scan-by-scan basis. The PMHT approach allows multiple measurement assignments to the same object,<sup>1</sup> which results in an efficient method using the Expectation-Maximization (EM) framework, see, e.g., [22, Ch. 9]. The RFS type approaches rely on modelling the objects and the measurements as random sets. A recent overview article about MTT, with a main focus on small, so-called point objects, is given in [197].

Today, there is still a huge variety of applications for which the “small object” assumptions are reasonable. However, due to rapid advances in sensor technology in the recent years, it is becoming increasingly common that objects occupy several sensor resolution cells. Furthermore, novel applications with objects in

<sup>1</sup>Note that allowing multiple assignments to the same object is in violation of the “small object” assumption, which assumes at most a single measurement per time frame/scan.

the near-field of sensors, e.g., in mobile robotics and autonomous driving, often render the “small object” assumptions invalid.

The tracking of an object that might occupy more than one sensor cell leads to the so-called *extended object tracking* or *extended target tracking* problem. In extended object tracking the objects give rise to a varying number of potentially noisy measurements from different spatially distributed measurement sources, also referred to as reflection points. The shape of the object is usually unknown and can even vary over time, and the objective is to recursively determine the shape of the object plus its kinematic parameters. Due to the non-linearity of the resulting estimation problem, already tracking a single extended object is in general a highly complex problem for which elaborate non-linear estimation techniques are required.

Although often misunderstood—extended object tracking, as defined above, is fundamentally different from typical contour tracking problems in computer vision [212]. In vision-based contour tracking [212], a complete red-green-blue (RGB) image is available at each time frame and one extracts a contour from each image that is tracked over time. In extended object tracking, one works with a few (usually two or three-dimensional) measurements per time step, i.e., a sparse point cloud. It is nearly always impossible to extract a shape only based on the measurement from one time instant. The object shape can only be determined if measurements over several time steps are systematically accumulated and fused under incorporation of the (unknown) object motion and sensor noise. An illustration of the difference between point object tracking, extended object tracking, and contour tracking is given in Figure 1.

In many practical applications it is necessary to track multiple extended objects, where no measurement-to-object associations are available. Unfortunately, data association becomes even more challenging in multiple *extended* object tracking as a huge number of association events are possible: all possible partitions of the set of measurements have to be enumerated, followed by all possible ways to assign partition cells to object estimates. The first computationally feasible multi-extended object tracking algorithms have recently been developed, and rely on approximations of the partitioning problem in the context of RFSS.

The objective of this article is to

- (i) provide an elaborate and up-to-date introduction to the extended object tracking problem,
- (ii) introduce basic concepts, models, and methods for shape estimation of a single extended object,
- (iii) introduce the basic concepts, models, and methods for tracking multiple extended objects,
- (iv) point out recent applications and future trends.

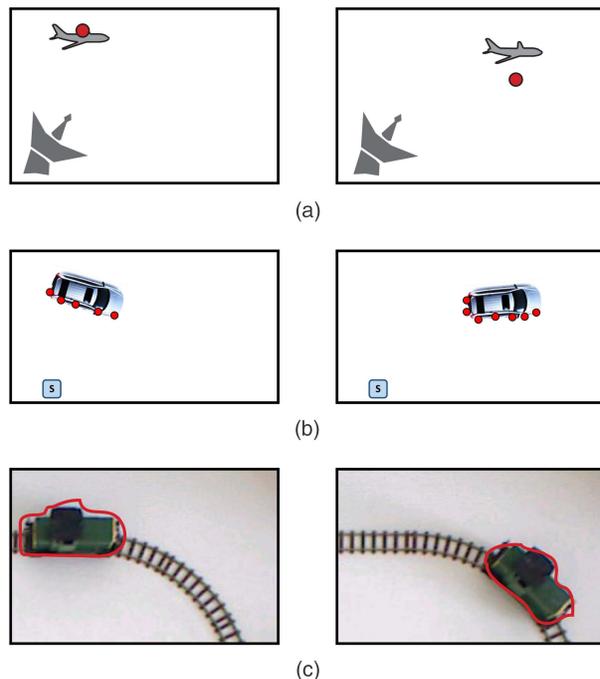


Fig. 1. Illustration of different types of tracking problems: a) Point object tracking example: Frame 1 (left) & Frame 2 (right). In point object tracking, at most one measurement (red markers) per frame is received. b) Extended object tracking example: Frame 1 (left) & Frame 2 (right). In extended object tracking, multiple measurements (red markers) from a varying number of measurement sources/reflection centers are obtained per frame. c) Contour Tracking example: Frame 1 (left) & Frame 2 (right). In contour tracking, a single contour (red) is extracted from each single image frame. Hence, one can say that in contour tracking, the measurements are contours, while in extended object tracking the measurements are (Cartesian) points. However, in both extended object tracking and contour tracking one aims at estimating the shape, i.e., a contour, based on the received measurements.

Historically, the first works on extended object tracking can be traced back to [42], [43]. Already in 2004, [199] gave a short literature overview of cluster (group) tracking and extended object tracking problems. However, since then, huge progress has been made in both shape estimation of a single object and multi-(extended)-object tracking. An overview of Sequential Monte Carlo (SMC) methods for group and extended object tracking can be found in [132]. The focus of [132] lies on group object tracking and SMC methods. Hence, the content of [132] is orthogonal to this article, and the two articles complement each other. A comparison of early versions of the random matrix and random hypersurface approach was performed in [17]. Since the publication of [17], both methods have been significantly further developed.

The rest of the article is organised as follows. In the next section some definitions are introduced, and modelling of object shape, number of measurements, and object dynamics is overviewed. Section III discusses two popular approaches to extent modelling and estimation: the random matrix model, Section III-A, and star-

convex models, Section III-B. Multiple extended object tracking is overviewed in Section IV, and in Section VI four applications are presented: tracking cars using a LIDAR, marine vessel tracking using X-band radar, tracking groups of pedestrians using a camera, and tracking complex shapes using a RGB-D sensor. The article is concluded in Section VII.

## II. DEFINITIONS AND EXTENDED OBJECT MODELLING

In this section we will first give a definition of the extended object tracking problem and some related types of object tracking. We will then overview extended object state modelling, measurement modelling, shape modelling, and dynamics modelling.

### A. Definitions

In tracking problems the physical, real-world-objects-of-interest always have spatial extents. This is true for relatively large objects-of-interest, like ships, boat, cars, bicyclists, humans and animals, and it is true for relatively small objects-of-interest, like cells. The differences between extended object tracking and point object tracking is due to sensor properties, especially the sensor resolution, rather than object properties such as spatial extent. If the resolution, relative to the size of the objects, is high enough, then an object may occupy several resolution cells. Thus, each object may generate multiple detections per time step in this case. In other words, depending on the sensor properties, specifically the sensor resolution, different types of object tracking will arise, and it is therefore instructive to distinguish between different types of object tracking problems. The following are definitions of types of tracking problems that are relevant to this article.

- *Point object tracking:*  
Each object generates at most a single measurement per time step, i.e., a single resolution cell is occupied by an object.
- *Extended object tracking:*  
Each object generates multiple measurements per time step and the measurements are spatially structured on the object, i.e., multiple resolution cells are occupied by an object.
- *Group object tracking:*  
A group object consists of two or more *subobjects* that share some common motion. Further, the objects are not tracked individually but are instead treated as a single entity. Thus, the group object occupies several resolution cells; each subobject may occupy either one or several resolution cells.

- *Tracking with multi-path propagation:*

Each object generates multiple measurements per time step that are due to multi-path propagation. Thus, the measurements are not spatially structured around the object.

All of the tracking approaches, except for point object tracking, assume the possibility of multiple measurements per target. Due to the required differences in motion and measurement modelling, we differentiate between the three tracking approaches rather than defining a single type called *multi-detection tracking*. Most literature considers one type of tracking problem, however, for the same sensor it can be the case that when an object is far away from the sensor it occupies at most one resolution cell, but when it is closer to the sensor it occupies several resolution cells.

The focus of the article lies on extended object tracking. However, we note that it is possible—and quite common—to employ extended object tracking methods to track the shape of a group object, see, e.g., [132] and the example in Section VI-A. It is easy to see that extended object tracking and group object tracking are two very similar problems. However, some distinctions can be made that warrant two definitions instead of just one.

In extended object tracking, each object is a single entity, e.g., a car, an airplane, a human, or an animal. Often the shape can be assumed to be a rigid body,<sup>2</sup> however, extended objects with deformable extents are also possible. In group object tracking, each object is a collection of (smaller) objects that share some common dynamics, while still allowing for individual dynamics within the group. For example, in a group of pedestrians, there is an overall group motion, but the individual pedestrians may also shift their positions within the group.

The measurements from an extended object are caused by *measurement sources*, which has different meaning depending on the sensor that is used and the types of objects that are tracked. In some cases, e.g., see [25], [26], [91], one can model a finite number of measurement sources, while in other cases it is better to model an infinite number of sources. For example, in [91] automotive radar is used to track cars, and the measurements are located around the wheelhouses of the tracked cars, i.e., there are four measurement sources. In [165], [167] LIDARs are used to track cars, and the measurements are then located on the chassi of the car. This can be interpreted as an infinite number of points that may act as measurement sources.

Note that certain sensors measure the object's cross-range and down-range extents (or similar object features), allowing for the extent (size and shape) of the object to be estimated, see e.g., [1], [162], [179–181],

<sup>2</sup>With the exception of the orientation of the extent, the size and shape of the object does not change over time.

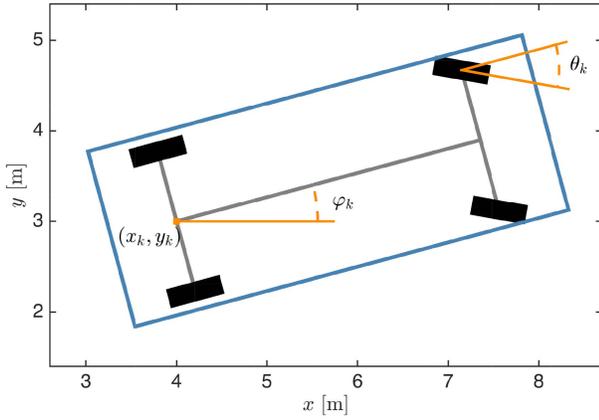


Fig. 2. Example illustration of car state. The state vector  $\mathbf{x}$  models position  $x, y$ , velocity  $v$ , heading  $\varphi$ , turning-angle  $\theta$ , length  $\ell$ , and width  $w$ . Note that velocity, length, and width are not marked in the illustration.

[223]. However, by the definitions used here, this is not extended object tracking unless there are multiple such measurements.

Lastly, multi-path phenomenon occur, e.g., when data from over-the-horizon-radar (OTHR) is used, see, e.g., [90], [164], [184]. An important difference between extended object tracking and tracking with multi-path phenomenon lies in the distribution of the measurements: for the plain multi-path problem a spatial distribution is not assumed.

## B. Object state

The extended object state models where the object is located, where it is going, and what its spatial extent (shape and size) is. The state typically includes the following:

- *Position*: Either  $(x, y)$ -position in 2D or  $(x, y, z)$ -position in 3D.
- *Kinematic state*: The motion parameters of the object, such as velocity, acceleration, heading, and turn-rate.
- *Extent state*: Parameters that determine the shape and the size of the object, as well as the orientation of the shape.

An example object state, appropriate for a car that is tracked using a horizontally mounted 2D LIDAR sensor [85], is illustrated in Figure 2. In this example the state vector at time step  $k$ , denoted  $\mathbf{x}_k$ , is

$$\mathbf{x}_k = [x_k \ y_k \ v_k \ \varphi_k \ \theta_k \ \ell_k \ w_k]^T \quad (1)$$

where  $x_k, y_k$  is 2D position, the kinematic state is comprised by velocity  $v_k$ , heading  $\varphi_k$  and turning angle  $\theta_k$ , and the extent state is comprised by length  $\ell_k$  and width  $w_k$ . Note that the shape of the car is assumed to be a rectangle, and the orientation of this rectangular shape is assumed to be aligned with the heading of the car. This state model is used in the car tracking example that is presented in Section VI-C.

In general, exactly what parameters the object state includes—e.g., 2D or 3D position? Which kinematics? Any assumed shape?—depends very much on the type of object that is tracked, the type of sensor data that is used, and the type(s) of object motion that one wishes to describe.

For example, for tracking cars it is often sufficient to only model the 2D position on the road, while airborne objects typically require 3D position. The position state may coincide with the objects centre-of-mass, however, this is not always the appropriate choice. When cars are tracked it is suitable to take the position as the mid-point on the rear-axle, because this facilitates the use of single-track-bicycle models in the motion modelling. Motion modelling, or dynamic modelling, for extended objects is address further in Section II-E.

If 2D position is modelled, the heading/orientation of the object can be described by a single angle, while 3D position may require more angles to accurately describe the heading/orientation, e.g., roll, pitch, and yaw angles. Often the orientation of the extent is aligned with the heading, however, this is not always the choice. For example, some motion models for cars include a so called slip angle that describes the angular difference between the heading of the car and the orientation of the shape of the car, see, e.g., [168] for an introduction to vehicle dynamics modelling.

The extent state is determined by the type of shape that one wishes to describe; it could be a simple geometric shape like the rectangle used in Figure 2, or it could be a more general shape. There are many different alternatives for this, and an overview is given in Section II-D.

## C. Measurement modelling

Depending on what type of sensor is used, where the measured object is located w.r.t. the sensor, and how the object is oriented, the sensor will produce a different number of detections, originating from different points on the object. In addition to this, sensor noise will affect the detections, and all these properties have to be taken into account in the measurement modelling.

An example with real-world LIDAR data is given in Figure 3. Here the 2D-LIDAR was used to track a car; in the Figure LIDAR detections from three different time steps are shown. We can see that the number of detections, as well as their locations relative to the target, changes with the sensor-to-target geometry.

Due to sensor noise and model uncertainties, the measurement modelling is typically handled using probabilistic tools. Let the extended object state be denoted  $\mathbf{x}$ , and let

$$\mathbf{Z} = \{\mathbf{z}^{(j)}\}_{j=1}^n \quad (2)$$

be a set of measurements that were caused by the object. Modelling the extended object measurements means to model the conditional distribution

$$p(\mathbf{Z} | \mathbf{x}), \quad (3)$$

often referred to as the extended object measurement likelihood. The likelihood (3) needs to capture the num-

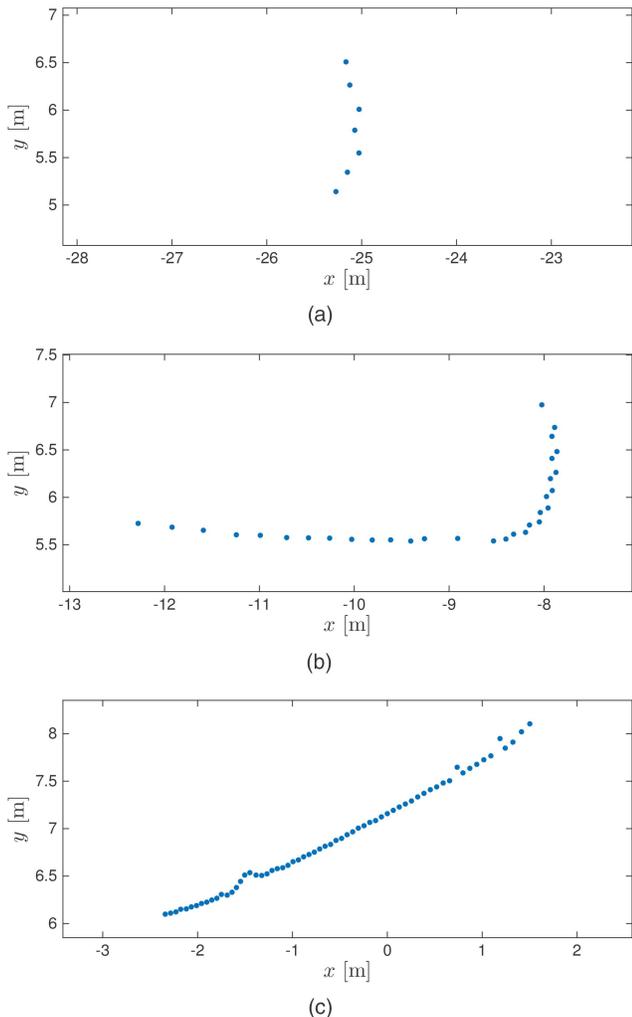


Fig. 3. Example of real-world LIDAR detections. The sensor is located in the origin, the measured object is a car. a)–c) shows detections from the same car from three different time steps. When the sensor-to-target geometry changes, the set of detections changes. In a) only the front side of the car is visible to the sensor, and the detections form a line. In b) the front and right sides are visible, and the detections (approximately) form an  $L$ -shape. In c) only the right side is visible to the sensor. Note that the car is farthest from the sensor in a), and closest in c).

ber of detections, and how the detections are spatially distributed around the target state  $\mathbf{x}$ . This modelling can be approached in several different ways; we overview the most common ways in the following.

1) Set of points on a rigid body: One way is to model that the extended object has some number  $L$  of reflection points<sup>3</sup> located on a rigid body shape, as described

<sup>3</sup>For some sensors, e.g., high resolution radar, the term *scattering* point may be a more accurate description of the underlying sensor properties. Further, reflection *source* may be a more accurate terminology in some cases, because the reflector may not be a discrete point but a larger structure, e.g., in automotive radar where the entire side of the car can be a reflector [29]. However, reflection point appears to be the more common expression in extended object tracking literature, so in the remainder of the paper we adhere to this terminology.

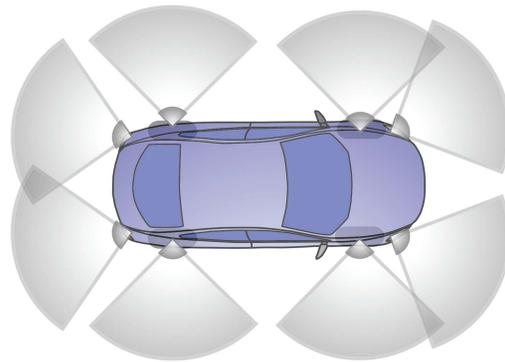


Fig. 4. Car with eight modelled radar reflection points: four points on the corners of the car, and four points on the wheel-houses. Also illustrated are the visibility regions. Image courtesy of Hammarstrand et al. [92].

in, e.g., [125, Sec. 12.7.1]. We denote this as a Set of Points on a Rigid Body (SPRB) model.

In SPRB models the reflection points are detected independently of each other, and the  $\ell$ th reflection point has a detection probability  $p_D^\ell$  that is a function of the object state. The measurement likelihood is [125, Eq. 12.208]

$$p(\mathbf{Z} | \mathbf{x}) = \sum_{\theta} \prod_{\theta_i=0} (1 - p_D^\ell) \prod_{\theta_i>0} p_D^\ell p^\ell(\mathbf{z}^{\theta_i} | \mathbf{x}) \quad (4)$$

if  $|\mathbf{Z}| \leq L$  and  $p(\mathbf{Z} | \mathbf{x}) = 0$  otherwise. Here  $|\mathbf{Z}|$  is the cardinality of the measurement set, and  $\theta$  is an assignment variable.<sup>4</sup> In mathematical terms, the measurement process for each reflection point can be described as a Bernoulli RFS [125], [127], and the measurement process for the extended object is a multi-Bernoulli RFS [125], [127].

SPRB models were used in some early work on extended object tracking, see, e.g., [27], [28], [39], [96], and were applied to data from vision sensors [27], [28]. SPRB modelling has also been applied to automotive radar, e.g., to model the reflection points on cars [29], [88], [92]. An illustration of the  $L = 8$  automotive radar reflection points modelled in [88], [92] is shown in Figure 4.

A challenge with the SPRB approach is that in a Bayesian estimation setting it requires data association between the  $L$  points on the extended object and the target detections, see the summation over the assignments  $\theta$  in (4). This association problem can be quite challenging in settings where the number of points, and their respective locations on the object, are (highly) uncertain. There are some standard methods for handling association problems, such as finding the best assignment using the auction algorithm [20], finding the  $M$  best

<sup>4</sup> $\theta_\ell = 0$  means that the  $\ell$ th point is not associated to any measurement, and  $\theta_\ell = j$  means that the  $\ell$ th point is associated to the  $j$ th measurement. Each measurement in  $\mathbf{Z}$  is associated to one of the reflection points, however, no reflection point is associated to more than one measurement.

assignments using Murty’s algorithm [135], or computing marginal association probabilities using, e.g., Probabilistic Data Association (PDA) [4] or fast-PDA [57]. A framework for handling the association uncertainty when automotive radar is used to track a single extended object is presented in [91]. In [25], [26], the association problem for the SPRB approach is by-passed by allowing more than one measurement from a point on the extended object and using the expectation maximization (EM) algorithm.

2) Spatial model: It was proposed by Gilholm et al. [66], [67] to model the target detections by an inhomogeneous Poisson Point Process (PPP). This models the number of detections as Poisson distributed with a rate  $\gamma(\mathbf{x})$  that is a function of the object’s state, and the detections are spatially distributed around the target. By this means, the data association problem is entirely avoided. The name *spatial model* derives from the assumption that the detections are spatially distributed. In this model the measurement likelihood is [125, Eq. 12.216]

$$p(\mathbf{Z} | \mathbf{x}) = e^{-\gamma(\mathbf{x})} \gamma(\mathbf{x})^{|\mathbf{Z}|} \prod_{z \in \mathbf{Z}} p(z | \mathbf{x}). \quad (5)$$

Using a PPP model is motivated in part by mathematical convenience—it is simple to use in both single object and multiple object scenarios, and avoiding an explicit summation over associations between measurements and points on the object is very attractive [66], [67].

The single measurement likelihood  $p(z | \mathbf{x})$  in (5) is called spatial distribution, and it captures the structure of the measurements by using a model of the object extent and a model of the sensor noise. One alternative is to model  $p(z | \mathbf{x})$  directly, e.g., using physics based modelling of the sensor. Another alternative is to model each detection  $\mathbf{z}$  as a noisy measurement of a source  $\mathbf{y}$  located somewhere on the object. The distribution  $p(\mathbf{z} | \mathbf{y})$  models the sensor noise, the distribution  $p(\mathbf{y} | \mathbf{x})$  models the extent and the spatial distribution  $p(\mathbf{z} | \mathbf{x})$  is given by the convolution

$$p(\mathbf{z} | \mathbf{x}) = \int p(\mathbf{z} | \mathbf{y}) p(\mathbf{y} | \mathbf{x}) d\mathbf{y}. \quad (6)$$

In other words, the measurement likelihood (6) is the marginalization of the reflection point  $\mathbf{y}$  out of the estimation problem. For the noise model  $p(\mathbf{z} | \mathbf{y})$  the Gaussian distribution is a common choice, however, other noise models are possible. An appropriate choice for the measurement source distribution  $p(\mathbf{y} | \mathbf{x})$  depends heavily on the type of sensor that is used and the representation of the object’s shape.

In [130, Sec. 2.3] the PPP model (5) is interpreted to imply that the extended object is far enough away from the sensor for the measurements to resemble a cluster of points, rather than a structured ensemble. However, the PPP model has been used successfully in multiple

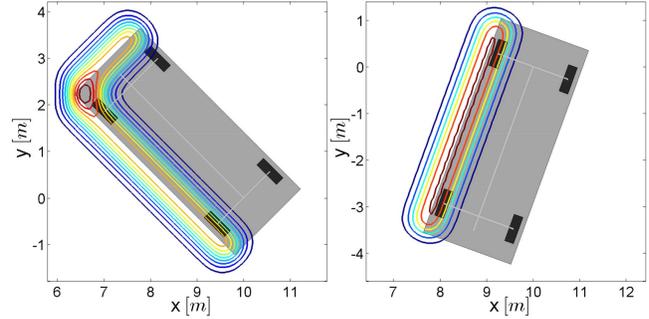


Fig. 5. Example of the spatial measurement model. The sensor is a 2D LIDAR located at the origin, and the tracked object is a car. The sensor can either receive measurements from two sides (example on left), or measurements from one side (example on right).

object scenarios where the object measurements show a high degree of structure, see, e.g., [70], [78], [85].

Multiple extended target tracking using the PPP model (5) has shown that the tracking results are sensitive to the state dependent Poisson rate  $\gamma(\mathbf{x})$ , see [74]. The Bayesian conjugate prior for an unknown Poisson rate is the gamma distribution, see, e.g., [64]. By augmenting the state distribution with a gamma distribution for the Poisson rate, an individual Poisson rate can be estimated for each extended object [79].

In [85] the PPP spatial model was used to track cars using data from a 2D LIDAR. The cars were modelled as rectangularly shaped, see (1) and Figure 2. The measurement modelling can be simplified by assuming that the LIDAR measurements are located along either one side of the assumed rectangular car, or along two sides. Example measurement likelihoods for these two cases are shown in Figure 5. The source density  $p(\mathbf{y} | \mathbf{x})$  is assumed uniform along the sides that are visible to the sensor, and a Gaussian density was used for the noise  $p(\mathbf{z} | \mathbf{y})$ .

A second alternative to the SPRB model with  $L$  reflection points is to use a spatial model where the number of detections is binomial distributed with parameters  $L$  and  $p_D$  [159], [160], i.e., there is an implicit assumption that the probabilities of detection are equal for all  $L$  points,  $p_D^\ell \equiv p_D, \forall \ell$ . As in the PPP model, the detections are spatially distributed around the target state. The measurement likelihood is [159, Eq. 5]

$$p(\mathbf{Z} | \mathbf{x}) = \frac{L!}{(L - |\mathbf{Z}|)!} p_D^{|\mathbf{Z}|} (1 - p_D)^{L - |\mathbf{Z}|} \prod_{z \in \mathbf{Z}} p(z | \mathbf{x}). \quad (7)$$

if  $|\mathbf{Z}| \leq L$  and  $p(\mathbf{Z} | \mathbf{x}) = 0$  otherwise. Note the considerable similarity to (5): the difference is in the assumed model for the number of detections, and the single measurement likelihood  $p(z | \mathbf{x})$  in (7) is analogous to  $p(z | \mathbf{x})$  in (5). For known  $L$ , the conjugate prior for an unknown  $p_D$  is the beta distribution. Bayesian approaches to estimating unknown  $L$  given a known  $p_D$ , or estimating both  $L$  and  $p_D$ , have to the best of our knowledge not been presented. However, a simple heuristic for

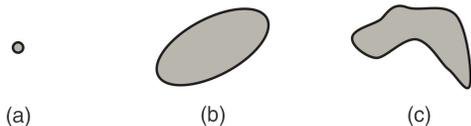


Fig. 6. Illustration of the three levels of shape complexity. a) No shape model is used, the point corresponds to, e.g., the centre-of-mass. b) A basic geometric shape, such as an ellipse, is used to represent the extent of the target. c) An arbitrary shape model is used for the extent of the target.

determining  $L$ , under the assumption that  $p_D$  is known, is given in [159].

In [66], [67] the Poisson assumption for the number of detections is not given much motivation using direct physical modelling of sensor properties. Similarly, in [159], [160] there is no physical modelling of sensor properties to motivate the binomial distribution model for the number of detections. Indeed, both models may be crude approximations for some sensor types, e.g., LIDAR. Nevertheless, experiments with real-world data show that both models are applicable to many different sensor types, regardless of whether or not the number of detections are actually Poisson/binomial distributed. The PPP model has been used successfully with data from LIDAR [70], [78], [85], radar [75], [76], and camera (see Section VI-A). The binomial model has been used successfully with camera data [159], [160].

3) Physics based modelling: In [29], [88], [91] SPRB models for car tracking using automotive radars are derived using a physics based approach. Naturally, it is possible to use physical modelling of the sensor properties—both the modelling of the number of detections, and the modelling of the single measurement likelihood—to derive models that do not fit into the SPRB model or the spatial model. For example, for a high resolution radar the number of measurements and their locations in the range-Doppler image can be reasonably predicted by deterministic electromagnetic theory, see, e.g., [21]. In [100] automotive radars are modelled using direct scattering, and this model is integrated into a multi-object framework in [166]. LIDAR sensors can be modelled precisely using ray-tracing [148] which facilitates the integration into multi-object tracking algorithms using the separable likelihood approach [167].

#### D. Shape modelling

When it comes to modelling the shape of the object, it is useful to distinguish different complexity levels for describing the shape, because different shape complexities might require different approximations and algorithms. The different ways to model this type of extended object tracking scenario are here divided into three complexity levels:

TABLE I  
Object shape (2D in 2D-space)

Stick	[7], [24], [67], [70], [186]
Circle	[11], [145], [146]
Ellipse	[2], [12], [38], [73], [102], [108], [155], [157], [171], [224]
Rectangle	[73], [85], [100]
Arbitrary shape	[9], [32], [86], [95], [109], [111], [121], [198]

- 1) The simplest level of modelling is to not model the shape at all, i.e., to only estimate the object’s kinematic properties. This approach has lowest computational complexity and the flexibility to track different type of objects is high because this model, even though it is simplistic in terms of object shape, is often applicable (with varying degree of accuracy).
- 2) A more advanced level of modelling is to assume a specific basic geometric shape for the object, such as an ellipse, a line, or a rectangle.
- 3) The most advanced approach is to construct a measurement model that is capable of handling a broad variety of both different shapes and different measurement appearances. While such a model would be most general, it could also prove to be overly computationally complex.

The three complexity levels are illustrated in Figure 6, and some references whose shape modelling fall into the latter two categories are listed in Table I.

The correct choice of complexity level is challenging and does not have a simple answer. In general, the more complex the shape, the more measurements (with less noise) are required to get a reasonable shape estimate. Furthermore, it depends on the type of sensor that is used, the types of objects, their motions, and what the tracking output will be used for. In some scenarios it may be sufficient to know the position of each object, in other scenarios it is necessary to have a detailed estimate of the size and shape of each object.

For example, in [70] it is shown that using LIDAR data bicycles can be tracked fairly accurately without modelling the extent. However, estimation performance<sup>5</sup> is improved by using a spatial distribution model where the measurement source distribution, cf.  $p(\mathbf{y} | \mathbf{x})$  in (6), is modelled by a stick shape and uniform distribution and the noise distribution, cf.  $p(\mathbf{z} | \mathbf{y})$  in (6), is modelled by a Gaussian distribution. Specifically, by modelling the shape it becomes possible to capture rotations of the shape, and thus capture the onset of turning maneuvers. Without a shape estimate, the turning is captured at a later time [70].

The 2D-LIDAR bicycle tracking results are also an example of how a simple geometric shape, in this case a stick, combined with a simple Gaussian noise model, is a suitable measurement likelihood. A 2D stick shape

<sup>5</sup>Video with tracking results: <https://youtu.be/sGTGNkrpts>.

is a crude approximation of the way a person riding a bicycle looks from a top-down perspective, however, here the stick shape is intended to model the measurement likelihood, and is not intended to be a nice visualization of the tracked bicyclist. Similarly, a rectangle shape is suitable when 2D-LIDAR is used to track cars, see, e.g., [73], [85], [148], even though many cars are only approximately rectangular. Another example is the ellipse-shape that is used to track boats and ships using marine radar in, e.g., [75], [76], [188–190]. Typically neither boats, nor ships, are shaped like ellipses, however, the ellipse shape is suitable for the measurement modelling, and the estimated major and minor axes of the object ellipses are accurate estimates of the real-world lengths and widths of the boats/ships [188–190].

In some scenarios the objects have extents with shapes that cannot accurately be represented by a simple geometric shape like an ellipse or a rectangle. For estimation of arbitrary object shapes, the literature contains at least two different types of approaches: either the shape is modelled as a curve with some parametrization [9], [32], [95], [121], [198], or the shape is modelled as combination of ellipses [86], [109], [111]. When the shape is given a curve parametrization the noisy detections can be modelled using Gaussian processes [95], [198]. Applied to car tracking using 2D-LIDAR [95], [198], this allows for shape modelling with rounded corners, which is a more accurate model of actual cars than a rectangle with sharp corners is. The price of a more accurate model is an increased complexity: a general shape requires more parameters than a simple geometric shape.

The increased complexity can be alleviated by utilizing the prior knowledge that cars are symmetric, see [51] for a general concept to incorporate symmetries and [95] for a Gaussian process model example. Another approach to handling the complexity is to use different models at different distances from the sensor; in [206] the priority of objects is ranked in three groups, specifying how accurately the different objects should be modelled. For example, for collision avoidance in autonomous driving, the objects closest to the ego-vehicle are more important than the distant objects, and this justifies “taking” computational resources from the distant objects and “spending” it on the closer objects.

In addition to modelling the shape itself, there are different ways to model how the measurements are spatially distributed over the shape. The types of extended object spatial distributions can be divided into two classes:

- Measurements along the boundary of the object’s extent. For measurements in 2D, this means that the measurements are noisy points on a curve. For measurements in 3D, the measurements are noisy points on either a curve or a surface. Measurements along the boundary are obtained, e.g., when LIDAR is used in automotive applications.

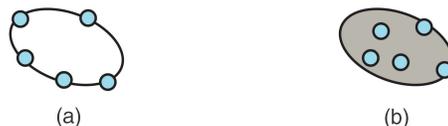


Fig. 7. a) Measurements from the boundary b) Measurements from the surface.

TABLE II  
Shape Dimensions

Curve in 2D/3D space:	[7], [24], [67], [70], [150], [222]
Surface in 2D space:	[12], [38], [73], [102], [108], [146], [155], [157], [171], [224]
Surface in 3D space:	[48], [52]

- Measurements inside the object’s extent, i.e., the measurements form a cluster. For example, two-dimensional radar detections of marine vessels can be interpreted as measurements from the inner of a two-dimensional shape, e.g., an ellipse, see [76] and Section VI-B for an experimental example.

In Table II some references are listed according to the shape dimension and measurement type, and Figure 7 provides an illustration. To our knowledge there is no explicit work about the estimation of 3D shapes in 3D space, probably because there are rarely sensors for this case. However, most algorithms for 2D shapes in 2D space can be generalized rather easily to the 3D case.

When the measurements lie on the boundary of the extended object, the resulting theoretical problem shares similarities with traditional curve fitting, where a curve is to be matched with noisy points [34], [58]. However, the curve fitting problem only considers static scenarios, i.e., non-moving curves. Additionally, the noise is usually isotropic and non-recursive non-Bayesian methods have been developed. Hence, curve fitting algorithms usually cannot directly be applied in the extended object tracking context. For a discussion of the rare Kalman filter-based approaches for curve fitting, we refer to [150], [222].

To summarize the discussion about shape modelling, we note that it is important that the shape model is not only a reasonable representation of the true object shape but is also suitable for the measurement modelling, and that the shape model has a complexity that is appropriate for the sensor, the tracked object, and the computational resources.

## E. Dynamics modelling

The object dynamic model describes how the object state evolves over time; for a moving object this describes how the object moves. This involves the position and the kinematic states that describe the motion—e.g., velocity, acceleration, turn-rate—however, it also involves descriptions of how the extent changes over time (typically it rotates when the object turns) and how

the number of measurements changes over time (often there are more measurements the closer to the sensor the object is).

There are two probabilistic parts to dynamics modelling that are important: the transition density and the Chapman-Kolmogorov equation. The transition density is denoted

$$p(\mathbf{x}_{k+1} | \mathbf{x}_k), \quad (8)$$

and describes the transition of the state from time step  $k$  to time step  $k + 1$ , i.e., from  $\mathbf{x}_k$  to  $\mathbf{x}_{k+1}$ . The Chapman-Kolmogorov equation

$$p(\mathbf{x}_{k+1}) = \int p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k) d\mathbf{x}_k. \quad (9)$$

describes how, given a prior state density  $p(\mathbf{x}_k)$  and a transition density, the predicted density  $p(\mathbf{x}_{k+1})$  is computed.

In many cases the dynamics for the position and the kinematic states can be modelled using any of the models that are standard in point object tracking, see [115] for a comprehensive overview. Examples include the constant velocity (CV) model, the constant acceleration (CA) model, and the coordinated, or constant, turn (CT) model. Detailed descriptions of CV, CA and CT models are given in [115]. When the tracked objects are cars, so called bicycle-models, introduced in [158], are suitable for describing the target motion, see, e.g., [168, Ch. 10–11] for an overview and introduction to bicycle-models.

When the extended object is a rigid body its size and shape does not change over time, however, the orientation of the shape (typically) rotates when the object turns. If the object is described by a set of points on a rigid body, see Section II-C.1, the point of rotation must be specified, and the centre-of-mass is a suitable choice. For the more common spatial models, see Section II-C.2, a typical assumption for the extent is to assume that its orientation is aligned with the heading of the object, e.g., this is the case in the bicycle models that are used in [70], [85]. When the heading and orientation are aligned the rotation of the extent does not have to be explicitly modelled as it is implicitly modelled by the object’s heading. However, if this is not the case, the point of the rotation must be specified—again a suitable choice is the object’s centre-of-mass.

When there are multiple objects present a common assumption is that the objects evolve independently of each other, resulting in the object estimates being predicted independently. Obviously, an independent prediction may result in physically impossible (e.g., overlapping/intersecting) object state estimates. To better model target interactions one can use, e.g., social force modelling [93]; this is done in [155], where LIDAR is used to track pedestrians. In group object tracking, where several objects form groups while remaining distinguishable, it is possible to apply, e.g., leader-follower

models, allowing for the individual objects to be predicted dependently, see e.g., [35], [143]. A Markov Chain Monte Carlo (MCMC) approach to inferring interaction strengths between targets in groups is presented in [134].

### III. TRACKING A SINGLE EXTENDED OBJECT

In this section we overview some widely-used approaches for single extended object tracking, namely random matrix models and star-convex models.

#### A. Random Matrix Approach

The random matrix model was originally proposed by Koch [102], and is an example of a spatial model (Section II-C.2). It models the extended object state as the combination of a kinematic state vector  $\mathbf{x}_k$  and an extent matrix<sup>6</sup>  $X_k$ . The vector  $\mathbf{x}_k$  represents the object’s position and its motion properties, such as velocity, acceleration, and turn-rate. The  $d \times d$  matrix  $X_k$  represents the object’s extent, where  $d$  is the dimension of the object;  $d = 2$  for tracking with 2D position and  $d = 3$  for tracking with 3D position. The matrix  $X_k$  is modelled as being symmetric and positive definite, which implies that the object shape is approximated by an ellipse. The ellipse shape may seem limiting, however, the model is applicable to many real scenarios, e.g., pedestrian tracking using LIDAR [78] and tracking of boats and ships using marine radar [75], [76], [171], [188]–[190].

1) Original measurement model: In the original model [102] the measurements are assumed independent, and conditioned on the object state  $\mathbf{x}_k, X_k$  the single measurement likelihood—cf. (5), (7)—is modelled as Gaussian,

$$p(\mathbf{z}_k | \mathbf{x}_k, X_k) = \mathcal{N}(\mathbf{z}_k; (H_k \otimes \mathbf{I}_d)\mathbf{x}_k, X_k). \quad (10a)$$

where  $\otimes$  is the Kronecker product,  $\mathbf{I}_d$  is an identity matrix of the same dimensions as the extent, the noise covariance matrix is the extent matrix, and  $(H_k \otimes \mathbf{I}_d)$  is a measurement model that picks out the Cartesian position from the kinematic vector  $\mathbf{x}_k$ .

For Gaussian measurements, the conjugate priors for unknown mean and covariance are the Gaussian and the inverse Wishart distributions, respectively. This motivates the object state distribution [102]

$$p(\mathbf{x}_k, X_k | \mathbf{Z}^k) = p(\mathbf{x}_k | X_k, \mathbf{Z}^k) p(X_k | \mathbf{Z}^k) \quad (10b)$$

$$= \mathcal{N}(\mathbf{x}_k; m_{k|k}, P_{k|k} \otimes X_k) \times \mathcal{IW}_d(X_k; v_{k|k}, V_{k|k}), \quad (10c)$$

<sup>6</sup>The book by Gupta and Nagar [89] is a good reference for various matrix variate distributions.

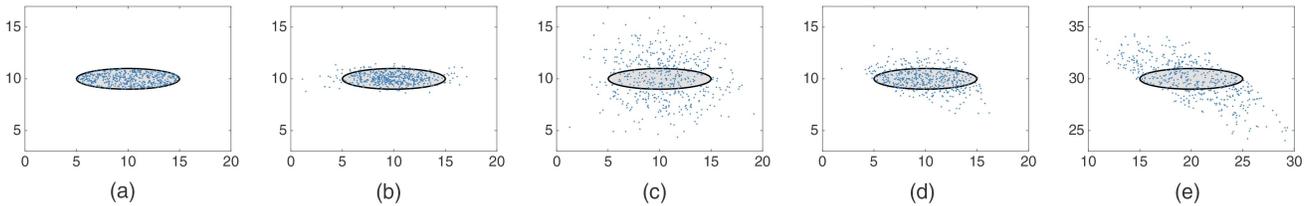


Fig. 8. Illustration of the random matrix measurement model. The sensor is located in the origin. a) Uniform reflection points, no noise. b) Gaussian approximation of uniform distribution. c) Uniform reflection points, Cartesian Gaussian noise. d)–e) Uniform reflection points, polar Gaussian noise. Note how the spread due to noise is larger when the object is further away (e).

where the kinematic vector is Gaussian distributed with mean  $m_{k|k}$  and covariance  $P_{k|k} \otimes X_k$ , and the extent matrix is inverse Wishart distributed with  $v_{k|k}$  degrees of freedom and scale matrix  $V_{k|k}$ . Owing to the specific form of the conditional Gaussian distribution, where the covariance is the Kronecker product of a matrix  $P_{k|k}$  and the extent matrix, non-linear dynamics, such as turn-rate, can not be included in the kinematic vector. In this model the kinematic state  $\mathbf{x}_k$  is limited to consist of a spatial state component  $\mathbf{r}_k$  that represents the center of mass (i.e., the object's position), and derivatives of  $\mathbf{r}_k$  (typically velocity and acceleration, although higher derivatives are possible) [102]. It follows from this that the motion modelling for the kinematic state is linear [102], see further in Section III-A.4.

The measurement update is linear without approximation [102], the details are given in Table III. For the kinematic state a Kalman-filter-like update is performed, and the extent state is updated with two matrices  $N$  and  $Z$ , where the matrix  $N$  is proportional to the spread of the centroid measurement  $\bar{\mathbf{z}}$  (mean measurement) around the predicted centroid  $(H_k \otimes \mathbf{I}_d)m$ , and the matrix  $Z$  is proportional to the sum of the spreads of the measurements around the centroid measurement.

2) Improved noise modelling: An implicit assumption of the original random matrix model (10) is that the measurement noise is negligible compared to the extent. In some scenarios this assumption does not hold, for example when marine X-band radar is used [188]. If the measurement noise is not modelled properly the filtering will lead to a biased estimate, see, e.g., [76].

To alleviate this problem Feldmann et al. [54]–[56] suggested to use a measurement likelihood that is a convolution of a source distribution and a noise distribution, see (6). The noise is modelled as zero mean Gaussian with constant covariance,

$$p(\mathbf{z}_k | \mathbf{y}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{y}_k, R), \quad (11)$$

and the measurement sources are modelled as uniformly distributed on the object,

$$p(\mathbf{y}_k | \mathbf{x}_k, X_k) = \mathcal{U}(\mathbf{y}_k; \mathbf{x}_k, X_k). \quad (12)$$

A uniform distribution is appropriate, e.g., when marine radar is used to track boats and ships, see [75], [76], [188]–[190]. The drawback of the uniform distribution is that the convolution (6) is not analytically tractable.

It is shown in [56] that for an elliptically shaped object the uniform distribution (12) can be approximated by a Gaussian distribution

$$p(\mathbf{y}_k | \mathbf{x}_k, X_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}_k \mathbf{x}_k, zX_k) \quad (13)$$

where  $z$  is a scaling factor and  $\mathbf{H}_k$  is a measurement model that picks out the position. A simulation study in [56] showed that  $z = 1/4$  is a good parameter setting; this result is experimentally verified in [188]. The difference between the uniform distribution (12) and its Gaussian approximation (13) is illustrated in Figure 8, see subfigures a and b.

With the Gaussian noise model (11) and the Gaussian approximation (13) the solution to the convolution (6) is

$$p(\mathbf{z}_k | \mathbf{x}_k, X_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k, zX_k + R). \quad (14)$$

An example with elliptic extent  $X$  and circular measurement noise covariance  $R$  is given in Figure 8, see subfigure c. The inclusion of the constant noise matrix  $R$  means that, with a Gaussian inverse Wishart prior of the form (10), the update is no longer analytically tractable. Feldmann et al. [54]–[56] proposed to approach this by modelling the extended object state with a factorised state density

$$p(\mathbf{x}_k, X_k | \mathbf{Z}^k) = p(\mathbf{x}_k | \mathbf{Z}^k) p(X_k | \mathbf{Z}^k) \quad (15a)$$

$$= \mathcal{N}(\mathbf{x}_k; m_{k|k}, P_{k|k}) \times \mathcal{IW}_d(X_k; v_{k|k}, V_{k|k}). \quad (15b)$$

Note the assumed independence between the kinematic state  $\mathbf{x}_k$  and  $X_k$  in (15b), an assumption that cannot be fully theoretically justified.<sup>7</sup>

Despite this theoretical drawback of a factorised density (15), there are some practical advantages to using the state distribution (15b), instead of (10c). The factorised model allows for a more general class of kinematic state vectors  $\mathbf{x}_k$ , e.g., including non-linear dynamics such as heading and turn-rate, and the Gaussian covariance is no longer intertwined with the extent matrix. Further, the measurement model is better when the size of the extent and the size of the sensor noise are within the same order of magnitude [56]. The assumed independence between  $\mathbf{x}_k$  and  $X_k$  is alleviated in practice

<sup>7</sup>After updating with a set of measurements  $\mathbf{Z}$  the kinematic state  $\mathbf{x}$  and extent state  $X$  are necessarily dependent.

TABLE III  
Random matrix update from [102]

**Input:** Parameters  $m, P, v, V$  of conditional state density (10), measurement model  $H$ , set of detections  $\mathbf{W}$ ,  $n = |\mathbf{W}|$

**Output:** Updated parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= m + (K \otimes \mathbf{I}_d)\varepsilon \\ P_+ &= P - KSK^T \\ v_+ &= v + n \\ V_+ &= V + N + Z \\ \varepsilon &= \bar{\mathbf{z}} - (H \otimes \mathbf{I}_d)m \\ \bar{\mathbf{z}} &= \frac{1}{n} \sum_{\mathbf{z}^i \in \mathbf{W}} \mathbf{z}^i \\ Z &= \sum_{\mathbf{z}^i \in \mathbf{W}} (\mathbf{z}^i - \bar{\mathbf{z}})(\mathbf{z}^i - \bar{\mathbf{z}})^T \\ S &= HPH^T + \frac{1}{n} \\ K &= PH^T S^{-1} \\ N &= S^{-1} \varepsilon \varepsilon^T \end{aligned}$$

by the measurement update which provides for the necessary interdependence between kinematics and extent estimation, see [56].

With the measurement likelihood (14) and the state density in (15) the updated extent estimate is unbiased, however, the measurement update requires approximation. The update presented in [56], for details see Table IV, is based on the assumption that the extent is approximately equal to the predicted estimate,

$$X_k \approx \hat{X}_{k|k-1} = E[X_k | \mathbf{Z}^{k-1}], \quad (16)$$

and on the approximation of non-linear functions of the extent using matrix square roots computed with Cholesky factorisation,  $\hat{X} = \hat{X}^{T/2} \hat{X}^{1/2}$ . After some clever approximations the update of the kinematic state is again a Kalman filter-like update, and the extent state shape matrix is again updated with two matrices  $\hat{N}$  and  $\hat{Z}$  proportional to the spreads around the predicted measurement and the centroid. Note that the difference to the original approach, see  $N$  and  $Z$  in Table III is in the scaling of the two matrices.

A simulation study in [56] shows that the noisy measurement model (14) and the factorised state model (15) does indeed outperform the original model (10) when the measurement noise is non-negligible. A performance analysis of the update in Table IV based on the posterior Cramér-Rao lower bounds can be found in [163].

For the models (14) and (15) two additional updates are presented in [3], [138]. The update presented in

TABLE IV  
Random matrix update from [56]

**Input:** Parameters  $m, P, v, V$  of factorised state density (15), measurement model  $\mathbf{H}$ , measurement noise covariance  $R$ , scaling factor  $z$ , set of detections  $\mathbf{W}$ ,  $n = |\mathbf{W}|$

**Output:** Updated parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= m + K\varepsilon \\ P_+ &= P - KSK^T \\ v_+ &= v + n \\ V_+ &= V + \hat{N} + \hat{Z} \\ \varepsilon &= \bar{\mathbf{z}} - \mathbf{H}m \\ \bar{\mathbf{z}} &= \frac{1}{n} \sum_{\mathbf{z}^i \in \mathbf{W}} \mathbf{z}^i \\ Z &= \sum_{\mathbf{z}^i \in \mathbf{W}} (\mathbf{z}^i - \bar{\mathbf{z}})(\mathbf{z}^i - \bar{\mathbf{z}})^T \\ S &= \mathbf{H}P\mathbf{H}^T + \frac{Y}{n} \\ K &= P\mathbf{H}^T S^{-1} \\ \hat{X} &= V(v - 2d - 2)^{-1} \\ Y &= z\hat{X} + R \\ \hat{N} &= \hat{X}^{1/2} S^{-1/2} \varepsilon \varepsilon^T (S^{-1/2})^T (\hat{X}^{1/2})^T \\ \hat{Z} &= \hat{X}^{1/2} Y^{-1/2} Z (Y^{-1/2})^T (\hat{X}^{1/2})^T \end{aligned}$$

[138] is based on variational Bayesian approximation,<sup>8</sup> where the unknown measurement sources  $\mathbf{y}$ , cf. (6), are estimated as so called hidden variables. The update is iterative, and can be run either for a fixed number of iterations, or until some convergence criterion is met. The details are given in Table V.

A simulation study in [138] shows that the variational update has smaller estimation error than the update based on Cholesky factorisation (Table IV), at the price of higher computational cost. It is reported that the update on average converges in 5 iterations, however, to be on the safe side 20 iterations were performed in each update in the simulation study [138].

An update based on linearisation of the natural logarithm of the measurement likelihood (14) is presented in [3], details are given in Table VI. A simulation study in [3] shows that the log-linearised update gives results that almost match the variational update, at a lower computational cost.

<sup>8</sup>Variational Bayes, or simply variational inference, is a type of approximate inference that builds upon approximating the true distribution with a factorised distribution, i.e., approximation under assumed independence. Thus, variational Bayes is a suitable estimation method for the state model (15b), since this model already makes the necessary factorisation assumption and approximates the distribution  $p(\mathbf{x}_k, X_k | \mathbf{Z}^k)$  with a factorised distribution  $p(\mathbf{x}_k | \mathbf{Z}^k)p(X_k | \mathbf{Z}^k)$ . Variational Bayes, and other approximate inference methods, are described further in, e.g., [22, Ch. 10].

TABLE V  
Random matrix update from [138]

**Input:** Parameters  $m, P, v, V$  of factorised state density (15), measurement model  $\mathbf{H}$ , measurement noise covariance  $R$ , scaling factor  $z$ , set of detections  $\mathbf{W}$ ,  $n = |\mathbf{W}|$

**Output:** Updated parameters  $m_+, P_+, v_+, V_+$   
Initialize

$$\begin{aligned} \mathbf{y}^{i,(0)} &= \mathbf{z}^i \\ \Sigma^{(0)} &= zV(v - 2d - 2)^{-1} \\ m_+^{(0)} &= m \\ P_+^{(0)} &= P \\ v_+ &= v + n \\ V_+^{(0)} &= V \end{aligned}$$

Iterate until convergence

$$\begin{aligned} \mathbf{y}^{i,(t+1)} &= \Sigma^{(t+1)}(v_+(zV_+^{(t)})^{-1}\mathbf{H}m_+^{(t)} + R^{-1}\mathbf{z}^i) \\ \Sigma^{(t+1)} &= (v_+(zV_+^{(t)})^{-1} + R^{-1})^{-1} \\ m_+^{(t+1)} &= P_+^{(t+1)} \left( P^{-1}m + n\mathbf{H}^T v_+(zV_+^{(t)})^{-1} \frac{1}{n} \sum_i \mathbf{y}^{i,(t)} \right) \\ P_+^{(t+1)} &= (P^{-1} + n\mathbf{H}^T v_+(zV_+^{(t)})^{-1} \mathbf{H})^{-1} \\ V_+^{(t+1)} &= V + \frac{1}{z} \sum_i (\mathbf{y}^{i,(t)} - \mathbf{H}m_+^{(t)})(\mathbf{y}^{i,(t)} - \mathbf{H}m_+^{(t)})^T \\ &\quad + \frac{n}{z} \mathbf{H}P^{(t)}\mathbf{H}^T + \frac{n}{z} \Sigma^{(t)} \end{aligned}$$

Output ( $T$  is the final iteration)

$$\begin{aligned} m_+ &= m_+^{(T)} \\ P_+ &= P_+^{(T)} \\ v_+ &= v + n \\ V_+ &= V_+^{(T)} \end{aligned}$$

To improve the measurement modelling for the original conditional state model (10c) the following measurement likelihood was proposed in [107], [108],

$$p(\mathbf{z}_k | \mathbf{x}_k, X_k) = \mathcal{N}(\mathbf{z}_k; (\mathbf{H}_k \otimes I)\mathbf{x}_k, B_k X_k B_k^T) \quad (17)$$

where  $B_k$  is a known parameter matrix. The update, see details in Table VII, builds upon the approximation [108, Eq. 28]

$$B_k X_k B_k^T \approx \gamma_k X_k \quad (18)$$

where  $\gamma_k$  is a scalar that is given by setting the determinants of both sides equal [108, Eq. 29]

$$\det(B_k X_k B_k^T) = \det(\gamma_k X_k) \Rightarrow \gamma_k = \det(B_k)^{2/d} \quad (19)$$

Under the assumption that the extent is approximately equal to the predicted estimate (16) the measurement model (17) can model additive Gaussian noise approximately by setting

$$B_k = (z\hat{X}_{k|k-1} + R)^{1/2} \hat{X}_{k|k-1}^{-1/2}. \quad (20)$$

TABLE VI  
Random matrix update from [3]

**Input:** Parameters  $m, P, v, V$  of factorised state density (15), measurement model  $\mathbf{H}$ , measurement noise covariance  $R$ , scaling factor  $z$ , set of detections  $\mathbf{W}$ ,  $n = |\mathbf{W}|$

**Output:** Updated parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= m + K\varepsilon \\ P_+ &= P - KSK^T \\ v_+ &= v + n \\ V_+ &= V + M \\ \varepsilon &= \bar{\mathbf{z}} - \mathbf{H}m \\ \bar{\mathbf{z}} &= \frac{1}{n} \sum_{\mathbf{z}^i \in \mathbf{W}} \mathbf{z}^i \\ Z &= \sum_{\mathbf{z}^i \in \mathbf{W}} (\mathbf{z}^i - \bar{\mathbf{z}})(\mathbf{z}^i - \bar{\mathbf{z}})^T \\ S &= \mathbf{H}P\mathbf{H}^T + \frac{z\hat{X} + R}{n} \\ K &= P\mathbf{H}^T S^{-1} \\ \hat{X} &= V(v - 2d - 2)^{-1} \\ C &= \mathbf{H}P\mathbf{H}^T + z\hat{X} + R \\ M &= n\hat{X} + nz\hat{X}C^{-1} \left( \frac{Z}{n} + \varepsilon\varepsilon^T - C \right) C^{-1} \hat{X} \end{aligned}$$

Note that similarly to the update presented in [56], this requires matrix square roots. In addition to modelling noise, the matrix  $B_k$  can be used to model distortion of the observed extent [108].

3) Non-linear measurements: Both the original measurement likelihood (10a) and the noise adapted measurement likelihoods (14) and (17) are linear with respect to the kinematic state  $\mathbf{x}_k$ , and the noise covariance in (14) and (17) is constant. However, when real-world data is used the measurement model is often non-linear, e.g., a radar measures range and azimuth to the object's position instead of measuring the position directly as in (10a) and (14). Further, due to the polar noise the noise covariance in Cartesian coordinates is not constant, but increases with increasing sensor-to-object distance.

In [188]–[190] non-linear radar measurements are handled by performing a polar to Cartesian conversion in a pre-processing step, and by modelling the the noise covariance  $R(\mathbf{y})$  as a function of the reflection point. The measurement noise model (11) is modified to

$$p(\mathbf{z}_k | \mathbf{y}_k, \mathbf{x}_k, X_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{y}_k, R(\mathbf{y}_k)). \quad (21)$$

After conversion to Cartesian coordinates the spread of the measurements due to noise is larger the further the object is from the sensor, see Figure 8, subfigures d and

TABLE VII  
Random matrix update from [107]

**Input:** Parameters  $m, P, v, V$  of conditional state density (10), measurement model  $H$ , parameter matrix  $B$ , set of detections  $\mathbf{W}$ ,  $n = |\mathbf{W}|$   
**Output:** Updated parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= m + (K \otimes \mathbf{I}_d)\varepsilon \\ P_+ &= P - KSK^T \\ v_+ &= v + n \\ V_+ &= V + N + \hat{Z} \\ \varepsilon &= \bar{\mathbf{z}} - (H \otimes \mathbf{I}_d)m \\ \bar{\mathbf{z}} &= \frac{1}{n} \sum_{\mathbf{z}^i \in \mathbf{W}} \mathbf{z}^i \\ Z &= \sum_{\mathbf{z}^i \in \mathbf{W}} (\mathbf{z}^i - \bar{\mathbf{z}})(\mathbf{z}^i - \bar{\mathbf{z}})^T \\ S &= HPH^T + \frac{\det(B)^{2/d}}{n} \\ K &= PH^T S^{-1} \\ N &= S^{-1} \varepsilon \varepsilon^T \\ \hat{Z} &= B^{-1} Z B^{-T} \end{aligned}$$

e. With the Gaussian noise model (21) and the Gaussian approximation (13), the convolution of the two (cf. (6))

$$\begin{aligned} p(\mathbf{z}_k | \mathbf{x}_k, X_k) \\ = \int \mathcal{N}(\mathbf{z}_k; \mathbf{y}_k, R(\mathbf{y}_k)) \mathcal{N}(\mathbf{y}_k; \mathbf{H}\mathbf{x}_k, Z_k) d\mathbf{y}_k, \end{aligned} \quad (22)$$

does not have an analytical solution. In [188]–[190] this is handled by approximating the noise covariance as

$$R(\mathbf{y}) \approx R(\hat{\mathbf{y}}_k), \quad (23)$$

$$\hat{\mathbf{y}}_k = \mathbf{H}\hat{\mathbf{x}}_{k|k-1} = \mathbf{H}E[\mathbf{x}_k | \mathbf{Z}^{k-1}]. \quad (24)$$

This allows any of the updates presented in [3], [56], [138] to be used (see Tables IV, V and VI).

Non-linear range and azimuth measurement for the conditional state model (10c) and the measurement likelihood (17) are modelled in [113], where linearisation and a Variational Bayes scheme are used to handle the non-linearities in the update. Radar doppler rate is integrated into the measurement modelling in [171].

4) Dynamic modelling: In the original random matrix model [102] the transition density is modelled as

$$\begin{aligned} p(\mathbf{x}_{k+1}, X_{k+1} | \mathbf{x}_k, X_k) \\ \approx p(\mathbf{x}_{k+1} | X_{k+1}, \mathbf{x}_k) p(X_{k+1} | X_k), \end{aligned} \quad (25a)$$

$$\begin{aligned} = \mathcal{N}(\mathbf{x}_{k+1}; (F_k \otimes \mathbf{I}_d)\mathbf{x}_k, D_k \otimes X_{k+1}) \\ \times \mathcal{W}_d(X_{k+1}; n_k, X_k/n_k) \end{aligned} \quad (25b)$$

TABLE VIII  
Random matrix prediction from [102]

**Input:** Parameters  $m, P, v, V$  of conditional state density (10), motion model  $F$ , motion noise covariance  $D$ , sampling time  $T_s$ , temporal decay constant  $\tau$   
**Output:** Predicted parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= (F \otimes \mathbf{I}_d)m \\ P_+ &= FPF^T + D \\ v_+ &= e^{-T_s/\tau}v \\ V_+ &= \frac{v_+ - 2d - 2}{v - 2d - 2}V \end{aligned}$$

and in [56] a slightly different transition density was proposed,

$$\begin{aligned} p(\mathbf{x}_{k+1}, X_{k+1} | \mathbf{x}_k, X_k) \\ \approx p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(X_{k+1} | X_k). \end{aligned} \quad (26a)$$

$$\begin{aligned} = \mathcal{N}(\mathbf{x}_{k+1}; \mathbf{F}_k \mathbf{x}_k, Q_k) \\ \times \mathcal{W}_d(X_{k+1}; n_k, X_k/n_k) \end{aligned} \quad (26b)$$

In both cases we have a linear Gaussian transition density for the kinematic vector, and for the extent a Wishart transition density where the parameter  $n_k > 0$  governs the noise level of the prediction: the smaller  $n_k$  is, the higher the process noise.

The predicted parameters of the kinematic state are simple to compute. For the extent state, rather than solving the Chapman-Kolmogorov equation, a simple heuristic is used in which the expected value is kept constant and the variance is increased [102]. This corresponds to exponential forgetting for the extent state, see [83] for additional discussion. The predicted parameters are given in Table VIII and Table IX.

This model for the extent's time evolution is sufficient when the object manoeuvres are sufficiently slow. In practice, this means that the object turns slowly enough for the rotation of the extent to be very small from one time step to another. The kinematics transition density  $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$  in (26) is assumed independent of the extent. This neglects factors such as wind resistance, which can be modelled as a function of the extent  $X_k$ , however, the assumption is necessary to retain the functional form (15b) in a Bayesian recursion.

An alternative to the heuristic extent predictions from [56], [102] is to analytically solve the Chapman-Kolmogorov equation (9) for a Wishart transition density, and approximate the resulting density with an inverse Wishart density. Different approaches to this is discussed in, e.g., [83], [102], [107], [108], [117].

In [107], [108] the following transition density is used, where transformations of the extent are allowed

TABLE IX  
Random matrix prediction from [56]

**Input:** Parameters  $m, P, v, V$  of factorised state density (15), motion model  $\mathbf{F}$ , motion noise covariance  $Q$ , sampling time  $T_s$ , temporal decay constant  $\tau$

**Output:** Predicted parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= \mathbf{F}m \\ P_+ &= \mathbf{F}PF^T + Q \\ v_+ &= 2d + 2 + e^{-T_s/\tau}(v - 2d - 2) \\ V_+ &= \frac{v_+ - 2d - 2}{v - 2d - 2}V \end{aligned}$$

via known parameter matrices  $A_k$ ,

$$\begin{aligned} p(\mathbf{x}_{k+1}, X_{k+1} \mid \mathbf{x}_k, X_k) \\ \approx p(\mathbf{x}_{k+1} \mid X_{k+1}, \mathbf{x}_k) p(X_{k+1} \mid X_k), \end{aligned} \quad (27a)$$

$$\begin{aligned} = \mathcal{N}(\mathbf{x}_{k+1}; (F_k \otimes \mathbf{I}_d)\mathbf{x}_k, D_k \otimes X_{k+1}) \\ \times \mathcal{W}_d(X_{k+1}; n_k, A_k X_k A_k^T) \end{aligned} \quad (27b)$$

The solution to the Chapman-Kolmogorov equation (9) is not Gaussian inverse Wishart of the form (10), however, using moment matching it can be approximated as such. The predicted parameters are given in Table X. The parameter matrices  $A_k$  correspond to, e.g., rotation matrices. Rotation matrices are useful for a turning target, because the extent rotates as the target turns. By using the prediction in Table X with three motion models, with different matrices  $A_k$  corresponding to i) no rotation, ii) clockwise rotation and iii) counter-clockwise rotation, the target motion can be predicted better compared to using the prediction in Table VIII, leading to improved estimation, see [108].

The extent transition density  $p(X_{k+1} \mid X_k)$  in (25), (26), and (27), assumes independence of the prior kinematic state  $\mathbf{x}_k$ . The extent of an object going through a turning manoeuvre will typically rotate during the turn, because the extent is aligned with the object's heading. This implies that the extent transition density should be dependent on the turn-rate, i.e., it should be dependent on the kinematic state  $\mathbf{x}_k$ .

The inverse Wishart transition density is generalized in [77], [83] to allow for transformation matrices  $M(\mathbf{x}_k)$  that are functions of the kinematic state, which means that the rotation angle can be coupled to, e.g., the turn-rate, and estimated online. The following transition density is used with the factorised state density (15),

$$\begin{aligned} p(\mathbf{x}_{k+1}, X_{k+1} \mid \mathbf{x}_k, X_k) \\ \approx p(\mathbf{x}_{k+1} \mid \mathbf{x}_k) p(X_{k+1} \mid \mathbf{x}_k, X_k). \end{aligned} \quad (28a)$$

$$\begin{aligned} = \mathcal{N}(\mathbf{x}_{k+1}; \mathbf{f}_k(\mathbf{x}_k), Q_k) \\ \times \mathcal{W}_d\left(X_{k+1}; n_k, \frac{M(\mathbf{x}_k)X_k M(\mathbf{x}_k)^T}{n_k}\right) \end{aligned} \quad (28b)$$

TABLE X  
Random matrix prediction from [107]

**Input:** Parameters  $m, P, v, V$  of conditional state density (10), motion model  $F$ , motion noise covariance  $D$ , motion noise degrees of freedom  $n$ , parameter matrix  $A$

**Output:** Predicted parameters  $m_+, P_+, v_+, V_+$

$$\begin{aligned} m_+ &= (F \otimes \mathbf{I}_d)m \\ P_+ &= FPF^T + D \\ v_+ &= \frac{2n(\lambda - 1)(\lambda - 2)}{\lambda(\lambda + n)} + 2d + 4 \\ V_+ &= \frac{n}{\lambda - 1}(v - 2d - 2)AVA^T \\ \lambda &= v - 2d - 2 \end{aligned}$$

Note that a non-linear motion model  $\mathbf{f}(\cdot)$  is used.

Similarly to (27), the solution to the Chapman-Kolmogorov equation is not of the desired form, i.e., not a factorised Gaussian inverse Wishart (15). By minimising the Kullback-Leibler divergence, the predicted density can be approximated as Gaussian inverse Wishart of the form (15). The parameters of the prediction are given in Table XI. The proof that the solution  $s$  to the non-linear equation is unique is given in [77].

A comparison of the predictions resulting from the transition densities (26), (27) and (28), i.e., the predictions in Tables IX, X, XI, is presented in [83]. For a target that moves according to a constant turn motion model, see, e.g., [115, Sec. V.A], the prediction in Table XI is shown to give lowest filtering and prediction errors when the true turn-rate is unknown. If the true turn-rate is assumed to be known, the two predictions in Tables X and XI perform similarly. Average cycle times for Matlab implementations are reported for the prediction in Table XI and the prediction in Table X; the prediction in Table XI is shown to be about three times faster than the prediction in Table X with three modes. Note that any prediction or update can be speeded up, e.g., by parallelising computations or implementing in a fast low level language, like C++. Because of this it is important to interpret any differences in average cycle time with care.

When there are many measurements per object the measurement update will dominate the prediction and compensate for dynamic motion modelling errors. However, when multiple objects are located next to each other the prediction is important even in scenarios with many measurements per object, and accurate motion modelling can be crucial for estimation performance, see [83], [86], [87].

5) Further extensions of the random matrix model: Multiple extended object tracking is overviewed in Section IV, here we briefly mention some MTT algorithms where the random matrix model has been used.

In [200]–[202] it is used in the Probabilistic Multi-Hypothesis Tracking (PMHT) framework [176] to track persons in video data. The random matrix model has also been used in several RFS-type filters for multiple extended object tracking in clutter [19], [68], [78], [122]. JPDA-type MTT algorithms are presented in [170], [171], [187]. Multi object tracking requires the predicted likelihood

$$p(\mathbf{Z}) = \int \int p(\mathbf{Z} | \mathbf{x}, X) p(\mathbf{x}, X) d\mathbf{x} dX \quad (29)$$

In [78, Appendix A] it is shown that for the original model [102] the predicted likelihood is proportional to a generalized matrix variate beta type 2 distribution. In MTT algorithms it is necessary to maintain several object hypotheses due to the many involved uncertainties. When the random matrix model is used the number of hypotheses can be reduced using the merging algorithm presented in [81].

Elliptically shaped group objects are tracked under kinematical constraints in [101]. A multiple model framework is used to handle different object types in [31], [112], leading to joint tracking and classification. New object spawning, and merging of two object's into a single object, is modelled within the random matrix framework in [82]. The MTT algorithms mentioned above all consider a single sensor. In [191] the multi sensor case is considered, and four different updates are derived and compared using marine radar data. A random matrix estimator based on a Rao-Blackwellised state density, with a Gaussian for the kinematic state density and a particle approximation for the extent state density, is shown to have best performance, albeit at higher average cycle time than the other estimators [191]. The random matrix model is applied to mapping in [53], where a batch measurement update is presented, allowing all data to be processed at once instead of sequentially.

The random matrix model assumes an ellipse shape for the object's extent. For objects with irregular, non-ellipsoidal, extents, the shape can be approximated as a combination of several elliptically shaped subobjects. Using multiple instances of a simpler shape alleviates the limitations posed by the implied elliptic object shape,<sup>9</sup> and also retains, on a subobject level, the relative simplicity of the random matrix model. In [111] a single extended object model is given where the extended object is a combination of multiple subobjects with kinematic state vectors  $\mathbf{x}_k^{(i)}$  and extent matrices  $X_k^{(i)}$ , and each subobject is modelled using (10c). Note that this model assumes independence between the subobjects. By modelling the subobject kinematic vectors as dependent random variables estimation performance can be improved significantly, see [86], [87]. In [110] the non-ellipsoidal extended object model [111] is used

<sup>9</sup>As the number of ellipses grows, their combination can form nearly any given shape.

TABLE XI  
Random matrix prediction from [83]

**Input:** Parameters  $m, P, v, V$  of factorised state density (15), motion model  $\mathbf{f}(\cdot)$ , motion noise covariance  $Q$ , motion noise degrees of freedom  $n$ , matrix transformation function  $M(\cdot)$

**Output:** Predicted parameters  $m_+, P_+, v_+, V_+$

$$m_+ = \mathbf{f}(m)$$

$$P_+ = \mathbf{F} P \mathbf{F}^T + Q$$

$$v_+ = (d+1) \left( 2 + \frac{(s-d-1)(n-d-1)(v-2d-2)}{sn(v-d-1) - (s-d-1)(n-d-1)(v-2d-2)} \right)$$

$$V_+ = \frac{v_+ - d - 1}{v - d - 1} \frac{s - d - 1}{s} \frac{n - d - 1}{n} C_2$$

$$\mathbf{F} = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x})|_{\mathbf{x}=m}$$

$$C_1 = E[\log(\det(M(\mathbf{x})VM(\mathbf{x})^T))] ]$$

$$C_2 = E[M(\mathbf{x})VM(\mathbf{x})^T]$$

where  $s$  is the unique solution to  $h(s) = 0$ ,

$$h(s) = d \log \left( \frac{s}{2} \right) - \sum_{i=1}^d \psi_0 \left( \frac{s-i+1}{2} \right) + C_1 - \log(\det(C_2))$$

and  $\psi_k(\cdot)$  is the poly-gamma function of order  $k$ . A solution to  $h(s) = 0$  can be found using numerical root-finding. The second order Halley's iteration is

$$s^{(t+1)} = s^{(t)} - \frac{2h(s^{(t)})h'(s^{(t)})}{2(h'(s^{(t)}))^2 - h(s^{(t)})h''(s^{(t)})}$$

where the first and second order differentiations of  $h(s)$  w.r.t.  $s$  are

$$h'(s) = \frac{d}{s} - \frac{1}{2} \sum_{i=1}^d \psi_1 \left( \frac{s-i+1}{2} \right)$$

$$h''(s) = -\frac{d}{s^2} - \frac{1}{4} \sum_{i=1}^d \psi_2 \left( \frac{s-i+1}{2} \right)$$

The expected values can be approximated using Taylor expansion,

$$C_1 \approx \log(\det(M(m)VM(m)^T))$$

$$+ \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \frac{d^2 \log(\det(M(\mathbf{x})VM(\mathbf{x})^T))}{d\mathbf{x}_i d\mathbf{x}_j} \Big|_{\mathbf{x}=m} P_{i,j}$$

$$C_2 \approx M(m)VM(m)^T$$

$$+ \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \frac{d^2(M(\mathbf{x})VM(\mathbf{x})^T)}{d\mathbf{x}_i d\mathbf{x}_j} \Big|_{\mathbf{x}=m} P_{i,j}$$

where  $\mathbf{x}_i$  is the  $i$ th element of  $\mathbf{x}$ ,  $P_{i,j}$  is the  $i, j$ th element of  $P$ , and the differentiations are  $(M_{\mathbf{x}} = M(\mathbf{x})$  for brevity)

$$\frac{d^2 \log(\det(W))}{d\mathbf{x}_i d\mathbf{x}_j} = \text{Tr} \left( W^{-1} \frac{d^2 W}{d\mathbf{x}_i d\mathbf{x}_j} - W^{-1} \frac{dW}{d\mathbf{x}_i} W^{-1} \frac{dW}{d\mathbf{x}_j} \right)$$

$$\frac{dM_{\mathbf{x}}VM_{\mathbf{x}}^T}{d\mathbf{x}_j} = \frac{dM_{\mathbf{x}}}{d\mathbf{x}_j} VM_{\mathbf{x}}^T + M_{\mathbf{x}} V \frac{dM_{\mathbf{x}}^T}{d\mathbf{x}_j}$$

$$\frac{d^2 M_{\mathbf{x}}VM_{\mathbf{x}}^T}{d\mathbf{x}_i d\mathbf{x}_j} = \frac{d^2 M_{\mathbf{x}}}{d\mathbf{x}_i d\mathbf{x}_j} VM_{\mathbf{x}}^T + \frac{dM_{\mathbf{x}}}{d\mathbf{x}_i} V \frac{dM_{\mathbf{x}}^T}{d\mathbf{x}_j}$$

$$+ \frac{dM_{\mathbf{x}}}{d\mathbf{x}_i} V \frac{dM_{\mathbf{x}}^T}{d\mathbf{x}_j} + M_{\mathbf{x}} V \frac{d^2 M_{\mathbf{x}}^T}{d\mathbf{x}_i d\mathbf{x}_j}$$

in a joint tracking and classification framework. The work [225] derives a multi-Bernoulli filter for extended targets based on sub-random matrices.

For performance evaluation of estimates computed using any of the random matrix predictions/updates, the Gaussian Wasserstein distance is a suitable performance measure [211]. For the random matrix prediction/update presented in [56], see Tables IV and IX, the posterior Cramér-Rao Lower Bound CRLB is given in [163].

## B. Star-Convex Shape Approaches

Star-convex shape approaches based on the random hypersurface model [8], [10] and its variant the Gaussian process model [95], [198] constitute an extended object tracking framework that employs

- a parametric representation of the shape contour,
- a Gaussian distribution for representing the uncertainty of the joint state vector of the kinematic and shape parameters, and
- non-linear Kalman filters for performing the measurement update.

In contrast to the random matrix model that inherently relies on the elliptic shape, the approaches in this subsection are designed for general star-convex shapes (without using multiple subobjects). However, the increased flexibility comes at the price of more complex closed-form formulas.

In the following, we first discuss the benefits of non-linear Kalman filters for extended object tracking. Next, the random hypersurface model and the Gaussian process model for star-convex shapes are introduced. Finally, an overview of recent developments and trends in the context of random hypersurface models is given.

1) Review—Non-linear Kalman Filtering: Consider a general non-linear measurement function (time index is omitted) in the form

$$\mathbf{z} = h(\mathbf{x}, \mathbf{v}), \quad (30)$$

which maps the state  $\mathbf{x}$  and the noise  $\mathbf{v}$  to the measurement  $\mathbf{z}$ . We assume that both the prior probability density function of the state and noise density are Gaussian, i.e.,  $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; m, P)$  and  $p(\mathbf{v}) = \mathcal{N}(\mathbf{v}; 0, R)$ . In order to calculate the posterior density function

$$p(\mathbf{x} | \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{x}) \cdot p(\mathbf{x})}{p(\mathbf{z})}, \quad (31)$$

it is necessary to determine the likelihood function  $p(\mathbf{z} | \mathbf{x})$  based on (30). Unfortunately, as the noise in (30) is non-additive, no general closed-form solution for the likelihood is available. As a consequence, non-linear estimators that work with the likelihood function (e.g., standard particle filters) cannot be applied directly to this kind of measurement equation. However, there are non-linear filters that do not explicitly calculate the likelihood function—instead they exclusively work with the

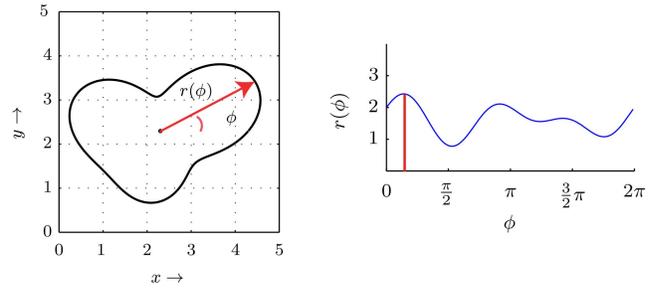


Fig. 9. Illustration of the representation of a star-convex contour (left) with a radius function  $r(\phi)$  (right).

measurement equation (30). The most prominent examples are *non-linear Kalman filters*, which directly apply the Kalman filter formulas to the non-linear measurement equation (30) in order to approximate the mean  $m^+$  and covariance  $P^+$  of the posterior density (31) as

$$m^+ = m + \text{Cov}[\mathbf{z}, \mathbf{x}]P^{-1}(\mathbf{z} - E[\mathbf{z}]) \quad (32)$$

$$P^+ = P - \text{Cov}[\mathbf{x}, \mathbf{z}]\text{Cov}[\mathbf{z}, \mathbf{z}]^{-1}\text{Cov}[\mathbf{z}, \mathbf{x}]. \quad (33)$$

Of course, in case of high non-linearity of the measurement equation, this can be a rough approximation. The exact posterior is only obtained in case of a linear measurement equation.

Analytic expressions for the required moments  $E[\mathbf{z}]$ ,  $\text{Cov}[\mathbf{z}, \mathbf{x}]$ , and  $\text{Cov}[\mathbf{z}, \mathbf{z}]$  in (32) and (33) are only available for special cases, e.g., polynomial measurement equations. However, a huge variety of approximate methods has been developed in the past such as the unscented transform [98]. An overview of recent methods is provided in [114].

2) Random Hypersurface Model: In the following, it is shown how the extended object tracking problem can be formulated as a measurement equation with non-additive noise (30) using the concept of a random hypersurface model. Based on the derived measurement equation, non-linear Kalman filters can be used to estimate the shape of extended objects as described above.

For this purpose, we first define a suitable parametrisation of a star-convex shape based on the so-called radius function  $r(\mathbf{p}_k, \phi)$ , which maps a shape parameter vector  $\mathbf{p}_k$  and an angle  $\phi$  to a contour point (relative to a centre  $\mathbf{d}_k$ ), see Figure 9 for an illustration. A reasonable (finite-dimensional) shape parameter vector  $\mathbf{p}_k$  can be defined by a Fourier series expansion [217] with  $N_F$  Fourier coefficients, i.e.,

$$r(\mathbf{p}_k, \phi) = R(\phi) \cdot \mathbf{p}_k,$$

where

$$R(\phi) = [\frac{1}{2}, \cos(\phi), \sin(\phi), \dots, \cos(N_F \phi), \sin(N_F \phi)],$$

$$\mathbf{p}_k = [a_k^{(0)}, a_k^{(1)}, b_k^{(1)}, \dots, a_k^{(N_F)}, b_k^{(N_F)}]^T.$$

Fourier coefficients with small indices capture coarse shape features while coefficients with larger indices represent finer details.

The overall state vector  $\mathbf{x}_k$  consists of the shape parameters  $\mathbf{p}_k$ , location  $\mathbf{d}_k$ , and further kinematic parameters  $\mathbf{c}_k$ , i.e.,

$$\mathbf{x}_k = [\mathbf{p}_k^T, \mathbf{d}_k^T, \mathbf{c}_k^T]^T. \quad (34)$$

A suitable measurement equation following the random hypersurface philosophy is formulated in polar form,

$$\mathbf{z}_k = s_k \cdot r(\mathbf{p}_k, \phi_k) + \mathbf{d}_k + \mathbf{v}_k, \quad (35)$$

where  $s_k \in [0, 1]$  is (multiplicative) noise that specifies the relative distance of the measurement source from the center, and  $\phi_k$  gives the angle to the measurement vector. In [9], it has been shown that  $s_k^2$  is uniformly distributed in case the measurement sources are uniformly distributed over the shape. It can be approximated by a Gaussian distribution with mean 0.8 and covariance  $\frac{1}{12}$ . By this means, the problem of estimating a (filled) shape has been reduced to a ‘‘curve fitting’’ problem, because for a fixed scaling factor  $s_k$ , (35) specifies a closed curve. See also the discussion in Section II-D.

The parameter  $\phi_k$  can be interpreted as a nuisance parameter (or latent variable) as in errors-in-variables models for regression and curve fitting. A huge variety of approaches for dealing with nuisance parameters has been developed in different areas. The most simple (and most inaccurate) approach is to replace the unknown  $\phi_k$  with a point estimate, e.g., the angle between  $\mathbf{d}_k - \mathbf{z}_k$  and the  $x$ -axis. This approach can be seen as greedy association model [50].

Having derived the measurement equation (35), a measurement update can be performed using the formulas (32) and (33). As (35) is polynomial for given  $\phi_k$ , closed-form formulas for the moments in the update equations are available.

As the greedy association model yields to a bias in case of high noise, a so-called *partial likelihood* has been developed, which outperforms the greedy association model in many cases [50], [52], e.g., high noise scenarios. For star-convex shapes, the partial likelihood model can be obtained from an algebraic reformulation of (35) and, hence, does not come with additional complexity [50], [52].

A further natural approach would be to assume  $\phi_k$  to be uniformly distributed on the interval  $[0, 2\pi]$ , however, a non-linear Kalman filter implicitly approximates a uniform distribution by a Gaussian distribution. Consequently, a reasonable mean for this Gaussian approximation is not obvious due to the circular nature of  $\phi_k$ .

Finally, we would like to note that due to the Gaussian state representation, prediction can be performed as usual in Kalman filtering, i.e., closed-form formulas are available for linear dynamic models and for non-linear dynamic models, non-linear Kalman filters can be employed.

### 3) Gaussian Process Model for Star-Convex Shapes:

Instead of using a Fourier series expansion for modelling the shape contour, [198] proposed to use Gaussian processes for star-convex shapes. A Gaussian Process [153] is a stochastic process, which is completely defined by a mean function  $\mu(u)$  and a kernel function  $k(u, u')$ :

$$f(u) \sim \mathcal{GP}(\mu(u), k(u, u')). \quad (36)$$

For a finite number of inputs  $u_1, \dots, u_n$ , a Gaussian process follows

$$[f(u_1) \cdots f(u_n)]^T \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}), \quad (37)$$

where

$$\boldsymbol{\mu} = [\mu(u_1) \cdots \mu(u_n)]^T \quad (38)$$

$$\mathbf{K} = \begin{bmatrix} k(u_1, u_1) & \cdots & k(u_1, u_n) \\ \vdots & \ddots & \vdots \\ k(u_n, u_1) & \cdots & k(u_n, u_n) \end{bmatrix}. \quad (39)$$

Gaussian processes are often used in machine learning. In contrast to machine learning approaches, where batch processing is typically applied, tracking applications require a recursive estimate of the Gaussian process for shape representation. Thus, the function  $f(u)$  is approximated by a finite number of function values or basis points which are updated over time. Consequently, the Gaussian process is described using a constant number of parameters which resembles the parameterization used in the random hypersurface model. However, the basis points are uniformly distributed over the angle interval, i.e., a separation of the basis points into points for coarse and fine shape features (cf. parameters for coarse and fine in (34)) is not possible.

The kernel function  $k$  restricts the kind of functions which can be represented by the Gaussian process, e.g., to symmetric functions [95], [198]. Besides the Kalman filter-based implementations, a Rao-Blackwellised particle filter implementation of the Gaussian process model for star-convex objects has been proposed in [142].

### 4) Further developments, extensions, and variations:

In the same manner as for star-convex shapes [9], the concept of a random hypersurface model can be applied for circular and elliptic shapes [12].

In many applications, the object to be tracked is symmetric, e.g., an aircraft or a vehicle. In this case specific improvements and adoptions can be performed in order to incorporate symmetry information [51], [95]. The concept of scaling the boundary of a curve in order to model an extended object has been combined with level sets in [213] in order to model arbitrary connected shapes. A closed-form likelihood for the use in non-linear filters based on the RHM measurement equation (35) has been derived in [174]. Elongated objects are considered in [214]. The RHM idea can be used in the same manner to model three-dimensional

shapes in three-dimensional space. In addition, two-dimensional shapes in three-dimensional space can also be modelled with RHM ideas [51], [52]. For example, in [52], measurements from a cylinder are modelled by means of translating a plane curve, i.e., a circle.

It is interesting to note that clutter detections that are not from the extended object, can improve shape estimation [215], [216] by modelling them as negative information. Furthermore, camera calibration can be performed by means of tracking an extended object [49].

5) **Multiplicative Error Model:** The basic idea of the RHM is to model one dimension of the spatial extent with a random scaling factor and the other one with, e.g., a greedy association model (GAM). By this means, Bayesian inference becomes tractable with a non-linear Kalman filter.

A recent line of work models both dimensions with a scaling factor [15], [209], [210], i.e., multiplicative noise. In this way, a uniform distribution can be matched better for simple shapes, such as circles or ellipses. The resulting model is called Multiplicative Error Model (MEM).

For tracking an elliptical shape, the state vector can be defined as

$$\mathbf{x}_k = [\mathbf{c}_k^T, \mathbf{p}_k^T]^T \quad (40)$$

where  $\mathbf{c}_k^T$  is the kinematic vector (here including the center) and

$$\mathbf{p}_k = [\alpha_k, l_{k,1}, l_{k,2}]^T$$

is the shape parameter vector with ellipse orientation  $\alpha_k$ , and semi-axes lengths  $l_{k,1}$  and  $l_{k,2}$ . Then the  $i$ th measurement at time  $k$  is modelled as

$$\mathbf{z}_k^i = \mathbf{H}\mathbf{c}_k + \mathbf{Rot}(\alpha_k) \begin{bmatrix} l_{k,1} & 0 \\ 0 & l_{k,2} \end{bmatrix} \begin{bmatrix} h_{k,1}^i \\ h_{k,2}^i \end{bmatrix} + \mathbf{v}_k^i \quad (41)$$

where  $\mathbf{H}_k$  is a matrix that picks out the object position from the kinematic state,

$$\mathbf{Rot}(\alpha_k) = \begin{bmatrix} \cos \alpha_k & -\sin \alpha_k \\ \sin \alpha_k & \cos \alpha_k \end{bmatrix} \quad (42)$$

is a rotation matrix,  $\mathbf{v}_k^i$  is additive sensor noise, and both  $h_{k,1}^i$  and  $h_{k,2}^i$  are (Gaussian) multiplicative noise terms that we assume to be mutually independent of all other random variables. Following the reasoning for the parameter  $z$  in (13), the variances of the multiplicative noise are set to  $\sigma_{h_1} = \sigma_{h_2} = \frac{1}{4}$  in order to match an elliptic uniform spatial distribution. In this manner, the multiplicative noise models the spatial distribution, i.e., the uncertainty of the measurement source. The corresponding likelihood to (41) coincides with the likelihood used in the random matrix approach, i.e., (14), but the ellipse parametrisation is different.

Unfortunately, it turns out that a direct application of the Kalman filter formulas to (41) does not give satisfying results [15] due to the strong linearities. A solution is to augment the original measurement equation (41)

TABLE XII

Update of the EKF for the multiplicative error model [210]. Source code: <http://github.com/Fusion-Goettingen>.

**Input:** Kinematic state prior mean  $m^c$  and covariance  $P^c$ , shape variable prior mean  $m^p$  and covariance  $P^p$  as defined in (40), measurement matrix  $\mathbf{H}$ , measurement noise covariance  $R$ , multiplicative noise variance  $\sigma_{h_1}$  and  $\sigma_{h_2}$ , measurement  $\mathbf{z}$

**Output:** Updated parameters  $m_+^c$ ,  $P_+^c$ ,  $m_+^p$  and  $P_+^p$ ,

$$m_+^c = m^c + \text{Cov}[\mathbf{x}, \mathbf{z}] (\text{Cov}[\mathbf{z}, \mathbf{z}])^{-1} (\mathbf{z} - \mathbb{E}[\mathbf{z}])$$

$$P_+^c = P^c - \text{Cov}[\mathbf{x}, \mathbf{z}] (\text{Cov}[\mathbf{z}, \mathbf{z}])^{-1} (\text{Cov}[\mathbf{x}, \mathbf{z}])^T$$

$$m_+^p = m^p + \text{Cov}[\mathbf{x}, \tilde{\mathbf{z}}] (\text{Cov}[\tilde{\mathbf{z}}, \tilde{\mathbf{z}}])^{-1} (\mathbf{z} - \mathbb{E}[\tilde{\mathbf{z}}])$$

$$P_+^p = P^p - \text{Cov}[\mathbf{x}, \tilde{\mathbf{z}}] (\text{Cov}[\tilde{\mathbf{z}}, \tilde{\mathbf{z}}])^{-1} (\text{Cov}[\mathbf{x}, \tilde{\mathbf{z}}])^T$$

$$\mathbb{E}[\mathbf{z}] = \mathbf{H}m^c$$

$$\text{Cov}[\mathbf{c}, \mathbf{z}] = P^c \mathbf{H}^T$$

$$\text{Cov}[\mathbf{z}, \mathbf{z}] = \mathbf{H}P^c \mathbf{H}^T + S \text{diag}(\sigma_{h_1}, \sigma_{h_2}) S^T + R$$

$$S = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \text{diag}(l_1, l_2)$$

$$[\alpha, l_1, l_2]^T = m^p$$

$$\tilde{\mathbf{z}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} ((\mathbf{z} - \mathbb{E}[\mathbf{z}]) \otimes (\mathbf{z} - \mathbb{E}[\mathbf{z}]))$$

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \text{Cov}[\mathbf{z}, \mathbf{z}]$$

$$\mathbb{E}[\tilde{\mathbf{z}}] = [\sigma_{11} \quad \sigma_{22} \quad \sigma_{12}]^T$$

$$\text{Cov}[\tilde{\mathbf{z}}, \tilde{\mathbf{z}}] = \begin{bmatrix} 3\sigma_{11}^2 & \sigma_{11}\sigma_{22} + 2\sigma_{12}^2 & 3\sigma_{11}\sigma_{12} \\ \sigma_{11}\sigma_{22} + 2\sigma_{12}^2 & 3\sigma_{22}^2 & 3\sigma_{22}\sigma_{12} \\ 3\sigma_{11}\sigma_{12} & 3\sigma_{22}\sigma_{12} & \sigma_{11}\sigma_{22} + 2\sigma_{12}^2 \end{bmatrix}$$

$$\text{Cov}[\mathbf{p}, \tilde{\mathbf{z}}] = P^p M^T$$

$$M = \begin{bmatrix} -\sin 2\alpha & \cos^2 \alpha & \sin^2 \alpha \\ \sin 2\alpha & \sin^2 \alpha & \cos^2 \alpha \\ \cos 2\alpha & \sin 2\alpha & -\sin 2\alpha \end{bmatrix} \cdot \begin{bmatrix} (l_1)^2 \sigma_{h_1} & -(l_2)^2 \sigma_{h_2} & 0 & 0 \\ 0 & 2l_1 \sigma_{h_1} & 0 & 0 \\ 0 & 0 & 2l_2 \sigma_{h_2} & 0 \end{bmatrix}$$

with the squared measurement  $\mathbf{z}^2$  using the Kronecker product and then apply a non-linear Kalman filter. In this way, higher order moments are incorporated in the update formulas. For this purpose, an Extended Kalman filter is derived in [210] that results in compact update formulas for the extent, which are depicted in Table XII. Exact prediction can be performed for linear models, see Table XIII.

#### IV. TRACKING MULTIPLE EXTENDED OBJECTS

In this section we overview multiple extended object tracking. Regardless of the type tracking problem—

TABLE XIII

Prediction of the EKF for the multiplicative error model prediction [210]. Source code: <http://github.com/Fusion-Goettingen>.

**Input:** Kinematic state mean  $m^c$  and covariance  $P^c$ , shape mean  $m^p$  and covariance  $P^p$ , process matrices  $F^c$ ,  $F^p$  with process noise covariances  $Q^c$  and  $Q^p$

**Output:** Parameters  $m_*^c, P_*^c, m_*^p$  and  $P_*^p$  for the prediction

$$m_*^c = F^c m^c$$

$$P_*^c = F^c P^c (F^c)^T + Q^c$$

$$m_*^p = F^p m^p$$

$$P_*^p = F^p P^p (F^p)^T + Q^p$$

point, extended, group, etc.—MTT is a problem that has many challenges:

- The number of objects is unknown and time varying.
- There are missed measurements, i.e., at each time step, some of the existing objects do not give measurements.
- The objects that are not missed give rise to an unknown number of detections.
- There are clutter measurements, i.e., measurements that were not caused by a target object.
- Measurement origin is unknown, i.e., the source of each measurement is unknown. This is often referred to as the “data association problem.”

For multiple point object tracking the literature is vast; recently a comprehensive overview of MTT algorithms, with a focus on point objects, was written by Vo et al. [197]. Since many of the existing extended object MTT algorithms are of the RFS type, we focus on these algorithms in the following (see IV-B.4 for selected approaches with other MTT algorithms). In the following subsections we will first give a brief overview of RFS filters, then we give examples of extended and group object MTT algorithms, and lastly we discuss the data association problem in extended object MTT.

#### A. Review—RFS filters

A random finite set (RFS) is a set whose cardinality is a random variable, and whose set members are random variables. In RFS based tracking algorithms both the set of objects and the sets of measurements are modelled as RFSS. Tutorials on RFS methods can be found in, e.g., [71], [126], [193], and in-depth descriptions of the RFS concept and of finite set statistics (FISST) are given in the books [125], [127].

The state of the set of objects that are present in the surveillance space is referred to as the *multi-object state*. Because of the computational complexity, specifically due to the data association problem, a full multi-object Bayes filter can be quite computationally demanding to run, and approximations of the data association problem are necessary. Computationally tractable filters include

the Probability Hypothesis Density (PHD) filter [128], the Cardinalized PHD (CPHD) filter [129], the Cardinality Balanced MeMber (CB-MeMber) filter [195], and the MTT conjugate priors [194], [205].

1) PHD and CPHD filters: The first order moment of the multi-object state is called the PHD,<sup>10</sup> and can be said to be to a random set as the expected value is to a random variable. A PHD filter recursively estimates the PHD under an assumed Poisson distribution for the cardinality. A consequence of the Poisson assumption is that the PHD filter’s cardinality estimate has high variance, a problem that manifests itself, e.g., where there are missed measurements [46]. The CPHD filter recursively estimates the PHD and a truncated cardinality distribution, and is known to have a better cardinality estimate compared to the PHD filter. The PHD and CPHD filters were first derived in [128], [129] using probability generating functionals.<sup>11</sup> In [63] it is shown that the PHD and CPHD filters can be derived by minimizing the Kullback-Leibler divergence [104] between the multi-object density and either a PPP density (PHD filter) or an iid cluster process density (CPHD filter).

In both the PHD filter and the CPHD filter the objects are independent identically distributed (iid); the normalised PHD is the estimated object pdf. When there are multiple objects the PHD has multiple modes (peaks), where each mode corresponds to one object. An exception to this is when two or more objects are located close to each other; in this case a mode can correspond to multiple objects, also called *unresolved* objects. The estimated number of objects located in an area, e.g., under one of the modes, is given by integrating the PHD over that area. Both the PHD filter and the CPHD filter are susceptible to a “spooky effect” [62], [127], a phenomenon manifested by PHD mass shifted from undetected objects to detected objects, even in cases when the objects are far enough away that they ought to be statistically insulated.

Ultimately the desired output from an MTT algorithm is a set of estimated trajectories (tracks), where a trajectory is defined as the sequence of states from the time the object appears to the time it disappears. In their most basic forms neither the PHD nor the CPHD formally estimate object trajectories. However, object trajectories can be obtained, e.g., using post-processing with labelling schemes [75], [76], [144].

2) CB-MeMber filter: The CB-MeMber filter [195] approximates the multi-object density with a multi-Bernoulli (MB) density [125, Ch. 17]. In an MB density the objects are independent but not identically distributed, compared to the PHD and CPHD filters where

<sup>10</sup>The first order moment is also called intensity function, see, e.g., [126], [192].

<sup>11</sup>The probability generating functional is an integral transform that can be used when working with RFS densities, see further in, e.g., [125], [127].

the objects are iid. The Bernoulli RFS density is a suitable representation of a single object, as it captures both the uncertainty regarding the object's state, as well as the uncertainty regarding the object's existence. As the name suggests, an MB density is the union of several independent Bernoulli densities, and it is therefore a suitable representation of multiple objects. The CB-MeMBer filter fixes the biased cardinality estimate of the MeMBer filter presented in [125, Ch. 17].

3) MTT conjugate priors: The concepts *conjugacy* and *conjugate prior* are central in Bayesian probability theory. In an MTT context, conjugacy means that if we begin with a multi-object density of a conjugate prior form, then all subsequent predicted and updated multi-object densities will also be of the conjugate prior form. Two MTT conjugate priors can be found in the literature, both based on multi-Bernoulli representations for the set of objects.

The first is based on labeled RFSS and is called Generalized Labeled Multi-Bernoulli (GLMB) [194]. In the GLMB filter the labels are used to obtain target trajectories. Because of the unknown measurement origin, the GLMB has a mixture representation, where each component in the mixture corresponds to one possible data association history. The GLMB filter performs well in challenging scenarios, however, it is computationally expensive. A computationally efficient approximation is the Labeled Multi-Bernoulli (LMB) filter [156], which approximates the GLMB mixture with a single labeled multi-Bernoulli density. Both the GLMB and LMB filters rely on handling the data association problem by computing the  $M$  top ranked assignments, an analysis of the approximation error incurred by this is presented in [196].

The second MTT conjugate prior is based on regular RFSS, i.e., unlabeled, and is called Poisson Multi-Bernoulli Mixture (PMBM) [205]. The PMBM conjugate prior allows an elegant separation of the set of objects into two disjoint subsets: objects that have been detected, and objects that have not yet been detected. A Poisson point process density is used for the undetected objects, and a multi-Bernoulli mixture is used for the detected objects. Explicitly modelling the objects that have not been detected is useful, e.g., when the sensor is susceptible to occlusions, or when the sensor is mounted to a moving platform. Similarly to the GLMB filter, in the PMBM filter the components in the multi-Bernoulli mixture corresponds to different data association histories. A variational Bayesian approach to approximating the multi-Bernoulli mixture density with a single multi-Bernoulli density is presented in [204], leading to the Variational Multi-Bernoulli (VMB) filter. Note that the variational approximation does not affect the Poisson part that models the undetected objects. The VMB filter can be understood to be to the PMBM filter, as the LMB filter is to the GLMB filter. However, it should be noted

that the approximations used in the VMB and LMB are not the same.

## B. Examples of extended and group MTT

1) PHD and CPHD filters: A PHD filter for extended objects under the Poisson model [66], see also Section II-C.2, was presented in [130]. Gaussian mixture implementations of this extended object PHD filter, for both linear and non-linear motion and measurement models, are presented in [72]–[74]. The resulting filters can be abbreviated ET-GM-PHD filters. A Gaussian inverse Wishart implementation, using the random matrix extended object model [102] (see also Section III-A), is presented in [78], [80], and the resulting filter is abbreviated GIW-PHD filter. A Gaussian mixture implementation using RHM (see Section III-B) was presented in [219]. Multiple model Gaussian mixture PHD filters can be found in [70], [85]; the filters are applied to tracking of cars and bicycles, under assumed rectangle and stick shape models, and it is shown that using multiple measurement models can improve the estimation results. Augmenting the implementations with gamma distributions makes it possible to estimate the unknown Poisson measurement rate for each object [79]. The resulting algorithms are then called gamma Gaussian (GG), or gamma Gaussian inverse Wishart (GGIW), respectively.

An approach to group object tracking based on a point object GM-PHD filter is presented in [35]. The extended object PHD filter presented in [182], [183] is derived for an object model different from the Poisson point process model [66]. The objects are modelled by a Poisson cluster process, a hierarchic process with a parent process and a daughter process. The parent process models a Poisson distributed number of objects. For each object a daughter process models a number of reflection points that generate measurements. An implementation is proposed where the object is assumed ellipse shaped and the reflection points are located on the edge of the ellipse.

At least two different CPHD filters have been presented. The CPHD filter for extended objects presented in [116] is derived under the assumption that “*relative to sensor resolution, the extended objects and the unresolved objects are not too close and the clutter density is not too large*” [116, Corollary 1]. However, this is an assumption that cannot be expected to hold in the general case. A CPHD filter capable of handling both spatially close objects and dense clutter is presented in [122], [139]–[141], and a GGIW implementation is also presented. A comparison shows that the GGIW-CPHD filter outperforms the GGIW-PHD filter, especially when the probability of detection is low, and/or the clutter density is high. The price for the increased performance is that the computational cost increases.

2) CB-MeMBer filters: An extension of the CB-MeMBer filter [195] to extended objects, using the PPP mea-

surement model overviewed in Section II-C.2, was presented in [218]. A Gaussian mixture implementation is presented in [218], and Sequential Monte Carlo (SMC) implementations of the CB-MEMBER for extended objects can be found in [118], [124]. An extended object CB-MEMBER filter with multiple models is presented in [97].

3) Conjugate priors: Labeled MB filters for extended object tracking are presented in [18], [19], both a GLMB filter and its approximation the LMB filter. GGIW implementations are presented, and simulation results show that the labelled MB filters outperform their PHD and CPHD counterparts. Additionally, the GLMB and LMB filters estimate object trajectories, which the PHD and CPHD filters do only if labeling is used in post processing, see, e.g., [75], [76]. The LMB filter was applied to LIDAR data for rectangular objects using the separable likelihood approach [167] and for star-convex objects using a modelling with Gaussian processes [95].

A PMBM filter for extended and group objects is derived and presented with a GGIW implementation in [68], [69]. A simulation study showed that the extended object PMBM filter outperforms the PHD, CPHD and LMB filters, and an experiment with LIDAR data illustrates that the PPP model can accurately represent the occluded areas of the surveillance space. The GGIW-PMBM model is applied to mapping in [53], where a batch measurement update is derived.

4) Non-RFS approaches: A Gaussian Mixture Markov Chain Monte Carlo filter for multiple extended object tracking is presented in [33]. The filter is compared to the linear ET-GM-PHD-filter [72], [74], and is shown to be less sensitive to clutter but also considerably more computationally costly (as measured by the average cycle time). The Probabilistic Multi-Hypothesis Tracker (PMHT) [176] allows more than one measurement per object, and the random matrix model (Section III-A) has been integrated in the PMHT framework, see [200]–[202]. A variational Bayesian Expectation Maximisation approach to mapping with extended objects is presented in [120].

### C. Multiple extended object data association

In MTT a data association specifies for each measurement the source from whence it came: either it is an object measurement or a clutter measurement. The possibility of multiple measurements per object means that in extended object MTT a data association can be split into two parts:

1) **Partition:** A partition of a set, denoted  $\mathcal{P}$ , is defined as a division of the elements of the set into non-empty subsets, called cells [130] and denoted  $\mathbf{W}$ , such that each element belongs to one and only one cell. The cells are to be understood to contain measurements that are from the same source, i.e., all measurements in the cell are from the same extended object, or they are all clutter.

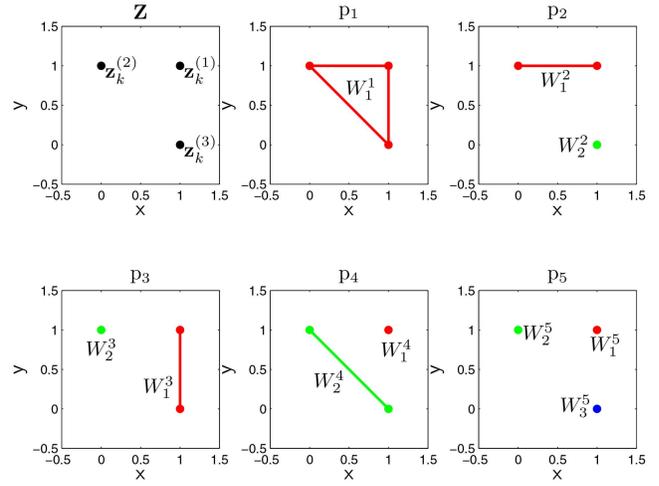


Fig. 10. Partition illustration. There are three measurements  $\mathbf{z}_k^{(1)}$ ,  $\mathbf{z}_k^{(2)}$ , and  $\mathbf{z}_k^{(3)}$ , which can be partitioned in five different ways. In the  $j$ th partition, denoted  $p_j$ , the  $i$ th cell is denoted  $W_i^j$ . With three measurements there is one partition with one cell, three partitions with two cells, and one partition with three cells. Note that the ordering of the partitions and cells is arbitrary; the particular ordering in this example is only used for notational simplicity.

2) **Cell association:** An association of the cells to a measurement source, either one of the objects or a clutter source.

Note that an association from measurement to cell, and from cell to source, defines an association from measurement to source.

For Bayes optimality it is necessary to consider all possible data associations in the MTT update. This means that in extended and group MTT it is necessary to consider all possible partitions of the set of measurements, and for each partition one has to consider all possible cell associations. Unless the measurement set contains a trivial number of measurements (i.e., extremely few) and there is a trivial number of objects, both of these problems are intractable because there are too many possible partitions, and too many possible cell associations. Fortunately, in the literature we can find methods that allow us to handle both of these problems. Below we first discuss the complexity of the partitions and the cell associations, and then we overview the solutions to these problems that can be found in the literature.

1) Complexity analysis: Let the set of measurements contain  $n$  measurements in total. The number of possible ways to partition a set of  $n$  measurements is given by the  $n$ th Bell number, denoted  $B(n)$  [161]. The sequence of Bell numbers is log-convex,<sup>12</sup> and  $B(n)$  grows very rapidly as  $n$  grows. For  $n = 3$  measurements there are  $B(3) = 5$  possible partitions; an example is shown

<sup>12</sup>The sequence of Bell numbers is logarithmically convex, i.e.,  $B(n)^2 \leq B(n-1)B(n+1)$  for  $n \geq 1$  [45]. If the Bell numbers are divided by the factorials,  $B(n)/n!$ , the sequence is logarithmically concave,  $(B(n)/n!)^2 \geq (B(n-1)/(n-1)!)(B(n+1)/(n+1)!)$ , for  $n \geq 1$  [30].

in Figure 10. For twice the number of measurements ( $n = 6$ ) there are  $B(6) = 203$  possible partitions, and for  $n = 90$  measurements there are  $B(90) > 10^{100}$  possible partitions. In other words, it is computationally intractable to consider all partitions, and approximations are necessary for implementation.

Let  $|\mathcal{P}|$  be the number of cells in the partition  $\mathcal{P}$ , and let  $m$  be the prior number of object estimates. Each cell can either be from one of the existing prior object estimates, or it could be from a new object. Thus, there are  $|\mathcal{P}| + m$  possible sources. The number of possible ways to associate  $|\mathcal{P}|$  cells to  $|\mathcal{P}| + m$  sources is

$$\binom{m + |\mathcal{P}|}{|\mathcal{P}|} = \frac{(m + |\mathcal{P}|)!}{m!|\mathcal{P}|!} \quad (43)$$

Similarly to the partitions, unless the number of cells and number of objects are very small, it is infeasible to consider all possible associations.

2) Complexity reduction: The MTT literature contains several different methods that can be used to alleviate the complexity, and that allows extended object MTT filters to be implemented using limited computational resources.

Gating, see, e.g., [5, Sec. 2.2.2.2], is a method that removes possible measurement-to-object associations by comparing the measurements to predictions of the objects' measurements. If the difference between the measurement and the predicted measurement is too large, the association is ruled out as infeasible. Gating has been used in a plethora of MTT algorithms, both for point targets and extended targets. Naturally, for extended targets the gates must take into account the position of the target, the size and shape of the target, as well as state uncertainties. Using gating it is possible to group the measurements and the objects into smaller groups that, given the gating decision, are independent. This way one can solve several smaller data association problems instead of one larger data association problem.

Even after gating, there are typically too many possible partitions and cell associations. An important contribution of [72], [74], [78] is to show how clustering can be used to find a subset of partitions. The basic insight behind the use of clustering lies in the definition of extended objects: the measurements are spatially distributed around the object. Therefore spatially close measurements are more likely to be from the same object, than spatially distant measurements. By only considering the partitions in which the cells contain spatially close measurements many partitions can be pruned, and the update becomes tractable.

Distance Partitioning [72], [74] is a simple method that puts measurements in the same cell if the distance between a measurement and its closest neighbour is less than a threshold. A detailed description of Distance Partitioning is given in [72], [74], [84]. By considering multiple thresholds, a subset of partitions is obtained.

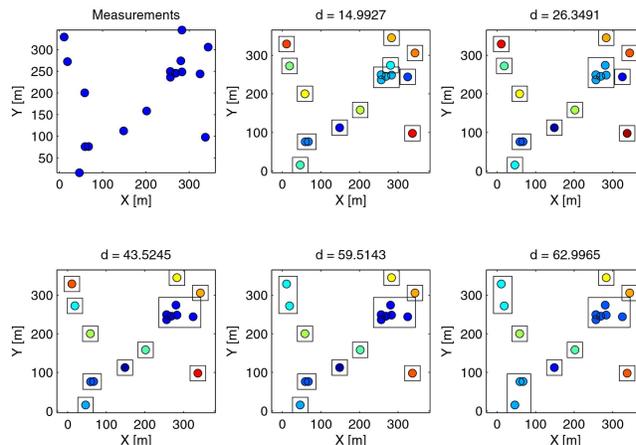


Fig. 11. Illustration of the output from Distance Partitioning, with 17 measurements. By clustering the measurements with progressively larger thresholds  $d$  different partitions are obtained. The smallest and largest threshold that are used are parameters of the clustering algorithm.

Finding a good subset of partitions is especially important when multiple extended objects are located in close vicinity of each other, see [19], [78], [122].

An example where Distance Partitioning is used is given in Figure 11. In this example there are 17 measurements, for which there are more than  $10^{10}$  possible partitions. Using Distance Partitioning this is limited to five partitions. Results from both simulations and experiments have shown that, despite the very drastic reduction in the number of partitions that are considered, performance is not sacrificed when clustering is used, see, e.g., [70], [76], [85], [165]. However, there may be scenarios where two objects are so close to each other, that their measurements may not be separated any more based on the distance. In these scenarios, prior information about the number of objects (e.g., based on the current cardinality estimate) may be used to improve partitioning (cf. [78]).

Distance Partitioning is an example of a hierarchical single linkage clustering algorithm, see, e.g., [22] for a discussion about clustering. Other clustering methods have also been used in an extended object MTT context, e.g., Gaussian Mixture Expectation Maximisation [78], spectral clustering [208], and fuzzy adaptive resonance theory [220], [221].

The extended object PHD and CPHD filters avoid the cell association through approximation, and instead the PHD is updated using all measurements. When the PHD has a distribution mixture representation, e.g., a Gaussian mixture, then the updated PHD is obtained by updating each Gaussian component in the PHD mixture with each measurement. In other extended object MTT filters, the number of cell associations can be reduced, either by computing association probabilities or by finding the best associations. Using association probabilities means that for each measurement-object-pair we compute the probability that the object is the origin of the measurement, and the probabilities are then used in the MTT

update. JPDA association probabilities are used in [171], [187]. Alternatively, one can find the best association assignment(s) by optimising a cost function that is related to the MTT predicted likelihood. The single best assignment can be found using the auction algorithm [20], and the  $M$  top ranked assignments can be found using Murty’s algorithm [135]. Finding optimal assignments is used in the implementations of the extended object conjugate priors [18], [19], [68], [69]. In [175], a JPDAF intensity filter that estimates an intensity function for each extended object is developed.

## V. METRICS AND PERFORMANCE EVALUATION

Regardless of the target type—point, extended, group or multi-path—it is important to be able to evaluate the performance of a target tracking algorithm, such that the estimates can be compared to the ground truth and different tracking algorithms can be compared to each other. For point targets the root mean squared error (RMSE) is a standard metric. For Gaussian assumed state estimates, the normalised estimation error squared (NEES) is another standard performance measure, that incorporates also the estimated covariance matrix and evaluates whether or not the estimate is consistent.

In extended object tracking the tracker output incorporates extent information, and because of this it is not trivial to answer the question: what is the distance between the estimate and the ground truth. It may seem tempting to use the RMSE, however, doing so is not always straightforward as the following two examples illustrate.

- 1) Consider an extended object with an assumed rectangular shape and state vector

$$\mathbf{x} = [x, y, l_1, l_2, \varphi]^T \quad (44a)$$

where  $x, y$  is the position,  $l_1$  and  $l_2$  are the dimensions of the two sides, and  $\varphi$  is the orientation of the side with length  $l_1$  (and does not specify the moving direction of the object). For this state vector the two estimates

$$\hat{\mathbf{x}}^{(1)} = [x, y, l_1, l_2, \varphi]^T, \quad (44b)$$

$$\hat{\mathbf{x}}^{(2)} = [x, y, l_2, l_1, \varphi + 0.5\pi]^T, \quad (44c)$$

where width and length are switched in  $\hat{\mathbf{x}}^{(2)}$ , define exactly the same shape in the Cartesian surveillance space, however, the RMSE errors would not be the same for the two estimates, which clearly violates intuition.

- 2) In the random matrix model the extended object state is a combination of a vector and a matrix. The estimated vector can be compared to the ground truth using the Euclidean norm. The matrix generalisation of the Euclidean norm for vectors is the Frobenius norm, and this norm can be used to compare the estimated matrix to the ground truth. In [122] it is suggested to use a weighted summation to combine

the vector norm and the matrix norm, however, this leads to a problem whereby one has to determine the weights in the summation.

In some works, see, e.g., [56], the extended object state is broken down into specific properties, such as position, velocity, orientation, extent area, and extent dimensions.<sup>13</sup> This facilitates easy interpretation of the results, however, by this means it is no longer possible to rank estimates from different trackers using a single score. Furthermore, standard multi-object metrics, such as the optimal sub-pattern assignment (OSPA) metric [169] and the generalized OSPA (G-OSPA) [152] build upon single object metrics that give a single output. In other words, breaking down the extended object state into different properties does not facilitate multi-object performance evaluation.

A widely-used measure in computer vision is the so-called Intersection-over-Union (IoU), which is defined as the area of the intersection between the estimated shape and the ground truth shape, divided by the area of the union of the two shapes. In the extended object tracking context, IoU has been used, e.g., for rectangular and elliptical extended objects [73]. For axis-aligned rectangles the IoU is simple to compute, however, for other shapes, or rectangles that are not axis-aligned, computing the IoU can be cumbersome. Furthermore, the IoU is always zero for non-overlapping objects,<sup>14</sup> meaning that the error measure is the same regardless of how big the translational error is. This goes against intuition, which tells us that the larger the translational difference is between two shapes, the larger the error should be.

One work in this direction is [211], which addresses performance metrics for elliptically shaped extended objects. By comparing several metrics and measures, the so-called Gaussian Wasserstein distance is identified as the most appropriate one. The Gaussian Wasserstein distance is available in closed-form, gives intuitive results, and is a true metric. Unfortunately, for general shapes, no analytic formulas for the Wasserstein distance exist, meaning that the Wasserstein metric is currently only suitable for objects with elliptic extents.

For star-convex shapes, the work [178] discusses a modified Hausdorff distance that fully incorporates different shape parametrisations.

While the existing extended object performance measures for non-elliptic shapes, such as decomposition into specific properties and IoU, have their applications, there is still a lot of work needed to specify a general single extended object performance evaluation criterion. However, for multiple extended object performance evaluation, given a chosen single object metric,

<sup>13</sup>For example, the semi-axes of an ellipse or the two sides of a rectangle.

<sup>14</sup>For two non-overlapping shapes, the intersection is empty, and thus the area of the intersection is zero.

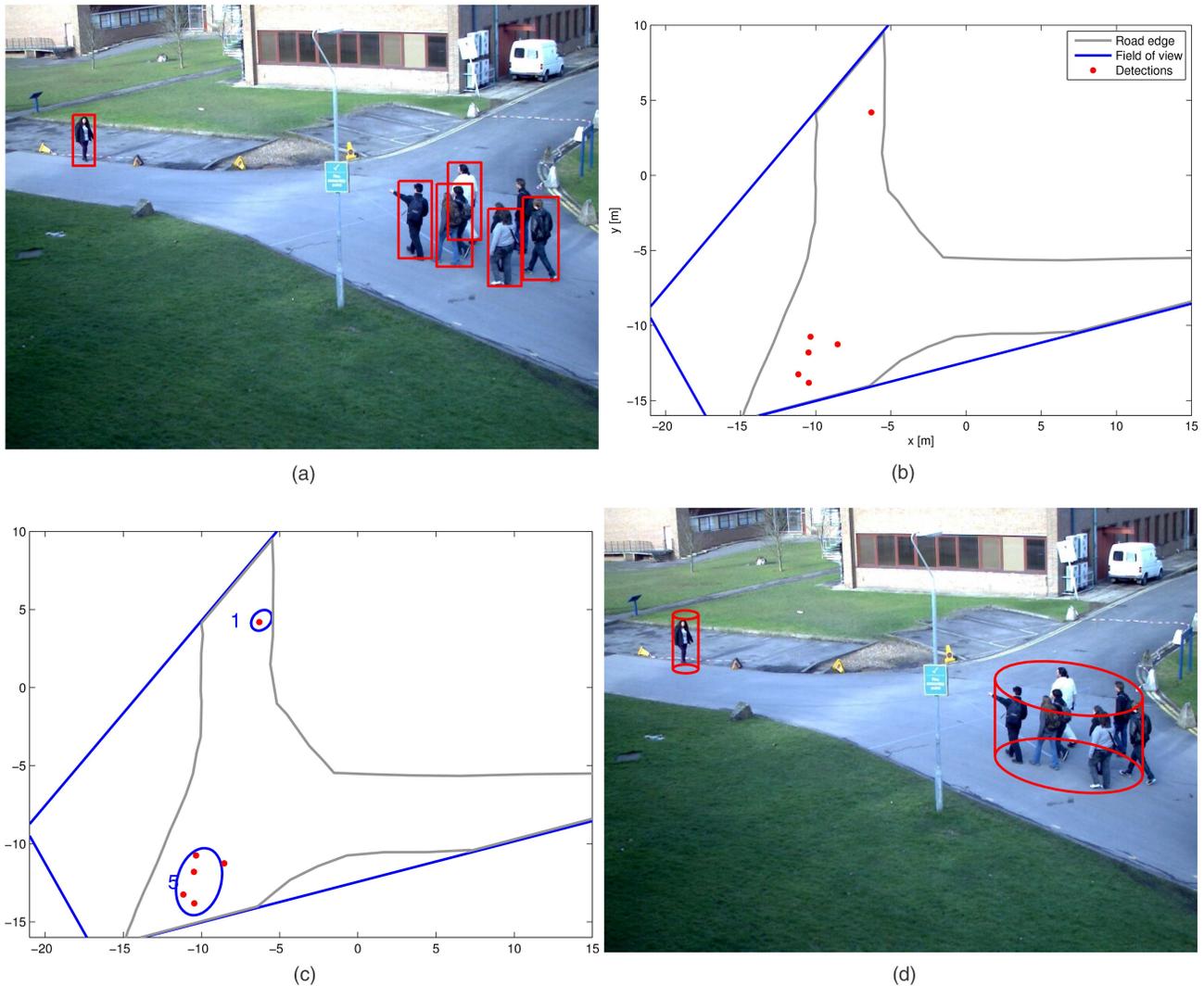


Fig. 12. Example application: tracking groups of pedestrians in video data. (a) Input image with pedestrian measurements in red. (b) Foot-print of measurements projected onto ground plane. (c) Group tracking output, numbers are estimated Poisson rates. (d) Output projected into input image, and visualised as elliptic cylinders.

the standard performance measures such as OSPA [169] and G-OSPA [152] are directly applicable.

## VI. EXTENDED OBJECT TRACKING APPLICATIONS

Extended object tracking algorithms have been applied in many different scenarios and have been evaluated using data from many different sensors such as LIDAR, camera, radar, RGB-depth (RGB-D) sensors, and unattended ground sensors (UGS). A list of references that contain experiments with real data is given in Table XIV. In this section we will present four example applications:

- Tracking groups of pedestrians using a camera overlooking a footpath.
- Tracking marine vessels using X-band radar.
- Tracking cars using a LIDAR mounted in the grille of an autonomous vehicle.
- Tracking objects with complex shapes using an RGB-D sensor.

TABLE XIV  
Experiments with different sensor types

Sensor	References
Automotive Radar	[29], [88], [91], [92], [100], [123], [166], [170]
Camera	[27], [28], [37], [44], [159], [160]
GMTI radar	[149]
Imaging Sonar	[103]
LIDAR	[19], [59], [60], [70], [73], [74], [78], [85], [119], [136], [140], [147], [155], [165], [172], [173], [201]
Marine Radar	[47], [75], [76], [171], [188], [189], [190]
RGB-D	[13], [14], [16], [49], [52]
Through wall radar	[65]
UGS, group tracking	[36]

These four examples are complementary in the sense that they illustrate different aspects of extended object tracking: different sensor modalities; the applicability of extended object methods to group object tracking;

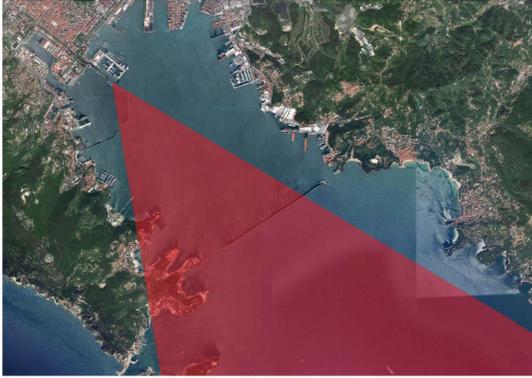


Fig. 13. Example application: tracking boats and ships using marine X-band radar. Aerial image of harbour, with sensor’s field of view shown in red.

object shapes of different complexities; and tracking in crowded scenarios with occlusions.

#### A. Tracking groups of pedestrians using camera

Automatic crowd surveillance is a complex task, and in scenes with a large number of persons it may be infeasible to track each person individually. In this case group object tracking using extended object MTT methods is a viable alternative, as this does not require tracking and identification of each individual. In the example presented here camera data is used to track groups of pedestrians that walk along a footpath. The online available PETS 2012 data set [185] is used for evaluation. For each image in the dataset a pedestrian detector [40], [41] is used, and the measurements are projected onto the ground plane using the camera parameters.

In this data the groups of pedestrians are loosely constructed and typically do not have a detailed shape that remains constant over time. Therefore the groups can be assumed to be elliptically shaped, and the random matrix measurement model can be used [102]. The ground plane measurements are input into a GGIW-PHD filter [78], [79], and the object extractions are projected back into the camera image for visualization. The GGIW-PHD filter is based on the Poisson model for the number of measurements from each group, i.e., for each group a Poisson rate parameter is estimated. This estimated rate can be taken as an estimate of the number of persons in the group.

Example results are shown in Figure 12.<sup>15</sup> The results show that the estimated ellipses are a good approximation of the pedestrian groups. The estimated Poisson rates tend to underestimate the number of persons in the group. The reason for this is that in groups with many persons, some individuals tend to be occluded and therefore are not detected. The estimated Poisson rate is more accurate when interpreted as a lower bound for the number of persons in the group, instead of interpreted as a count of the number of persons in the group.

<sup>15</sup>Video with tracking results: <https://youtu.be/jN-KXQqargE>.

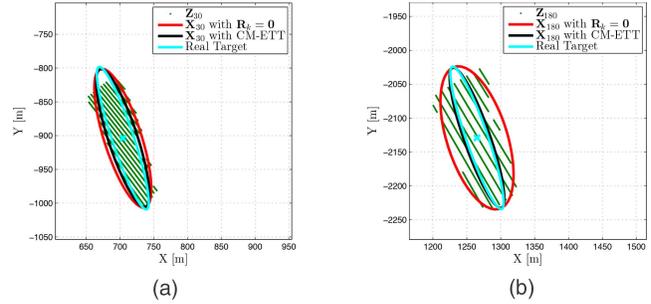


Fig. 14. Example application: tracking boats and ships using marine X-band radar. a) and b) Example detections (green dots), ground truth (real ellipse), and two estimates (red and black ellipses). The black ellipse is when the noise is correctly modelled, and the red ellipse shows the estimate when the noise is not modelled.

#### B. Tracking marine vessels using X-band radar

Harbours are busy places where many vessels share the water, from small boats to large ships. To keep track of where all the vessels are, marine X-band radar can be used [76]. These sensors produce high-resolution data that allow the tracking algorithm to estimate the size of the vessel, further allowing the possibility to classify the tracked vessels using prior information about the size of different ships and boats. An example is given in Figure 13, where the field of view of the sensor is overlaid on an aerial image of a harbour.

The raw sensor data is pre-processed using a Constant False Alarm Rate (CFAR) detector, producing polar detections (range and azimuth) [76]. Because boats and ships are best modelled in Cartesian coordinates, polar detections are converted to Cartesian coordinates [76]. The pre-processed data is suitable for use with the random matrix model, meaning that the shapes of the vessels are assumed to be ellipses. Typically neither boats, nor ships, are elliptically shaped, however, the major and minor axes of the estimated ellipses correspond to the length and width of the vessel. If measurement noise is modelled correctly low estimation errors can be achieved, however, if the noise is not modelled the size of the vessel is overestimated, especially in the cross-range dimension [188]. The significant difference between modelling the noise correctly, or not, is shown in Figure 14. If multiple radar sensors are used, the tracking results can be improved further [191].

#### C. Tracking cars using LIDAR

Autonomous active safety features are standard in many modern cars, and in both research and industry there is a considerable push towards fully driverless vehicles, see, e.g., [105]. For safe operation in dense scenes, such as inner city and other urban environments, an autonomous vehicle must be capable of keeping track of other objects, to avoid collisions. To this end, high resolution sensors such as LIDAR and extended object tracking algorithms can be used.

The high angular resolution of LIDAR sensors typically results in a large number of measurements for

each object. Thus, if an extended object tracking filter is not used, preprocessing is necessary to update the object estimates. Such preprocessing commonly consists of segmentation and clustering [94], [131], [151], shape fitting [133], or feature extraction [137]. The drawback of using such algorithms is that they are heavily dependent on parametrization, and often suffer from over-or under-segmentation. Especially in scenarios in which the environment changes, or when there are different object types, it is very difficult to find appropriate parameters. Because the tracking builds upon the data that is input, any error during segmentation and clustering will manifest itself as a tracking error.

In this section we will present experimental results where LIDAR sensors and an extended object PHD filter have been used to track cars; the results presented here are a subset of the results presented in [85]. The LIDAR sensor is assumed to be mounted in the grille of the ego vehicle, and the cars are assumed to be rectangular, with unknown length and width. The measurement modelling that was used is shown in Figure 5. The tracking problem is cast as a multiple model problem, and a multiple model PHD filter is used to track multiple cars. A full description of the tracking algorithm can be found in [85]. When there are multiple cars in the sensor’s field of view the cars may occlude each other, either partially or fully. To avoid losing track of cars that are occluded a non-homogeneous probability of measurement can be used. This is illustrated in Figure 15. Similar approaches to occlusion modelling are taken in [74], [78], [155], [207].

Experimental results in [85] show that the lateral position of the tracked cars can be estimated with an average error of less than 5 cm, while the average longitudinal position error is slightly larger, around 10 to 30 cm for different datasets. The shape parameters are estimated with an average error around 2 cm for the width, and around 20 cm for the length. The increased error in object length is due to the limited observability of the object length due to the aspect angle. Example detections and tracking results for a scenario with four cars is given in Figure 16, snapshots of this data are also shown in Figure 3.

#### D. Tracking complex shapes using RGB-D sensor

In this subsection we present an experimental setup where complex object shapes are estimated using RGB-D sensor data. This experiment has been published first in [13], [14], [16]. Specifically, a moving miniature railway vehicle is to be tracked from a bird’s eye view with the help of a RGB-D camera. An optical flow algorithm determines the velocity of each image point incorporating both the RGB and depth image sequences. Based on a threshold on the velocity, we obtain measurements, i.e., points classified as “moving,” that originate from the moving object. In this manner, a varying number

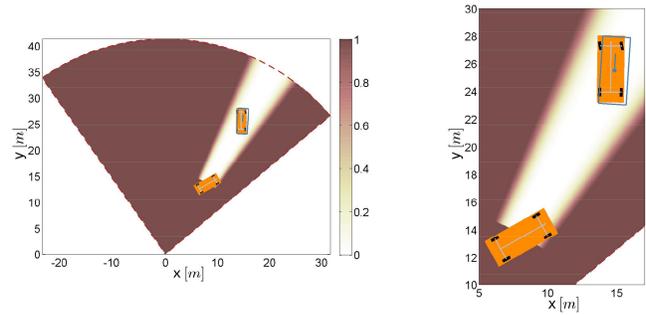


Fig. 15. Occlusion example. The sensor is located in the origin; darker color means higher probability of measurement; estimates in orange, ground truth in blue. Thanks to the use of an occlusion model the occluded car can be tracked with high accuracy while it traverses an area where it cannot be detected.

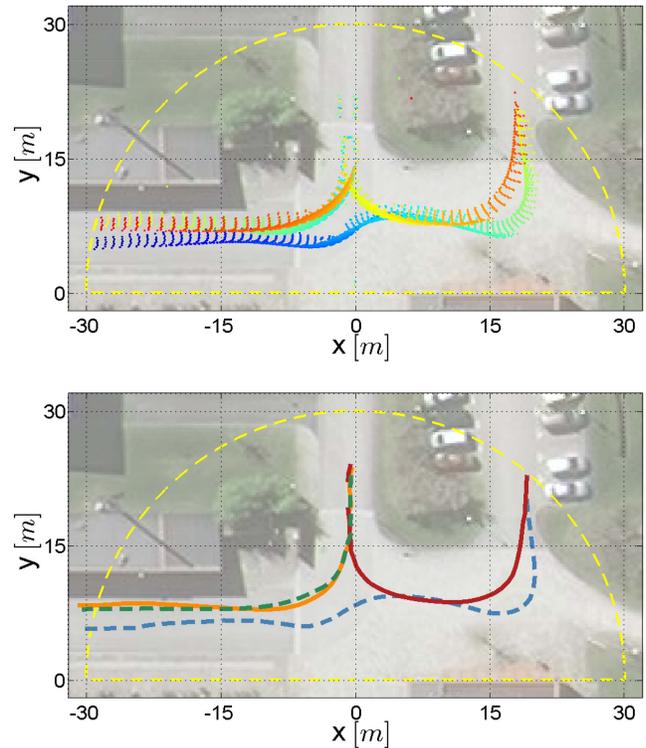


Fig. 16. Results from scenario with four cars. Top: sensor data, color coded according to time. Bottom: Estimated positions.

of noisy measurements from the object’s surface is received at each frame, see Figure 17 for an example frame. Due to the noisy images and inaccuracy of the optical flow algorithm, the measurements are noisy and do not completely fill the object surface. In fact, this is a typical extended object tracking problem where measurements come from a two-dimensional shape in two-dimensional space. Figure 18 shows example results with an implementation of the star-convex random hypersurface approach as discussed in Section III. Also, Figure 18 shows the result obtained from an active contour (snake) algorithm [99], which is a standard algorithm in computer vision. In general, an active contour model works with intensity/RGB images and not with point measurements. It calculates a contour by mini-

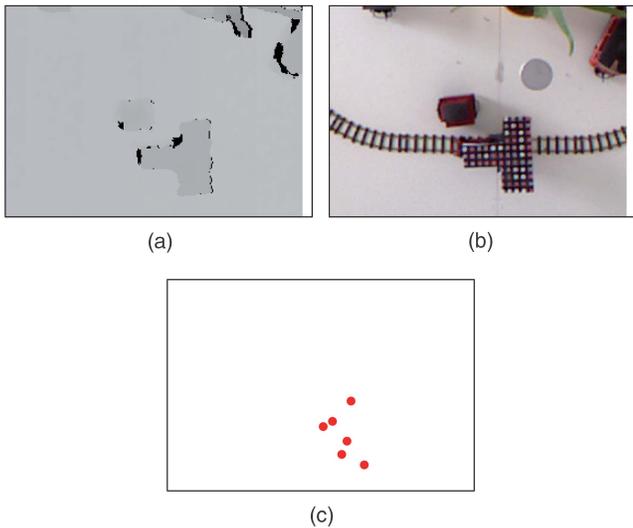


Fig. 17. Tracking a railway vehicle using a RGBD camera from a bird's eye view [14]. (a) Depth image. (b) RGB image. (c) Measurements.

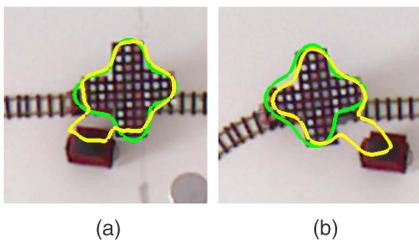


Fig. 18. Result for “+”-shaped vehicle: RHM (green) vs. active contour model using depth images (yellow) [14].

mizing an energy function [99] that is composed of an external force for pushing the contour to image features and an internal force for regularization. In this scenario, active contours are applied to the depth image and hence, can be unreliable in case the vehicle passes objects with similar depth, see Figure 18.

Alternatively, active contours can be applied to a “smoothed” version of the point measurements: the measurements are interpreted as an intensity image by placing a Gaussian kernel function at each measurement location. As indicated by Figure 19, active contours then aim at determining an enclosing curve of the point measurements in each frame. As the vehicle’s surface is not covered completely by the measurements in a single frame, active contours do not give a reasonable shape estimate. Active contours are not capable of systematically accumulating individual point measurements over time—without this capability no reasonable shape estimate can be expected.

## VII. SUMMARY AND CONCLUDING REMARKS

In this article we gave an introduction to extended object tracking, a comprehensive up-to-date overview of state-of-the-art research, and illustrated the methods using several different sensors and object types. Increasing

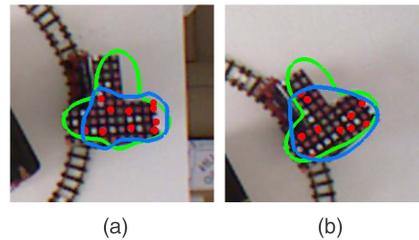


Fig. 19. Results for a “T”-shaped vehicle: RHM (green) vs. active contour model using (smoothed) point measurements (blue) [14].

sensor resolutions mean that there will be an increasing number of scenarios in which extended object methods can be applied. It is possible to cluster/segment the data in pre-processing and then apply standard point object methods, however, this requires careful parameter tuning, thereby increasing the risk for errors. Extended object tracking, on the other hand, uses Bayesian models for the multiple measurements per object, meaning that the tracking performance is much less dependent on clustering/segmentation.

During the last ten years an impressive number of new methods and applications have appeared in the literature, covering different approaches to extent modelling and multiple object tracking. This trend can be expected to continue, as there are many open questions to solve, and improvements that can be made. Due to the high non-linearity and high dimensionality of the problem, estimation of arbitrary shapes is still very much challenging. There is a need for performance bounds for extended object tracking methods: for a given shape model, how many measurements are required in order for the estimation algorithm to converge to an estimate with small error? Performance bounds may help in answering the question of which shape complexity is suitable when modelling the object. Naturally, in most applications one is interested in a shape description that is as precise as possible.

For arbitrary object shapes, the determination of suitable performance metrics for the evaluation of the shape estimate is still an open research question. Further, existing works on extended object tracking focus on single sensor systems (or perhaps systems with several very similar sensors). However, the fusion of complementary sensors like camera and LIDAR in an extended object tracking algorithm raises new challenges due to the different measurement principles and perception capabilities.

## REFERENCES

- [1] D. Angelova and L. Mihaylova “Extended object tracking using Monte Carlo methods,” *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 825–832, Feb. 2008.
- [2] D. Angelova, L. Mihaylova, N. Petrov, and A. Gning “A convolution particle filtering approach for tracking elliptical extended objects,” in *Proceedings of the International Conference on Information Fusion*, Jul. 2013, pp. 1542–1549.

- [3] T. Ardeshiri, U. Orguner, and F. Gustafsson  
“Bayesian inference via approximation of log-likelihood for priors in exponential family,”  
*CoRR*, vol. abs/1510.01225, 2015. [Online]. Available: <http://arxiv.org/abs/1510.01225>.
- [4] Y. Bar-Shalom  
“Extension of the probabilistic data association filter to multi-target tracking,”  
in *Proceedings of the Fifth Symposium on Nonlinear Estimation*, San Diego, CA, USA, Sep. 1974.
- [5] Y. Bar-Shalom and W. D. Blair  
*Multitarget-multisensor tracking: applications and advances, vol III*.  
Artech House, 2000.
- [6] Y. Bar-Shalom, F. Daum, and J. Huang  
“The probabilistic data association filter,”  
*IEEE Control Systems*, vol. 29, no. 6, pp. 82–100, Dec. 2009.
- [7] M. Baum, F. Faion, and U. Hanebeck  
“Modeling the target extent with multiplicative noise,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2406–2412.
- [8] M. Baum and U. Hanebeck  
“Random hypersurface models for extended object tracking,”  
in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Ajman, United Arab Emirates, Dec. 2009, pp. 178–183.
- [9] ———  
“Shape tracking of extended objects and group targets with star-convex RHMs,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 338–345.
- [10] ———  
“Extended object tracking with random hypersurface models,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 149–159, Jan. 2013.
- [11] M. Baum, V. Klumpp, and U. Hanebeck  
“A novel Bayesian method for fitting a circle to noisy points,”  
in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.
- [12] M. Baum, B. Noack, and U. Hanebeck  
“Extended object and group tracking with elliptic Random Hypersurface Models,”  
in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.
- [13] M. Baum  
“Student research highlight: Simultaneous tracking and shape estimation of extended targets,”  
*IEEE Aerospace and Electronic Systems Magazine*, vol. 27, no. 7, pp. 42–44, July 2012.
- [14] ———  
*Simultaneous tracking and shape estimation of extended objects*.  
KIT Scientific Publishing, 2013, vol. 13.
- [15] M. Baum, F. Faion, and U. D. Hanebeck  
“Modeling the target extent with multiplicative noise,”  
in *Proceedings of the 15th International Conference on Information Fusion (Fusion 2012)*, Singapore, Jul. 2012.
- [16] ———  
“Tracking ground moving extended objects using RGBD data,”  
in *Proceedings of the 2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2012)*, Hamburg, Germany, Sep. 2012.
- [17] M. Baum, M. Feldmann, D. Fränken, U. D. Hanebeck, and W. Koch  
“Extended object and group tracking: A comparison of Random Matrices and Random Hypersurface Models,”  
in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2010)*, Leipzig, Germany, Oct. 2010.
- [18] M. Beard, S. Reuter, K. Granström, V. B.-T., B.-N. Vo, and A. Scheel  
“A generalised labelled multi-Bernoulli filter for extended multi-target tracking,”  
in *Proceedings of the International Conference on Information Fusion*, Washington, DC, USA, Jul. 2015, pp. 991–998.
- [19] ———  
“Multiple extended target tracking with labelled random finite sets,”  
*IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1638–1653, Apr. 2016.
- [20] D. Bertsekas  
“The auction algorithm: A distributed relaxation method for the assignment problem,”  
*Annals of Operations Research*, vol. 14, no. 1, pp. 105–123, 1988.
- [21] R. Bhalla and H. Ling  
“A fast algorithm for signature prediction and image formation using the shooting and bouncing ray technique,”  
*IEEE Transactions on Antennas and Propagation*, vol. 43, no. 7, pp. 727–731, Jul. 1995.
- [22] C. M. Bishop  
*Pattern recognition and machine learning*.  
New York, USA: Springer, 2006.
- [23] S. Blackman and R. Popoli  
*Design and Analysis of Modern Tracking Systems*.  
Norwood, MA, USA: Artech House, 1999.
- [24] Y. Boers, H. Driessen, J. Torstensson, M. Trieb, R. Karlsson, and F. Gustafsson  
“A track before detect algorithm for tracking extended targets,”  
*IEEE Proceedings Radar, Sonar and Navigation*, vol. 153, no. 4, pp. 345–351, Aug. 2006.
- [25] S. Bordonaro, P. Willett, Y. Bar-Shalom, T. Luginbuhl, and M. Baum  
“Extended object tracking with exploitation of range rate measurements,”  
*Journal of Advances in Information Fusion*.
- [26] S. Bordonaro, P. Willett, Y. B. Shalom, M. Baum, and T. Luginbuhl  
“Extracting speed, heading and turn-rate measurements from extended objects using the EM algorithm,”  
in *IEEE Aerospace Conference*, Big Sky, Montana, USA, Mar. 2015.
- [27] T. J. Broida, S. Chandrashekhar, and R. Chellappa  
“Recursive 3-d motion estimation from a monocular image sequence,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 26, no. 4, pp. 639–656, Jul. 1990.
- [28] T. J. Broida and R. Chellappa  
“Estimating the kinematics and structure of a rigid object from a sequence of monocular images,”  
*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 497–513, Jun. 1991.
- [29] M. Buhren and B. Yang  
“Simulation of automotive radar target lists using a novel approach of object representation,”  
in *IEEE Intelligent Vehicles Symposium*, Jun. 2006, pp. 314–319.

- [30] E. R. Canfield  
“Engel’s inequality for bell numbers,”  
*Journal of Combinatorial Theory, Series A*, vol. 72, no. 1,  
pp. 184–187, 1995.
- [31] W. Cao, J. Lan, and X. R. Li  
“Extended object tracking and classification based on re-  
cursive joint decision and estimation,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Jul. 2013, pp. 1670–1677.
- [32] X. Cao, J. Lan, and X. R. Li  
“Extension-deformation approach to extended object track-  
ing,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Jul. 2016, pp. 1185–1192.
- [33] A. Carmi, F. Septier, and S. J. Godsill  
“The Gaussian mixture MCMC particle algorithm for dy-  
namic cluster tracking,”  
*Automatica*, vol. 48, no. 10, pp. 2454–2467, Oct. 2012.
- [34] N. Chernov  
*Circular and Linear Regression: Fitting Circles and Lines by  
Least Squares*.  
CRC Press, 2010.
- [35] D. E. Clark and S. J. Godsill  
“Group target tracking with the Gaussian mixture probabil-  
ity hypothesis density filter,”  
in *International Conference on Intelligent Sensors, Sensor  
Networks and Information (ISSNIP)*, Melbourne, Australia,  
Dec. 2007, pp. 149–154.
- [36] T. Damarla and L. Kaplan  
“A fusion architecture for tracking a group of people using  
a distributed sensor network,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Istanbul, Turkey, July 2013, pp. 1776–1783.
- [37] S. Davey, M. Wieneke, and H. Vu  
“Histogram-PMHT unfettered,”  
*IEEE Journal of Selected Topics in Signal Processing*, vol. 7,  
no. 3, pp. 435–447, June 2013.
- [38] J. Degerman, J. Wintenby, and D. Svensson  
“Extended target tracking using principal components,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Chicago, IL, USA, Jul. 2011, pp. 330–337.
- [39] J. C. Dezert  
“Tracking maneuvering and bending extended target in  
cluttered environment,”  
in *Signal and Data Processing of Small Targets*, O. E.  
Drummond, Ed., vol. 3373, Orlando, FL, USA, Apr. 1998,  
pp. 283–294.
- [40] P. Dollar, S. Belongie, and P. Perona  
“The fastest pedestrian detector in the west,”  
in *Proceedings of the British Machine Vision Conference*,  
2010, pp. 68.1–68.11.
- [41] P. Dollár, Z. Tu, P. Perona, and S. Belongie  
“Integral channel features,”  
in *Proceedings of the British Machine Vision Conference*,  
2009.
- [42] O. E. Drummond, S. S. Blackman, and K. C. Hell  
“Multiple sensor tracking of clusters and extended objects,”  
*Technical Proceedings: 1988 Tri-Service Data Fusion Sym-  
posium*, 1988.
- [43] O. E. Drummond, S. S. Blackman, and G. C. Pretrisor  
“Tracking clusters and extended objects with multiple sen-  
sors,”  
*Signal and Data Processing of Small Targets 1990*, vol. 1305,  
no. 1, pp. 362–375, 1990.
- [44] V. Edman, M. Andersson, K. Granström, and F. Gustafsson  
“Pedestrian group tracking using the GM-PHD filter,”  
in *Proceedings of the European Signal Processing Confer-  
ence*, Marrakech, Morocco, Sep. 2013.
- [45] K. Engel  
“On the average rank of an element in a filter of the partition  
lattice,”  
*Journal of Combinatorial Theory, Series A*, vol. 65, no. 1,  
pp. 67–78, 1994.
- [46] O. Erdinc, P. Willett, and Y. Bar-Shalom  
“Probability hypothesis density filter for multitarget multi-  
sensor tracking,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Philadelphia, CA, USA, Jul. 2005, pp. 146–  
153.
- [47] B. Errasti-Alcala and P. Braca  
“Track before detect algorithm for tracking extended targets  
applied to real-world data of X-band marine radar,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Salamanca, Spain, Jul. 2014.
- [48] F. Faion, M. Baum, and U. D. Hanebeck  
“Tracking 3D shapes in noisy point clouds with random  
hypersurface models,”  
in *Proceedings of the 15th International Conference on In-  
formation Fusion (Fusion 2012)*, Singapore, Jul. 2012.
- [49] F. Faion, M. Baum, A. Zea, and U. D. Hanebeck  
“Depth sensor calibration by tracking an extended object,”  
in *Proceedings of the 2015 IEEE International Conference  
on Multisensor Fusion and Integration for Intelligent Systems  
(MFI 2015)*, San Diego, California, USA, Sep. 2015.
- [50] F. Faion, A. Zea, M. Baum, and U. D. Hanebeck  
“Partial likelihood for unbiased extended object tracking,”  
in *Proceedings of the 18th International Conference on Infor-  
mation Fusion (Fusion 2015)*, Washington D.C., USA, Jul.  
2015.
- [51] ———  
“Symmetries in Bayesian extended object tracking,”  
*Journal of Advances in Information Fusion*, vol. 10, no. 1,  
pp. 13–30, Jun. 2015.
- [52] F. Faion, A. Zea, J. Steinbring, M. Baum, and U. D.  
Hanebeck  
“Recursive Bayesian pose and shape estimation of 3D  
objects using transformed plane curves,”  
in *Proceedings of the IEEE ISIF Workshop on Sensor Data  
Fusion: Trends, Solutions, Applications (SDF 2015)*, Bonn,  
Germany, Oct. 2015.
- [53] M. Fatemi, K. Granström, L. Svensson, F. Ruiz, and L. Ham-  
marstrand  
“Poisson multi-Bernoulli radar mapping using Gibbs sam-  
pling,”  
*IEEE Transactions on Signal Processing*.
- [54] M. Feldmann and D. Fränken  
“Tracking of extended objects and group targets using  
random matrices—A new approach,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Cologne, Germany, Jul. 2008.
- [55] M. Feldmann and D. Fränken  
“Advances on tracking of extended objects and group tar-  
gets using random matrices,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Seattle, USA, Jul. 2009, pp. 1029–1036.

- [56] M. Feldmann, D. Fränken, and J. W. Koch  
“Tracking of extended objects and group targets using random matrices,”  
*IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, Apr. 2011.
- [57] R. Fitzgerald  
“Development of practical PDA logic for multitarget tracking by microprocessor,”  
in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom, Ed. Artech House, 1990, pp. 1–23.
- [58] A. Fitzgibbon, M. Pilu, and R. B. Fisher  
“Direct least square fitting of ellipses,”  
*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 476–480, May 1999.
- [59] B. Fortin, R. Lherbier, and J. Noyer  
“A labeled PHD filter for extended target tracking in lidar data using geometric invariance properties: Vehicular application,”  
in *Proceedings of the International Conference on Information Fusion*, Istanbul, Turkey, July 2013, pp. 1744–1751.
- [60] ———  
“A track-before-detect approach for extended target tracking in multi-lidar systems using a low-level centralized fusion framework,”  
in *Proceedings of the International Conference on Information Fusion*, Salamanca, Spain, July 2014, pp. 1–8.
- [61] T. Fortmann, Y. Bar-Shalom, and M. Scheffe  
“Sonar tracking of multiple targets using joint probabilistic data association,”  
*IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–184, Jul. 1983.
- [62] D. Fränken, M. Schmidt, and M. Ulmke  
“‘Spooky action at a distance’ in the Cardinalized Probability Hypothesis Density Filter,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 4, pp. 1657–1664, Oct. 2009.
- [63] A. F. Garcia-Fernandez and B. N. Vo  
“Derivation of the PHD and CPHD filters based on direct Kullback-Leibler Divergence minimization,”  
*IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5812–5820, Nov. 2015.
- [64] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin  
*Bayesian Data Analysis*, ser. Texts in Statistical Science. Chapman & Hall/CRC, 2004.
- [65] G. Gennarelli, G. Vivone, P. Braca, F. Soldovieri, and M. G. Amin  
“Multiple extended target tracking for through-wall radars,”  
*IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6482–6494, Dec. 2015.
- [66] K. Gilholm, S. Godsill, S. Maskell, and D. Salmond  
“Poisson models for extended target and group tracking,”  
in *Proceedings of Signal and Data Processing of Small Targets*, vol. 5913. San Diego, CA, USA: SPIE, Aug. 2005, pp. 230–241.
- [67] K. Gilholm and D. Salmond  
“Spatial distribution model for tracking extended objects,”  
*IEE Proceedings of Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 364–371, Oct. 2005.
- [68] K. Granström, M. Fatemi, and L. Svensson  
“Gamma Gaussian inverse-Wishart Poisson multi-Bernoulli filter for extended target tracking,”  
in *Proceedings of the International Conference on Information Fusion*, Heidelberg, Germany, Jul. 2016.
- [69] ———  
“Poisson multi-Bernoulli conjugate prior for multiple extended object estimation,”  
*arXiv pre-print*, 2016. [Online]. Available: [arxiv.org/abs/1605.06311](http://arxiv.org/abs/1605.06311).
- [70] K. Granström and C. Lundquist  
“On the use of multiple measurement models for extended target tracking,”  
in *Proceedings of the International Conference on Information Fusion*, Istanbul, Turkey, Jul. 2013.
- [71] K. Granström, C. Lundquist, F. Gustafsson, and U. Orguner  
“Random set methods: Estimation of multiple extended objects,”  
*IEEE Robotics and Automation Magazine*, vol. 21, no. 2, pp. 73–82, Jun. 2014.
- [72] K. Granström, C. Lundquist, and U. Orguner  
“A Gaussian mixture PHD filter for extended target tracking,”  
in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.
- [73] ———  
“Tracking rectangular and elliptical extended targets using laser measurements,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 592–599.
- [74] ———  
“Extended target tracking using a Gaussian Mixture PHD filter,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.
- [75] K. Granström, A. Natale, P. Braca, G. Ludeno, and F. Serafino  
“PHD extended target tracking using an incoherent X-band Radar: Preliminary real-world experimental results,”  
in *Proceedings of the International Conference on Information Fusion*, Salamanca, Spain, Jul. 2014.
- [76] ———  
“Gamma Gaussian inverse Wishart probability hypothesis density for extended target tracking using X-band marine radar data,”  
*IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6617–6631, Dec. 2015.
- [77] K. Granström and U. Orguner  
“Properties and approximations of some matrix variate probability density functions,”  
Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Tech. Rep. LiTH-ISY-R-3042, Dec. 2011. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-88735>.
- [78] ———  
“A PHD filter for tracking multiple extended targets using random matrices,”  
*IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5657–5671, Nov. 2012.
- [79] ———  
“Estimation and maintenance of measurement rates for multiple extended target tracking,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2170–2176.
- [80] ———  
“Implementation of the GIW-PHD filter,”  
Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Tech. Rep. LiTH-ISY-R-3046, Mar. 2012. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-94585>.
- [81] ———  
“On the reduction of Gaussian inverse Wishart mixtures,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2162–2169.

- [82] ———  
“On spawning and combination of extended/group targets modeled with random matrices,”  
*IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 678–692, Feb. 2013.
- [83] ———  
“New prediction for extended targets with random matrices,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1577–1589, Apr. 2014.
- [84] K. Granström, U. Orguner, R. Mahler, and C. Lundquist  
“Corrections on: ‘Extended target tracking using a Gaussian-Mixture PHD filter’,”  
*IEEE Transactions on Aerospace and Electronic Systems*.
- [85] K. Granström, S. Reuter, D. Meissner, and A. Scheel  
“A multiple model PHD approach to tracking of cars under an assumed rectangular shape,”  
in *Proceedings of the International Conference on Information Fusion*, Salamanca, Spain, Jul. 2014.
- [86] K. Granström, P. Willett, and Y. Bar-Shalom  
“An extended target tracking model with multiple random matrices and unified kinematics,”  
in *Proceedings of the International Conference on Information Fusion*, Washington, DC, USA, Jul. 2015, pp. 1007–1014.
- [87] K. Granström  
“An extended target tracking model with multiple random matrices and unified kinematics,”  
*CoRR*, vol. abs/1406.2135, 2014. [Online]. Available: <http://arxiv.org/abs/1406.2135>.
- [88] J. Gunnarsson, L. Svensson, L. Danielsson, and F. Bengtsson  
“Tracking vehicles using radar detections,”  
in *IEEE Intelligent Vehicles Symposium*, Jun. 2007, pp. 296–302.
- [89] A. K. Gupta and D. K. Nagar  
*Matrix variate distributions*,  
ser. Chapman & Hall/CRC monographs and surveys in pure and applied mathematics. Chapman & Hall, 2000.
- [90] B. Habtemariam, R. Tharmarasa, T. Thayaparan, M. Mallick, and T. Kirubarajan  
“A multiple-detection joint probabilistic data association filter,”  
*IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 461–471, Jun. 2013.
- [91] L. Hammarstrand, M. Lundgren, and L. Svensson  
“Adaptive radar sensor model for tracking structured extended objects,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 1975–1995, Jul. 2012.
- [92] L. Hammarstrand, L. Svensson, F. Sandblom, and J. Sörstedt  
“Extended object tracking using a radar resolution model,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2371–2386, Jul. 2012.
- [93] D. Helbing and P. Molnár  
“Social force model for pedestrian dynamics,”  
*Physical Review E*, vol. 51, pp. 4282–4286, May 1995.
- [94] M. Himmelsbach, F. v Hundelshausen, and H. Wuensche  
“Fast segmentation of 3d point clouds for ground vehicles,”  
in *IEEE Intelligent Vehicles Symposium (IV)*, June 2010, pp. 560–565.
- [95] T. Hirscher, A. Scheel, S. Reuter, and K. Dietmayer  
“Multiple extended object tracking using Gaussian processes,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2016, pp. 868–875.
- [96] T. S. Huang and A. N. Netravali  
“Motion and structure from feature correspondences: a review,”  
*Proceedings of the IEEE*, vol. 82, no. 2, pp. 252–268, Feb 1994.
- [97] T. Jiang, M. Liu, Z. Fan, and S. Zhang  
“On multiple-model extended target multi-Bernoulli filters,”  
*Digital Signal Processing*, vol. 59, pp. 76–85, 2016.
- [98] S. J. Julier and J. K. Uhlmann  
“Unscented filtering and nonlinear estimation,”  
in *Proceedings of the IEEE*, vol. 92, no. 3, 2004, pp. 401–422.
- [99] M. Kass, A. Witkin, and D. Terzopoulos  
“Snakes: Active contour models,”  
*International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [100] C. Knill, A. Scheel, and K. Dietmayer  
“A direct scattering model for tracking vehicles with high-resolution radars,”  
in *IEEE Intelligent Vehicles Symposium*, Jun. 2016, pp. 298–303.
- [101] J. W. Koch and M. Feldmann  
“Cluster tracking under kinematical constraints using random matrices,”  
*Robotics and Autonomous Systems*, vol. 57, no. 3, pp. 296–309, Mar. 2009.
- [102] W. Koch  
“Bayesian approach to extended object and cluster tracking using random matrices,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, Jul. 2008.
- [103] D. Krout, W. Kooiman, G. Okopal, and E. Hanusa  
“Object tracking with imaging sonar,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, July 2012, pp. 2400–2405.
- [104] S. Kullback and R. A. Leibler  
“On information and sufficiency,”  
*The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [105] F. Kunz, D. Nuss, J. Wiest, H. Deusch, S. Reuter, F. Gritschneider, A. Scheel, M. Stäbler, M. Bach, P. Hatzelmann, C. Wild, and K. Dietmayer  
“Autonomous driving at Ulm university: A modular, robust, and sensor-independent fusion approach,”  
in *IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2015, pp. 666–673.
- [106] T. Kurien  
“Issues in the design of practical multitarget tracking algorithms,”  
*Chapter 3 in Multitarget-Multisensor Tracking: Advanced Applications*, Ed. Y. Bar-Shalom, Artech House, pp. 43–83, 1990.
- [107] J. Lan and X. R. Li  
“Tracking of extended object or target group using random matrix: New model and approach,”  
*IEEE Transactions on Aerospace and Electronic Systems*.
- [108] ———  
“Tracking of extended object or target group using random matrix—part I: New model and approach,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2177–2184.
- [109] ———  
“Tracking of extended object or target group using random matrix—part II: Irregular object,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2185–2192.

- [110] ———  
“Joint tracking and classification of non-ellipsoidal extended object using random matrix,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2014.
- [111] ———  
“Tracking maneuvering non-ellipsoidal extended object or target group using random matrix,”  
*IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2450–2463, 2014.
- [112] ———  
“Joint tracking and classification of extended object using random matrix,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2013, pp. 1550–1557.
- [113] ———  
“Extended object or group target tracking using random matrix with nonlinear measurements,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2016, pp. 901–908.
- [114] T. Lefebvre, H. Bruyninckx, and J. De Schutter  
“Kalman filters for non-linear systems: A comparison of performance,”  
*International Journal of Control*, vol. 77, no. 7, pp. 639–653, 2004.
- [115] X. R. Li and V. Jilkov  
“Survey of maneuvering target tracking: Part I. Dynamic models,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
- [116] F. Lian, C. Han, W. Liu, J. Liu, and J. Sun  
“Unified cardinalized probability hypothesis density filters for extended targets and unresolved targets,”  
*Signal Processing*, vol. 92, no. 7, pp. 1729–1744, 2012.
- [117] F. Lian, C.-Z. Han, W.-F. Liu, X.-X. Yan, and H.-Y. Zhou  
“Sequential Monte Carlo implementation and state extraction of the group probability hypothesis density filter for partly unresolvable group targets-tracking problem,”  
*IET Radar, Sonar and Navigation*, vol. 4, no. 5, pp. 685–702, Oct. 2010.
- [118] M. Liu, T. Jiang, and S. Zhang  
“The sequential Monte Carlo multi-Bernoulli filter for extended targets,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2015, pp. 984–990.
- [119] M. Luber, G. D. Tipaldi, and K. O. Arras  
“Place-dependent people tracking,”  
*International Journal for Robotics Research*, vol. 30, no. 3, pp. 280–293, Jan. 2011.
- [120] M. Lundgren, L. Svensson, and L. Hammarstrand  
“Variational Bayesian expectation maximization for radar map estimation,”  
*IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1391–1404, 2016.
- [121] C. Lundquist, K. Granström, and U. Orguner  
“Estimating the shape of targets with a PHD filter,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 49–56.
- [122] ———  
“An extended target CPHD filter and a Gamma Gaussian inverse Wishart implementation,”  
*IEEE Journal of Selected Topics in Signal Processing, Special Issue on Multi-target Tracking*, vol. 7, no. 3, pp. 472–483, Jun. 2013.
- [123] C. Lundquist, U. Orguner, and F. Gustafsson  
“Estimating polynomial structures from radar data,”  
in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.
- [124] D. Ma, F. Lian, and J. Liu  
“Sequential Monte Carlo implementation of cardinality balanced multi-target multi-Bernoulli filter for extended target tracking,”  
*IET Radar, Sonar Navigation*, vol. 10, no. 2, pp. 272–277, 2016.
- [125] R. Mahler  
*Statistical Multisource-Multitarget Information Fusion*.  
Norwood, MA, USA: Artech House, 2007.
- [126] ———  
“‘Statistics 102’ for multisensor-multitarget tracking,”  
*IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 376–389, Jun. 2013.
- [127] ———  
*Advances in Multisource-Multitarget Information Fusion*.  
Norwood, MA, USA: Artech House, 2014.
- [128] ———  
“Multitarget Bayes filtering via first-order multi target moments,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
- [129] ———  
“PHD filters of higher order in target number,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, Oct. 2007.
- [130] ———  
“PHD filters for nonstandard targets, I: Extended targets,”  
in *Proceedings of the International Conference on Information Fusion*, Seattle, WA, USA, Jul. 2009, pp. 915–921.
- [131] D. Meissner, S. Reuter, and K. Dietmayer  
“Combining the 2d and 3d world: A new approach for point cloud based object detection,”  
in *Intelligent Signal Processing Conference*, 12 2013.
- [132] L. Mihaylova, A. Carmi, F. Septier, A. Gning, S. Pang, and S. Godsill  
“Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking,”  
*Digital Signal Processing*, vol. 25, no. 0, pp. 1–16, Feb. 2014.
- [133] M. Munz  
“Generisches Sensorfusionsframework zur gleichzeitigen Zustands—und Existenzschätzung für die Fahrzeugumfeldererkennung,”  
Ph.D. dissertation, Ulm University, 2011.
- [134] J. Murphy, E. Özkan, P. Bunch, and S. J. Godsill  
“Sparse structure inference for group and network tracking,”  
in *2016 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 1208–1214.
- [135] K. Murty  
“An algorithm for ranking all the assignments in order of increasing cost,”  
*Operations Research*, vol. 16, no. 3, pp. 682–687, 1968.
- [136] L. E. Navarro-Serment, C. Mertz, and M. Hebert  
“Pedestrian Detection and Tracking Using Three-dimensional LADAR Data,”  
*International Journal for Robotics Research*, vol. 29, no. 12, pp. 1516–1528, Oct. 2010.
- [137] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart  
“A comparison of line extraction algorithms using 2d laser rangefinder for indoor mobile robotics,”  
in *Proceedings of the 2005 IEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 1929–1934.
- [138] U. Orguner  
“A variational measurement update for extended target tracking with random matrices,”  
*IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3827–3834, Jul. 2012.

- [139] U. Orguner, C. Lundquist, and K. Granström  
“Extended target tracking with a cardinalized probability hypothesis density filter.”  
Department of Electrical Engineering, Linköping University, Sweden, Linköping, Sweden, Tech. Rep. LiTH-ISY-R-2999, Mar. 2011. [Online]. Available: <http://www.control.isy.liu.se/research/reports/2011/2999.pdf>.
- [140] ———  
“Extended target tracking with a Cardinalized Probability Hypothesis Density filter,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 65–72.
- [141] U. Orguner  
(2010, Nov.) CPHD filter derivation for extended targets. ArXiv:1011.1512v2. [Online]. Available: <http://arxiv.org/abs/1011.1512v2>.
- [142] E. Özkan, N. Wahlström, and S. J. Godsill  
“Rao-Blackwellised particle filter for star-convex extended target tracking models,”  
in *Proceedings of the International Conference on Information Fusion*, July 2016, pp. 1193–1199.
- [143] S. K. Pang, J. Li, and S. Godsill  
“Detection and tracking of coordinated groups,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 472–502, Jan. 2011.
- [144] K. Panta, D. Clark, and B.-N. Vo  
“Data association and track management for the Gaussian mixture probability hypothesis density filter,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1003–1016, Jul. 2009.
- [145] N. Petrov, A. Gning, L. Mihaylova, and D. Angelova  
“Box particle filtering for extended object tracking,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2012, pp. 82–89.
- [146] N. Petrov, L. Mihaylova, A. Gning, and D. Angelova  
“A novel sequential Monte Carlo approach for extended object tracking based on border parametrisation,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 306–313.
- [147] N. Petrov, M. Ulmke, L. Mihaylova, A. Gning, M. Schikora, M. Wieneke, and W. Koch  
“On the performance of the box particle filter for extended object tracking using laser data,”  
in *Workshop on Sensor Data Fusion: Trends, Solutions, Applications*, Sep. 2012, pp. 19–24.
- [148] A. Petrovskaya and S. Thrun  
“Model based vehicle tracking for autonomous driving in urban environments,”  
in *Proceedings of Robotics: Science and Systems*, Zürich, Switzerland, Jun. 2008.
- [149] E. Pollard, B. Pannetier, and M. Rombaut  
“Performances in multitarget tracking for convoy detection over real GMTI data,”  
in *13th Conference on Information Fusion (FUSION)*, Edinburgh, Scotland, July 2010.
- [150] J. Porrill  
“Fitting ellipses and predicting confidence envelopes Using a bias corrected Kalman filter,”  
*Image Vision Computing*, vol. 8, pp. 37–41, 1990.
- [151] C. Premebida and U. Nunes  
“Segmentation and geometric primitives extraction from 2d laser range data for mobile robot applications,”  
in *Actas do Encontro Científico Robótica 2005*, 2005, pp. 17–25.
- [152] A. S. Rahmathullah, A. F. García-Fernández, and L. Svensson  
“Generalized optimal sub-pattern assignment metric,”  
*arXiv pre-print*, 2016. [Online]. Available: [arxiv.org/abs/1601.05585](http://arxiv.org/abs/1601.05585).
- [153] C. E. Rasmussen and C. K. I. Williams  
*Gaussian Processes for Machine Learning*.  
Boston, MA, USA: MIT Press, 2006.
- [154] D. Reid  
“An algorithm for tracking multiple targets,”  
*IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.
- [155] S. Reuter and K. Dietmayer  
“Pedestrian tracking using random finite sets,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 1101–1108.
- [156] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer  
“The Labeled Multi-Bernoulli Filter,”  
*IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, Jul. 2014.
- [157] S. Reuter, B. Wilking, and K. Dietmayer  
“Methods to model the motion of extended objects in multi-object Bayes filters,”  
in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 527–534.
- [158] P. Riekert and T. E. Schunck  
“Zur Fahrmechanik des gummibereiften Kraftfahrzeugs,”  
*Ingenieur-Archiv*, vol. 11, no. 3, pp. 210–224, 1940.
- [159] B. Ristic and J. Sherrah  
“Bernoulli filter for detection and tracking of an extended object in clutter,”  
in *Proceedings of 2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Apr. 2013, pp. 306–311.
- [160] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina  
“A tutorial on Bernoulli filters: Theory, implementation and applications,”  
*IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3406–3430, Jul. 2013.
- [161] G.-C. Rota  
“The number of partitions of a set,”  
*The American Mathematical Monthly*, vol. 71, no. 5, pp. 498–504, May 1964.
- [162] D. J. Salmond and M. C. Parr  
“Track maintenance using measurements of target extent,”  
*IEE Proceedings—Radar, Sonar and Navigation*, vol. 150, no. 6, pp. 389–395, Dec. 2003.
- [163] E. Saritas and U. Orguner  
“Posterior Cramer-Rao lower bounds for extended target tracking with random matrices,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2016, pp. 1485–1492.
- [164] T. Sathyan, T.-J. Chin, S. Arulampalam, and D. Suter  
“A multiple hypothesis tracker for multitarget tracking with multiple simultaneous measurements,”  
*IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 448–460, Jun. 2013.
- [165] A. Scheel, K. Granström, D. Meissner, S. Reuter, and K. Dietmayer  
“Tracking and data segmentation using a GGIW filter with mixture clustering,”  
in *Proceedings of the International Conference on Information Fusion*, Salamanca, Spain, Jul. 2014.
- [166] A. Scheel, C. Knill, S. Reuter, and K. Dietmayer  
“Multi-sensor multi-object tracking of vehicles using high-resolution radars,”  
in *IEEE Intelligent Vehicles Symposium*, Jun. 2016, pp. 558–565.
- [167] A. Scheel, S. Reuter, and K. Dietmayer  
“Using separable likelihoods for laser-based vehicle tracking with a labeled multi-Bernoulli filter,”  
in *Proceedings of the International Conference on Information Fusion*, Heidelberg, Germany, Jul. 2016, pp. 1200–1207.

- [168] D. Schramm, M. Hiller, and R. Bardini  
*Vehicle Dynamics: Modeling and Simulation*.  
Berlin, Heidelberg, Germany: Springer Berlin Heidelberg,  
2014.
- [169] D. Schuhmacher, B.-T. Vo, and B.-N. Vo  
“A consistent metric for performance evaluation of multi-  
object filters,”  
*IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp.  
3447–3457, Aug. 2008.
- [170] M. Schuster, J. Reuter, and G. Wanielik  
“Probabilistic data association for tracking extended group  
targets under clutter using random matrices,”  
*Journal of Advances in Information Fusion*.
- [171] ———  
“Probabilistic data association for tracking extended group  
targets under clutter using random matrices,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Washington, DC, USA, Jul. 2015, pp. 961–968.
- [172] M. Schutz, N. Appenrodt, J. Dickmann, and K. Dietmayer  
“Multiple extended objects tracking with object-local occu-  
pancy grid maps,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Salamanca, Spain, July 2014.
- [173] L. Spinello, R. Triebel, and R. Siegwart  
“Multiclass multimodal detection and tracking in urban  
environments,”  
*International Journal for Robotics Research*, vol. 29, no. 12,  
pp. 1498–1515, Oct. 2010.
- [174] J. Steinbring, M. Baum, A. Zea, F. Faion, and U. Hanebeck  
“A closed-form likelihood for particle filters to track ex-  
tended objects with star-convex RHMs,”  
in *Proceedings of the 2015 IEEE International Conference on  
Multisensor Fusion and Information Integration (MFI 2015)*,  
San Diego, California, USA, Sep. 2015.
- [175] R. Streit  
“JPDA intensity filter for tracking multiple extended ob-  
jects in clutter,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*. ISIF, 2016, pp. 1477–1484.
- [176] R. Streit and T. E. Luginbuhl  
“A probabilistic multihypothesis tracking algorithm with-  
out enumeration and pruning,”  
in *In proceedings of the Sixth Joint Service Data Fusion  
Symposium*, Laurel, MD, USA, Jun. 1993, pp. 1015–1024.
- [177] R. L. Streit and T. E. Luginbuhl  
“Probabilistic Multi-Hypothesis Tracking,”  
DTIC Document, Tech. Rep., 1995.
- [178] L. Sun, S. Zhang, B. Ji, and J. Pu  
“Performance evaluation for shape estimation of extended  
objects using a modified Hausdorff distance,”  
in *IEEE International Conference on Information and Au-  
tomation (ICIA)*, Aug. 2016, pp. 780–784.
- [179] L. Sun, J. Lan, and X. R. Li  
“Joint tracking and classification of extended object based  
on support functions,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*. IEEE, 2014, pp. 1–8.
- [180] ———  
“Modeling for tracking of complex extended object using  
Minkowski addition,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*. IEEE, 2014, pp. 1–8.
- [181] L. Sun, X. R. Li, and J. Lan  
“Modeling of extended objects based on support functions  
and extended Gaussian images for target tracking,”  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 50, no. 4, pp. 3021–3035, 2014.
- [182] A. Swain and D. Clark  
“Extended object filtering using spatial independent cluster  
processes,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Edinburgh, UK, Jul. 2010.
- [183] ———  
“The PHD filter for extended target tracking with estimable  
shape parameters of varying size,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Singapore, Jul. 2012.
- [184] X. Tang, X. Chen, M. McDonald, R. Mahler, R. Tharmarasa,  
and T. Kirubarajan  
“A multiple-detection probability hypothesis density filter,”  
*IEEE Transactions on Signal Processing*, vol. 63, no. 8, pp.  
2007–2019, Apr. 2015.
- [185] University of Reading  
“PETS 2012 dataset S1: Person count and density estima-  
tion,”  
Jan. 2012. [Online]. Available: [http://www.cvg.rdg.ac.uk/  
PETS2012/a.html](http://www.cvg.rdg.ac.uk/PETS2012/a.html).
- [186] J. Vermaak, N. Ikoma, and S. J. Godsill  
“Sequential Monte Carlo framework for extended object  
tracking,”  
*IEE Proceedings of Radar, Sonar and Navigation*, vol. 152,  
no. 5, pp. 353–363, Oct. 2005.
- [187] G. Vivone and P. Braca  
“Joint probabilistic data association tracker for extended  
target tracking applied to x-band marine radar data,”  
*IEEE Journal of Oceanic Engineering*, vol. 41, no. 4, pp.  
1007–1019, Oct. 2016.
- [188] G. Vivone, P. Braca, K. Granström, A. Natale, and J. Chanus-  
sot  
“Converted measurements bayesian extended target track-  
ing applied to X-band marine radar data,”  
*Journal of Advances in Information Fusion*.
- [189] ———  
“Converted measurements random matrix approach to ex-  
tended target tracking using X-band marine radar data,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Washington, DC, USA, Jul. 2015, pp. 976–983.
- [190] G. Vivone, P. Braca, K. Granström, and P. Willett  
“Multistatic Bayesian extended target tracking,”  
*IEEE Transactions on Aerospace and Electronic Systems*.
- [191] G. Vivone, P. Braca, K. Granström, and P. Willett  
“Multiple sensor Bayesian extended target tracking fusion  
approaches using random matrices,”  
in *Proceedings of the International Conference on Informa-  
tion Fusion*, Heidelberg, Germany, Jul. 2016.
- [192] B.-N. Vo, S. Singh, and A. Doucet  
“Sequential Monte Carlo methods for multitarget filtering  
with random finite sets,”  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 41, no. 4, pp. 1224–1245, Oct. 2005.
- [193] B.-N. Vo, B.-T. Vo, and D. Clark  
“Bayesian multiple target filtering using random finite  
sets,”  
in *Integrated Tracking, Classification, and Sensor Manage-  
ment*, M. Mallick, V. Krishnamurthy, and B.-N. Vo, Eds.  
New York, NY, USA: Wiley, 2014.
- [194] B.-T. Vo and B.-N. Vo  
“Labeled random finite sets and multi-object conjugate  
priors,”  
*IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp.  
3460–3475, Apr. 2013.

- [195] B.-T. Vo, B.-N. Vo, and A. Cantoni  
“The cardinality balanced multi-target multi-Bernoulli filter and its implementations,”  
*IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 409–423, Feb. 2009.
- [196] B.-T. Vo, B.-N. Vo, and D. Phung  
“Labeled random finite sets and the Bayes multi-target tracking filter,”  
*IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6554–6567, Dec. 2014.
- [197] B.-N. Vo, M. Mallick, Y. Bar-Shalom, S. Coraluppi, R. Osborne, R. Mahler, and B.-T. Vo  
“Multitarget tracking,”  
*Wiley Encyclopedia of Electrical and Electronics Engineering*, Sep. 2015.
- [198] N. Wahlström and E. Özkan  
“Extended target tracking using Gaussian processes,”  
*IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4165–4178, Aug. 2015.
- [199] M. J. Waxman and O. E. Drummond  
“A bibliography of cluster (group) tracking,”  
*Signal and Data Processing of Small Targets 2004*, vol. 5428, no. 1, pp. 551–560, 2004.
- [200] M. Wieneke and S. Davey  
“Histogram PMHT with target extent estimates based on random matrices,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 1–8.
- [201] M. Wieneke and W. Koch  
“A PMHT approach for extended objects and object groups,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2349–2370, 2012.
- [202] W. Wieneke and J. W. Koch  
“Probabilistic tracking of multiple extended targets using random matrices,”  
in *Proceedings of SPIE Signal and Data Processing of Small Targets*, Orlando, FL, USA, Apr. 2010.
- [203] P. Willett, Y. Ruan, and R. Streit  
“PMHT: Problems and some solutions,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 738–754, Jul 2002.
- [204] J. Williams  
“An efficient, variational approximation of the best fitting multi-Bernoulli filter,”  
*IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 258–273, Jan. 2015.
- [205] ———  
“Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based MeMBeR,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 1664–1687, Jul. 2015.
- [206] K. Wyffels and M. Campbell  
“Priority-based tracking of extended objects,”  
*Journal of Advances in Information Fusion*.
- [207] ———  
“Negative information for occlusion reasoning in dynamic extended multiobject tracking,”  
*IEEE Transactions on Robotics*, vol. 31, no. 2, pp. 425–442, Apr. 2015.
- [208] J. Yang, F. Liu, H. Ge, and Y. Yuan  
“Multiple extended target tracking algorithm based on gmphd filter and spectral clustering,”  
*EURASIP Journal on Advances in Signal Processing*, Jul. 2014.
- [209] S. Yang and M. Baum  
“Second-order extended Kalman filter for extended object and group tracking,”  
in *Proceedings of the International Conference on Information Fusion*, July 2016, pp. 1178–1184.
- [210] ———  
“Extended Kalman filter for extended object tracking,”  
in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, USA, Mar. 2017.
- [211] S. Yang, M. Baum, and K. Granström  
“Metrics for performance evaluation of elliptic extended object tracking methods,”  
in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*, Baden-Baden, Germany, Sep. 2016.
- [212] A. Yilmaz, O. Javed, and M. Shah  
“Object tracking: A survey,”  
*ACM Computing Surveys*, vol. 38, no. 4, Dec. 2006.
- [213] A. Zea, F. Faion, M. Baum, and U. Hanebeck  
“Level-set Random Hypersurface Models for tracking non-convex extended objects,”  
in *Proceedings of the International Conference on Information Fusion*, Istanbul, Turkey, Jul. 2013.
- [214] A. Zea, F. Faion, and U. D. Hanebeck  
“Tracking elongated extended objects using splines,”  
in *Proceedings of the International Conference on Information Fusion*, July 2016, pp. 612–619.
- [215] ———  
“Exploiting clutter: Negative information for enhanced extended object tracking,”  
in *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)*, Washington D. C., USA, Jul. 2015.
- [216] A. Zea, F. Faion, J. Steinbring, and U. D. Hanebeck  
“Exploiting negative measurements for tracking star-convex extended objects,”  
in *Proceedings of the 2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2016)*, Baden-Baden, Germany, Sep. 2016.
- [217] D. Zhang and G. Lu  
“Study and evaluation of different Fourier methods for image retrieval,”  
*Image and Vision Computing*, vol. 23, no. 1, pp. 33–49, 2005.
- [218] G. Zhang, F. Lian, and C. Han  
“CBMeMBeR filters for nonstandard targets, i: Extended targets,”  
in *Proceedings of the International Conference on Information Fusion*, Jul. 2014.
- [219] H. Zhang, H. Xu, X.-Y. Wang, and W. An  
“A PHD filter for tracking closely spaced objects with elliptic Random Hypersurface models,”  
in *Proceedings of the International Conference on Information Fusion*. IEEE, 2013, pp. 1558–1565.
- [220] Y. Zhang and H. Ji  
“A novel fast partitioning algorithm for extended target tracking using a Gaussian mixture PHD filter,”  
*Signal Processing*, vol. 93, no. 11, pp. 2975–2985, 2013.
- [221] ———  
“Robust Bayesian partition for extended target Gaussian inverse Wishart PHD filter,”  
*IET Signal Processing*, vol. 8, no. 4, pp. 330–338, Jun. 2014.
- [222] Z. Zhang  
“Parameter estimation techniques: A tutorial with application to conic fitting,”  
*Image and Vision Computing*, vol. 15, no. 1, pp. 59–76, 1997.
- [223] Z. Zhong, H. Meng, and X. Wang  
“Extended target tracking using an IMM based Rao-Blackwellised Unscented Kalman filter,”  
in *Proceedings of the International Conference on Signal Processing*, Oct. 2008, pp. 2409–2412.

- [224] H. Zhu, C. Han, and C. Li  
“An extended target tracking method with random finite set observations,”  
in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 73–78.

- [225] P. Zong and M. Barbary  
“Improved multi-Bernoulli filter for extended stealth targets tracking based on sub-random matrices,”  
*IEEE Sensors Journal*, vol. 16, no. 5, pp. 1428–1447, March 2016.



**Karl Granström** (M’08) is a postdoctoral research fellow at the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden. He received the MSc degree in Applied Physics and Electrical Engineering in May 2008, and the PhD degree in Automatic Control in November 2012, both from Linköping University, Sweden. He previously held postdoctoral positions at the Department of Electrical and Computer Engineering at University of Connecticut, USA, from September 2014 to August 2015, and at the Department of Electrical Engineering of Linköping University from December 2012 to August 2014. His research interests include estimation theory, multiple model estimation, sensor fusion and target tracking, especially for extended targets. He has received paper awards at the Fusion 2011 and Fusion 2012 conferences.



**Marcus Baum** is Juniorprofessor (Assistant Professor) at the University of Goettingen, Germany. He received the Diploma degree in computer science from the University of Karlsruhe (TH), Germany, in 2007, and graduated as Dr.-Ing. (Doctor of Engineering) at the Karlsruhe Institute of Technology (KIT), Germany, in 2013. From 2013 to 2014, he was postdoc and assistant research professor at the University of Connecticut, CT, USA. His research interests are in the field of data fusion, estimation, and tracking. Marcus Baum is associate administrative editor of the “Journal of Advances in Information Fusion (JAIF)” and served as local arrangement chair of the “19th International Conference on Information Fusion (FUSION 2016).” He received the best student paper award at the FUSION 2011 conference.



**Stephan Reuter** is Juniorprofessor (Assistant Professor) at Ulm University, Germany. He received the Diploma degree (equivalent to M.Sc. degree) and the Dr.-Ing. degree (equivalent to Ph.D.) in electrical engineering from Ulm University, Germany, in 2008 and 2014, respectively. Since 2008 he is a research assistant at the Institute of Measurement, Control and Microtechnology at Ulm University. He received the best Student Paper award at FUSION 2014 and the Uni-DAS award for an excellent PhD thesis in the field of driver assistance systems. His main research topics are sensor data fusion, multi-object tracking, extended object tracking, environment perception for intelligent vehicles, and sensor data processing.

# Multi Detection Joint Integrated Probabilistic Data Association Using Random Matrices with Applications to Radar-Based Multi Object Tracking

MICHAEL SCHUSTER  
JOHANNES REUTER  
GERD WANIELIK

**In extended object tracking, a target is capable to generate more than one measurement per scan. Assuming the target being of elliptical shape and given a point cloud of measurements, the Random Matrix Framework can be applied to concurrently estimate the target's dynamic state and extension. If the point cloud contains also clutter measurements or origins from more than one target, the data association problem has to be solved as well. However, the well-known joint probabilistic data association method assumes that a target can generate at most one detection. In this article, this constraint is relaxed, and a multi-detection version of the joint integrated probabilistic data association is proposed. The data association method is then combined with the Random Matrix framework to track targets with elliptical shape. The final filter is evaluated in the context of tracking smaller vessels using a high resolution radar sensor. The performance of the filter is shown in simulation and in several experiments.**

Manuscript received October 31, 2015; revised May 31, 2016; released for publication October 17, 2016.

Refereeing of this contribution was handled by Karl Granstrom.

Authors' addresses: M. Schuster and J. Reuter, Institute of System Dynamics, Konstanz University of Applied Sciences (E-mail: michael.schuster,johannes.reuter@htwg-konstanz.de). G. Wanielik, Professorship of Communication Engineering, Chemnitz University of Technology (E-mail: gerd.wanielik@etit.tu-chemnitz.de).

1557-6418/17/\$17.00 © 2017 JAIF

## I. INTRODUCTION

Radar systems have become standard for many automotive applications like adaptive cruise control or lane change assistance. Customarily, these sensors have the advantage that they need low mounting space, have low power consumption, are available at low cost, and still have a good resolution for ranges up to approx. 200 m. These features make the sensor also interesting for alternative applications. In this paper, these kind of radar sensors are considered for application in marine environment. Radars in this context typically operate with 3 GHz or 9 GHz and, as a consequence, have rather large apertures and high energy consumption. Small unmanned surface vessels (USV) usually do not have sufficient space or energy resources for such systems. On the other hand, very often these types of vessels operate in harbors or on rivers and in general require only short range surveillance [1]. Thus, automotive radar sensors (ARS) are an interesting alternative.

When applying these sensors in marine environment, the extension of a scanned vessel in comparison to sensor resolution is very high. Hence, at each scan, a point cloud of detections from an object is provided by the sensor. This leads to an extended target tracking problem. In order to solve this, numerous algorithms have already been proposed, see e.g. the surveys in [2] and [3].

Assuming the sensor does not generate stable but fast fluctuating reflection centers, an estimation of the target extent can be obtained by analyzing the noise distribution. If the measurements are randomly distributed over the target extent, or the noise of the measurements is correlated with the target's size, [4] presented an approach for simultaneously estimating the state and extension of a target. There, the target's physical extension is assumed to be of an elliptical shape and is represented by a symmetric positive definite random matrix. For many real sensors, the measurement spread is only partially dependent on the target's extent and also on the sensor's accuracy. Thus, [5] made the proposition to model this spread as a linear combination of extension noise and measurement noise. Due to the heuristics in [5], [6] derived a more complex filter update step which improves the estimation results. A unification of [4] and [5] was proposed in [7] and further extended for non-elliptical models in [8]. An alternative for arbitrary shapes is presented in [9], where Random Hypersurface Models are used to estimate the extent of an object.

Besides the pure state estimation task, the measurement to track data association problem has to be solved. A typical problem in Extended Target Multi Object Tracking (ET-MOT) is shown in Figure 1. Each object generates several detections, and, in combination with the clutter measurements, using the detections only it is unclear which target created how many detections, and where these measurements are located. A first data association method in the context of the Random Matrix

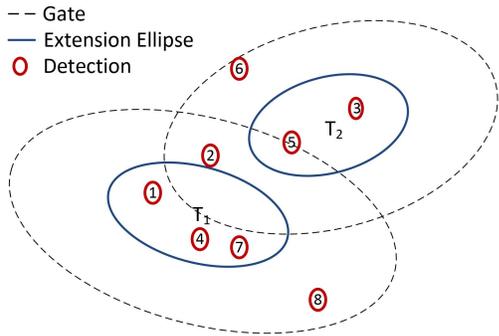


Fig. 1. Typical situation in extended object tracking if the targets have an elliptical shape. The dashed lines indicate a typical association threshold. Only detections within are considered for update of the central track.

framework is given in [10] and [11], where the Probabilistic Multi-Hypothesis algorithm is applied. Using Random Finite Sets, various Multi Object trackers have been presented, e.g. in [12]–[14]. Although the results of the Labeled Multi Bernoulli in [14] are very promising, in the authors opinion it is worth to also take a look at more traditional approaches and to investigate how these can be modified to be applied in ET-MOT.

One of the most popular methods for assigning measurements to a track is the Probabilistic Data Association (PDA) filter [15]. It performs a measurement update for each possible association and computes for each association the corresponding likelihood that the selected measurement is correct, i.e. was originated by the target. The a posteriori estimate of the track is then given by the weighted sum of the updated state estimates. In case that more than one object is present, the PDA is extended to the Joint PDA (JPDA), which also reflects the partitioning of the measurements to the tracks when calculating the association likelihood.

Besides the state estimate, for a multi-object tracker, the existence probability is of relevance, as well. Therefore, [16] modified the PDA to integrate also the estimation of the existence likelihood into the filtering and data association process. The new methods are named IPDA for the single object case and JIPDA for the multi-object case [17], respectively, where the letter ‘T’ stands for “Integrated.”

All these PDA algorithms rely on the common assumption that an object can generate at most one measurement during a sensor scan cycle. As already stated in the abstract, for extended object tracking, this assumption does not hold in general. To make PDA feasible in this context, spatial clustering of the detections in combination with a JPDA is proposed in [18]. A version of the PDA that can handle more than one detection was presented in [19] and a Multi-Detection JPDA in [20]. The MD-JPDA was used to handle multi-path reflections from over-the-horizon radars. With the Generalized PDA, also the existence estimation was introduced into the MD-PDA by [21]. Since for multi-target tracking algorithms, the existence likelihood estimation is

essential, in this paper, a Multi-Detection JIPDA (MD-JIPDA) is derived, as an extension to the JIPDA to assign more than one measurement to an object.

The structure of this paper is as follows: First, the general derivation of the MD-JIPDA is given in section II. In section III, an implementation of the MD-JIPDA that uses the concept of Random Matrices is proposed. It gives a short introduction to Random Matrices and also illustrates an approximation scheme of the MD-JIPDA to make it real-time applicable. Some results of Monte Carlo Simulations for different sensor characteristics and scenarios are shown in section IV. In the context of tracking smaller vessels in close distance using an automotive radar sensor on USV, also some experimental results are presented in section V, and conclusions are drawn in section VI.

## II. MULTI DETECTION-JOINT INTEGRATED PROBABILISTIC DATA ASSOCIATION

In the original Integrated Probabilistic Data Association (IPDA) approach, a track is considered to consist of two components [22, p. 142]: The object’s dynamic states and the object’s existence estimates. While the dynamic state is a continuous random variable, the existence is a binary variable that only can take the values “object exists” and “object does not exist.” One central assumption of the IPDA filter, as proposed in [16], is that an object can generate at most one measurement at time  $k$ . The Generalized Probabilistic Data Association Filter relaxes this constraint in such a way that up to  $n_{\max}$  measurements can be originated by a single object. The obtained GPDA filter can be applied for multi-object cases only if the objects are well separated in the measurement space. If the targets are in close proximity to each other, the GPDA will tend to merge tracks. To avoid this, in this section, a new algorithm, that can be used to consider joint track to measurement associations is described.

In the following, the dynamic state of track  $t$  at time  $k$  will be referred to as  $\mathbf{x}_k^t$ , and  $\chi_k^t$  describes the event that the object exists and  $\bar{\chi}_k^t$  the complementary event. If a set of new measurements is received, for each track, only those measurements are considered that fall into some elliptical association gate with constant probability  $P_G$ . Tracks that share at least one common measurement in their association gate are put together in a new association cluster, and a joint cluster volume  $V_k$  can be computed. Since gating and track clustering are commonly known techniques, the reader is referred to the literature for further information, e.g. [23, p. 334ff]. The set of  $n_k$  gated measurements will be denoted as  $\mathbf{Z}_k = \{\mathbf{z}_k^1, \dots, \mathbf{z}_k^{n_k}\}$  and  $\mathbf{Z}^k$  the measurements received up to time  $k$ .

From the Bayes' filter update equation, the posterior probability of the target existence is given by:

$$p(\chi_k | \mathbf{Z}^k) = \frac{p(\mathbf{Z}_k | \chi_k^t) p(\chi_k | \mathbf{Z}^{k-1})}{p(\mathbf{Z}_k | \bar{\chi}_k^t) p(\bar{\chi}_k^t | \mathbf{Z}^{k-1}) + p(\mathbf{Z}_k | \chi_k^t) p(\chi_k^t | \mathbf{Z}^{k-1})} \quad (1)$$

The predicted existence probability from time  $k-1$  is computed from

$$p(\chi_k^t | \mathbf{Z}^{k-1}) = p_{\text{surv}} \cdot p(\chi_k^t | \mathbf{Z}^{k-1}), \quad (2)$$

where, for simplicity, the probability of object survival  $p_{\text{surv}}$  is here considered as state independent. Using a proper existence time constant  $\tau_\chi$  for a track, the probability of survival is only conditioned on the sample time  $T$  with

$$p_{\text{surv}} = e^{-T/\tau_\chi}. \quad (3)$$

To evaluate the likelihoods in (1), the following assumptions are made:

- Clutter is uniformly distributed over the volume  $V$ :

$$p_s^{Cl}(n) = V^{-n}$$

- The number of clutter measurements is Poisson distributed with constant clutter rate  $\lambda$ :

$$p_c^{Cl}(n) = \frac{(\lambda V)^n}{n!} e^{-\lambda V}$$

- An object is detected by the sensor with probability  $P_D$ .
- The number of measurements  $m$  generated by one object  $t$  follows an arbitrary distribution  $p_c^t(m)$ .
- The spatial model for a single measurement of an object  $t$  is denoted  $p_s^t(\mathbf{z}_k | \mathbf{x}_k)$ , which is typically a Gaussian of the form

$$p_s^t(\mathbf{z}_k | \mathbf{x}_k^t) = \frac{1}{P_G} \mathcal{N}(\mathbf{z}_k; \mathbf{H}\mathbf{x}_k, \Sigma_k),$$

where  $\mathbf{H}$  is the observation matrix and  $\Sigma_k$  the corresponding covariance.

#### A. Joint Association Probability

Let  $\mathcal{H}_{i,j}$  denote a joint association event that describes one hypothesis how the  $n_k$  measurements within an association cluster have been created. The joint association events can be grouped into joint detection events  $\mathcal{D}_i$  which assign the same detection count pattern. One pattern assigns a specific measurement count  $m_i$  to track  $t$ , e.g. for the situation in Figure 1 the event  $\mathcal{H}_{i,1} = \{\{4, 7\}, \{3, 5, 6\}\}$  that assigns measurements  $M4$  and  $M7$  to track one, and  $M3, M5, M6$  to track two belong to the same detection event  $\mathcal{D}_i$  as  $\mathcal{H}_{i,2} = \{\{1, 7\}, \{3, 5, 6\}\}$ .

The a posteriori probability of a joint event  $\mathcal{H}_{i,j}$  is defined as

$$\begin{aligned} p(\mathcal{H}_{i,j} | \mathbf{Z}^k) &= \eta_{\mathcal{H}} \cdot p(\mathbf{Z}_k, n_k, \mathcal{H}_{i,j}, \mathcal{D}_i | \mathbf{Z}^{k-1}) \\ &= \eta_{\mathcal{H}} \cdot p(\mathbf{Z}_k | n_k, \mathcal{H}_{i,j}, \mathcal{D}_i, \mathbf{Z}^{k-1}) \\ &\quad \times p(n_k | \mathcal{H}_{i,j}, \mathcal{D}_i, \mathbf{Z}^{k-1}) \\ &\quad \times p(\mathcal{H}_{i,j} | \mathcal{D}_i, \mathbf{Z}^{k-1}) \cdot p(\mathcal{D}_i | \mathbf{Z}^{k-1}). \end{aligned} \quad (4)$$

With this definition, in order to compute the individual probabilities for each joint detection event, the tracks can be separated in two sets:

- 1)  $\mathcal{T}_{\text{mis}}(\mathcal{D}_i)$ : Set of tracks with no allocated measurements
- 2)  $\mathcal{T}_{\text{hit}}(\mathcal{D}_i)$ : Set of tracks with at least one allocated measurement

Let further  $\mathcal{A}_{\text{mis}}^t$  denote an event, where no measurement is assigned to track  $t$  and  $\mathcal{A}_{\text{hit}}^t$  for at least one assigned detection.

For each track  $t$  in  $\mathcal{T}_{\text{mis}}(\mathcal{D}_i)$  the prior probability is

$$p(\mathcal{A}_{\text{mis}}^t | \mathbf{Z}^{k-1}) = 1 - P_D^t P_G^t p(\chi_k^t | \mathbf{Z}^{k-1}),$$

and if one or more measurements are assigned

$$p(\mathcal{A}_{\text{hit}}^t | \mathbf{Z}^{k-1}) = P_D^t P_G^t p(\chi_k^t | \mathbf{Z}^{k-1}).$$

With these two definitions, the prior probability of a joint detection event is [22, p. 161]

$$\begin{aligned} p(\mathcal{D}_i | \mathbf{Z}^{k-1}) &= \prod_{t \in \mathcal{T}_{\text{mis}}(\mathcal{D}_i)} (1 - P_D^t P_G^t p(\chi_k^t | \mathbf{Z}^{k-1})) \\ &\quad \times \prod_{t \in \mathcal{T}_{\text{hit}}(\mathcal{D}_i)} (P_D^t P_G^t p(\chi_k^t | \mathbf{Z}^{k-1})). \end{aligned} \quad (5)$$

In the next step, consider that  $\mathcal{D}_i$  assigns  $m_1$  measurements to track one,  $m_2$  to track two, and  $m_{n_T}$  up to track  $n_T$ . The total number of combinations for joint association events  $\mathcal{H}_{i,j}$  in  $\mathcal{D}_i$  is given by the multinomial coefficient. Since all events are a priori equally likely, the a priori probability that event  $\mathcal{H}_{i,j}$  is true is given by the inverse of the multinomial:

$$P(\mathcal{H}_{i,j} | \mathcal{D}_i, \mathbf{Z}^{k-1}) = \binom{n_k}{n_k - m_T, m_1, \dots, m_{n_T}}^{-1} = n_M^{-1}, \quad (6)$$

where  $m_T = m_1 + \dots + m_{n_T}$  is the total number of assigned measurements in the detection event  $\mathcal{D}_i$ .

In the next step, the probability of receiving  $n_k$  measurements is defined using the cardinality models  $p_c^{Cl}(n)$  for clutter, and  $p_c^t(n)$  for the target:

$$p(n_k | \mathcal{H}_{i,j}, \mathcal{D}_i, \mathbf{Z}^{k-1}) = p_c^{Cl}(n_k - m_T) \prod_{t \in \mathcal{T}_{\text{hit}}(\mathcal{D}_i)} p_c^t(m_t) \quad (7)$$

In the last step, the probability for the newly received measurement set  $\mathbf{Z}_k$ , given all the quantities above, is

defined as

$$\begin{aligned} p(\mathbf{Z}_k | n_k, \mathcal{H}_{i,j}, \mathcal{D}_i, \mathbf{Z}^{k-1}) \\ = p_s^{Cl}(n_k - m_T) \prod_{t \in \mathcal{T}_{\text{hit}}(\mathcal{D}_i)} \prod_{l \in \mathcal{M}} p'_s(\mathbf{z}_k^l | \mathbf{x}_k^l). \end{aligned} \quad (8)$$

In this equation,  $p_s^{Cl}(n)$  is the spatial distribution of clutter and  $p(\mathbf{z}_k^l | \mathbf{x}_k^l)$  is the spatial measurement model. The set  $\mathcal{M}$  with cardinality  $m_t$  comprises of the indices of the measurements assigned to track  $t$  in hypothesis  $\mathcal{H}_{i,j}$ .

Inserting (5), (6), (7) and (8) into (4) finally yields:

$$\begin{aligned} P(\mathcal{H}_{i,j} | \mathbf{Z}^k) \\ = \eta_{\mathcal{H}} \cdot p_s^{Cl}(n_k - m_T) p_c^{Cl}(n_k - m_T) \cdot n_M^{-1} \\ \times \prod_{t \in \mathcal{T}_{\text{mis}}(\mathcal{D})} (1 - P_G^t P_D^t p(\chi_k^t | \mathbf{Z}^{k-1})) \\ \times \prod_{t \in \mathcal{T}_{\text{hit}}(\mathcal{D})} P_G^t P_D^t p_c^t(m_t) p(\chi_k^t | \mathbf{Z}^{k-1}) \prod_{l \in \mathcal{M}} p'_s(\mathbf{z}_k^l | \mathbf{x}_k^l) \end{aligned} \quad (9)$$

Since all feasible association hypotheses are mutually exclusive and form an exhaustive set, the normalization constant  $\eta_{\mathcal{H}}$  can be derived by demanding

$$\sum_{\mathcal{H}} P(\mathcal{H}_{i,j} | \mathbf{Z}^k) = 1.$$

### B. Track-Based Association Probability

From this point, the MD-JIPDA is derived exactly the same way as the JIPDA in [17]: The hypotheses set  $\mathcal{H}$  now, in general, contains several hypotheses that assign the same measurement combination for the  $t$ th track. Let  $\mathcal{A}_i^{m_i}$  denote the  $i$ th combination hypothesis of assigning  $m$  measurements to a track  $t$  with  $i = [1, \dots, \binom{m_k}{m_i}]$ . For example from Figure 1, the first two combinations are  $\mathcal{A}_1^2 = \{1, 4\}$ , which assigns detections  $M1, M4$  and  $\mathcal{A}_2^2 = \{4, 7\}$  with detections  $M4, M7$ .

Let  $\tilde{\mathcal{H}} \in \mathcal{H}$  denote the set of hypotheses with a specific combination  $\mathcal{A}_i^{m_i}$  assigned to track  $t$ . For each combination in  $\mathcal{A}_i^{m_i} \setminus \{\mathcal{A}_1^0\}$ , the probability that it was generated by  $t$  and the object exists, is then given by

$$p(\chi_k^t, \mathcal{A}_i^{m_i} | \mathbf{Z}^k) = \sum_{\tilde{\mathcal{H}} \in \mathcal{H}} P(\tilde{\mathcal{H}} | \mathbf{Z}^k). \quad (10)$$

The set of hypotheses, where no detection is assigned to a track is denoted  $\mathcal{H}^0$ . Then, in case of a missed detection, following the probability that the object exists is given by

$$p(\chi_k^t, \mathcal{A}_1^0 | \mathbf{Z}^k) = \frac{(1 - P_D^t P_G^t) p(\chi_k^t | \mathbf{Z}^{k-1})}{(1 - P_D^t P_G^t p(\chi_k^t | \mathbf{Z}^{k-1}))} \sum_{\tilde{\mathcal{H}}^0 \in \mathcal{H}} P(\tilde{\mathcal{H}}^0 | \mathbf{Z}^k). \quad (11)$$

The final object existence is given by summing up over all possible measurement combinations  $\mathcal{A}_i^{m_i} \in \mathcal{A}$ :

$$p(\chi_k^t | \mathbf{Z}^k) = \sum_{\mathcal{A}} p(\chi_k^t, \mathcal{A}_i^{m_i} | \mathbf{Z}^k) \quad (12)$$

The association likelihoods are then given by

$$p(\mathcal{A}_i^{m_i} | \mathbf{Z}^k) = p(\mathcal{A}_i^{m_i} | \chi_k^t, \mathbf{Z}^k) = \frac{p(\chi_k^t, \mathcal{A}_i^{m_i} | \mathbf{Z}^k)}{p(\chi_k^t | \mathbf{Z}^k)}, \quad (13)$$

since assigning a measurement to a track requires the underlying assumption that the track also exists. Since (10)–(13) are basically the same equations as for a point target, the reader is referred to [22, p. 162ff.] for a more detailed derivation.

With the association likelihoods, the new posterior state estimate is computed by

$$p(\mathbf{x}_k^t | \mathbf{Z}^k) = \sum_{\mathcal{A}} p(\mathbf{x}_k^t | \mathbf{Z}^k, \mathcal{A}_i^{m_i}) p(\mathcal{A}_i^{m_i} | \mathbf{Z}^k), \quad (14)$$

where  $p(\mathbf{x}_k^t | \mathbf{Z}^k, \mathcal{A}_i^{m_i})$  is the computed measurement update for a specific combination e.g. obtained via standard Kalman filtering.

### III. IMPLEMENTATION USING RANDOM MATRICES

For the derivation of the MD-JIPDA given above, very few assumptions regarding the sensor model have been made. Since an ET-MOT has to concurrently estimate the object's kinematic state and extension, a large variety for the spatial model  $p'_s(\mathbf{z}_k^{(j)} | \mathbf{x}_k^t)$  and measurement cardinality model  $p'_c(m_i)$  are possible. As one possible approach, in this section, the implementation of MD-JIPDA using the Random Matrix framework is briefly described.

It is assumed that each object is of elliptical shape and its extension is described by a symmetric positive definite random matrix  $\mathbf{X}_k$ . The extension is considered to be statistically independent of the kinematic state and of the cardinality model as well. It shall only influence the spatial model which is then rewritten as  $p'_s(\mathbf{z}_k^{(j)} | \mathbf{x}_k^t, \mathbf{X}_k)$ .

For the measurement cardinality model  $p'_c(m_i)$ , using a Poisson distribution is the most common way. However, the expected number of detections per scan can be different for each object. Thus, this parameter, denoted as measurement rate  $\gamma_k$ , has to be estimated in parallel as well.

With these two new quantities, the posterior state estimate is now given by

$$p(\mathbf{x}_k^t, \mathbf{X}_k^t, \gamma_k^t | \mathbf{Z}^k) = \sum_{\mathcal{A}} p(\mathbf{x}_k^t, \mathbf{X}_k^t, \gamma_k^t | \mathbf{Z}^k, \mathcal{A}_i^{m_i}) p(\mathcal{A}_i^{m_i} | \mathbf{Z}^k), \quad (15)$$

The individual steps for solving this equation are explained in the following: First some details on the random matrix framework are given and some considerations on the measurement cardinality model are exemplified. Since for real-time applications, evaluating all possible combinations  $\mathcal{A}_i^{m_i}$  may require too much time, in this section, a proposal is made to handle the exponential increase of the hypotheses trees. Finally, the algorithm applied for track birth and deletion is presented as well.

## A. Random Matrices

The seminal work for the Random Matrices framework for estimation of extended objects was presented in [4]. Using the Bayes' filter, a concept to estimate kinematic target state and its physical extension in parallel was established. Therefore, the following assumptions on the target characteristics are made: First, it is assumed that the shape of the target can be represented by an ellipse. Further, the direction of the object's motion shall be independent of the orientation of the ellipse.

Finally, as the most important assumption, the noise of the measurements is mainly caused by the physical extension. For a measurement  $\mathbf{z}_k^j$  at time  $k$  it is assumed that it can be described by a linear function of the state  $\mathbf{x}_k$ , superimposed by a normally distributed noise term  $\mathbf{w}_k$ :

$$\mathbf{z}_k^j = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k \quad (16)$$

Assuming that the noise part of the measurement is mainly due to the size of the object, the probability density function for a set of measurements  $\mathbf{Z}_k = \{\mathbf{z}_k^1, \dots, \mathbf{z}_k^{n_k}\}$  is defined as:

$$p(\mathbf{Z}_k | n_k, \mathbf{x}_k, \mathbf{X}_k) = \prod_{j=1}^{n_k} \mathcal{N}(\mathbf{z}_k^j; \mathbf{H}\mathbf{x}_k, \mathbf{X}_k) \quad (17)$$

Substituting this relationship in the Bayes' filter recursion leads to an analytic solution for state expectation and covariance update as well as for the update of  $\mathbf{X}_k$ .

However, for many real-world sensors, the extension driven noise is superimposed by some non-negligible sensor driven measurement noise. For example, radar detections are generally in polar coordinates with range  $r$  and detection angle  $\phi$ . Thus, if targets are detected in greater distance, this leads to a larger spread of the measurements in azimuth. Disregarding this fact for the estimation of the physical extension would lead to an overestimation of the true size when the object is far away.

To include the contribution of the sensor error to the measurement spread, [5] proposed the probability density function in the following way:

$$p(\mathbf{Z}_k | n_k, \mathbf{x}_k, \mathbf{X}_k) = \prod_{j=1}^{n_k} \mathcal{N}(\mathbf{z}_k^j; \mathbf{H}\mathbf{x}_k, c\mathbf{X}_k + \mathbf{R}_k) \quad (18)$$

However, for this model, no exact analytical solution for  $p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^k)$  can be found. To obtain a recursive update scheme, in [5] the assumption is made that the target extent is predicted with sufficient accuracy, which makes it possible to separate kinematic and extension updates.

As already mentioned in the introduction, a more general update scheme using the sensor model as in (18) was presented in [6]. According to [6], this approach has a significant better extension estimate, but at the price of a small decrease in position accuracy, and only if the kinematic state uncertainty is sufficiently

small. Some analysis by the authors indicated that in combination with a GPDA, it is also quite vulnerable to false associations. The approach by [5] seems to be more robust, so only this approach is further considered here.

The integration of both update schemes into the GPDA is straight forward: For computation of  $p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^k, A_t^m)$  the update schemes can be implemented exactly as given in the cited papers. For the track prediction, the method proposed in [5] is used. For the sake of completeness, the prediction and update equations are given in the appendix.

The final tracker is designed for radar sensors, so for filtering, the detections have to be transformed from polar to Cartesian space using

$$\mathbf{z}_k = \begin{bmatrix} r_k \cos(\phi_k) \\ r_k \sin(\phi_k) \end{bmatrix}. \quad (19)$$

Since the polar measurement standard deviation  $\sigma_r$  for range and  $\sigma_\phi$  for the detection angle will be small, the associated covariance matrix in Cartesian coordinates is approximated using [24]

$$\mathbf{R}_k \approx \frac{1}{2}(\sigma_r^2 - r_k^2 \sigma_\phi^2) \begin{bmatrix} b + \cos(2\phi_k) & \sin(2\phi_k) \\ \sin(2\phi_k) & b - \cos(2\phi_k) \end{bmatrix} \\ b = \frac{\sigma_r^2 + r_k^2 \sigma_\phi^2}{\sigma_r^2 - r_k^2 \sigma_\phi^2}. \quad (20)$$

This makes the measurement noise state dependent, which may have serious impact in the resulting extension estimate.

For the computation of association likelihoods in (9) with the Random Matrix framework, the spatial model is modified to  $p_s^t(\mathbf{z}_k^{(j)} | \mathbf{x}_k^t, \mathbf{X}_k^t)$ . With respect to (18), it is defined to be a normal distribution, with expectation  $\mathbf{H}\mathbf{x}_{k|k-1}^t$  and the covariance matrix given by

$$\Sigma_k = \mathbf{H}\mathbf{P}_{k|k-1}^t \mathbf{H}^T + c\mathbf{X}_{k|k-1}^t + \mathbf{R}_k.$$

## B. Cardinality Model

In general, a target is considered to give birth to a random number of detections in each scan. This number is in general Poisson distributed with nearly constant mean  $\gamma$ . In a multi-object scenario, the individual targets may also have different values for  $\gamma$ , which are not known a priori.

As long as the number of detections remains a small single digit value (typ.  $\gamma < 5$ ), it can be sufficient to use an average number over all targets for  $\gamma$ : The Poisson distribution is rather indifferent for those values e.g. for expecting  $\gamma = 3$  the measurement probabilities  $p_c^t(m)$  for  $m = 1 \dots 6$  are between [0.05, 0.22]; if expecting 4, the values are in the same interval. In practice, these little differences have no large impact on the association likelihoods, compared to the spatial models. For example, if  $m$  changes, the clutter model with its factor  $V^{n_k - m}$  has, in general, an impact factor in the

region to the power of ten, while the cardinality model only changes by a factor of three at most.

However, in the case of higher detection counts, the values of the Poisson distribution differ by orders of magnitude as well. If the expected values are variable, it is necessary to estimate them together with kinematic state and extension. In [13], it was proposed to model the distribution of  $\gamma$  by a Gamma Distribution and estimate its parameters in parallel by assuming that it is actually independent of the kinematic and extension densities. This concept can be applied to the MD-JIPDA as well.

The Gamma p.d.f. is defined with the two parameters  $\alpha_k > 0$  and  $\beta_k > 0$  as

$$P_{GAM}(\gamma; \alpha_k, \beta_k) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} \gamma^{\alpha_k-1} e^{-\beta_k \gamma}.$$

The expected cardinality is obtained from the expectation value  $\gamma_k = \alpha_k / \beta_k$ . The cardinality model for the MD-JIPDA is then obtained from the joint likelihood of the Poisson distribution with parameter  $\gamma$  and the gamma distribution using

$$\begin{aligned} p'_c(m) &= \tilde{\eta}_c \cdot \int P_{POI}(m | \gamma) P_{GAM}(\gamma | \mathbf{Z}^{k-1}) d\gamma \quad (21) \\ &= \eta_c \cdot \frac{1}{m!} \sum_{\gamma=1}^{n_{\max}} \gamma^{m+\alpha'_{k|k-1}-1} e^{-\gamma(\beta'_{k|k-1}+1)}. \end{aligned}$$

The normalization constant  $\eta_c$  has to account for the fact that the constraint  $\sum_{m=1}^{n_{\max}} p'_c(m) = 1$  still must be fulfilled. The computation of the predicted parameters  $\alpha'_{k|k-1}, \beta'_{k|k-1}$  of the gamma distribution and also its update equations are given in the appendix as well.

### C. Hypotheses Generation

It is easy to see that the MD-JIPDA suffers even more from the exponential increase of possible association hypotheses than the standard JIPDA does. To make the MD-JIPDA computationally feasible for complex scenarios with several tracks within one association gate, or if each track can evoke a large number of measurements, an approximation scheme has to be found in such a way that not all theoretically possible combinations have to be evaluated. This reduction problem is a known issue for an ET-MOT. To solve this problem, the RFS approaches, cited in the introduction, use a combination of distance and Expectation Maximization partitioning to cluster nearby measurements and remove unlikely combinations.

For simplification of the MD-JIPDA, a similar concept is proposed in this paper: First, a k-means clustering is applied to the gated measurements for each track individually. The number of clusters should still be larger than the number of tracks in the specific gate. When the number of clusters is chosen too small, the effect that measurements from different objects are put in the same cluster is very likely to happen, as was pointed

out in detail in [12]. This would also somehow foil the idea of a PDA since the association tree would be very small. The authors have made good experience if the k-means creates at least as many clusters as three times the number of tracks. This ensures that only very few clusters contain measurements of several objects since each object is sufficiently often split. Of course, dependent on the clutter rate or distribution of the measurements on target, a higher cluster number may be required.

If  $j$  clusters have been created, then  $C_{\mathcal{A}} \leq 2^j$  single object association combinations are possible. Given  $n_T$  tracks in the joint association gate, the total number of joint multi-object hypotheses is  $C_{\mathcal{H}} < (C_{\mathcal{A}})^{n_T}$ . Even in the case of five clusters per track and three tracks in a common gate, this can lead to several hundred thousand joint hypotheses, which might be beyond of a real-time implementation. Thus, as a second step, it is considered that by the user, a maximum value for  $C_{\mathcal{H}}$  is given, from which with the relation above, a maximum value for  $C_{\mathcal{A}}$  is derived. Since the number of single object combinations created after the k-means cluster can be significantly higher, from these combinations the  $C_{\mathcal{A}}$  best are selected according to the cardinality model (21). Sampling proportional to the cardinality model  $p'_c(m)$  has the advantage that combinations with highly probable detection count  $m$  are preferred. For example, if  $\gamma = 5$  detections were expected and  $C_{\mathcal{A}} = 20$ , assuming a pure Poisson distribution, 4 combinations that assign  $m = 5$  detections are selected, but only one for  $m = 1$ . For each count  $m$ , the best combinations from the spatial model are chosen. If for a specific count  $m$  more combinations are desired than actually available, the next available count is selected. Only for these selected single association combinations, the joint association hypotheses are built. Please note that when using a reduced number of joint hypotheses, the multinomial coefficient in (9) has to be replaced by the actually generated combination count of each joint detection association event.

### D. Birth Model

The basis of a new track is formed by those measurements that have not been assigned to a track during the data association process. Based on these detections, first a DBSCAN algorithm is executed to cluster closely spaced measurements. For each cluster, a new track is initialized. The initial existence likelihood is computed based on the number of measurements in a cluster and its distance to existing tracks. Consider a measurement cluster with mean  $\bar{\mathbf{z}}$  that contains  $n_z$  measurements, then the initial probability is

$$p(\chi_0) = \beta_{\text{birth}} \cdot p_{\text{birth}}^{\text{card}}(n_z) \cdot p_{\text{birth}}^{\text{sp}}(\bar{\mathbf{z}}). \quad (22)$$

The constant  $\beta_{\text{birth}}$  denotes the general likelihood, typically selected as the average number of new born objects per scan. The likelihood  $p_{\text{birth}}^{\text{card}}(\cdot)$  accounts for the probability that a specific number of detections are part

of a new track. Here, it is proposed to use a typical measurement cardinality model  $p_c^{tref}$  as basis and set

$$p_{\text{birth}}^{\text{card}}(n_z) = \sum_{i=1}^{n_z} p_c^{tref}(i). \quad (23)$$

This model ensures that if the number of detections in the cluster increases, also the probability of a new track is increased.

The spatial model accounts for the distance between the centroid and already existing tracks:

$$p_{\text{birth}}^{\text{sp}}(\bar{\mathbf{z}}) = 1 - \exp\left(-\min_{t \in \{1, \dots, n_T\}} \{\|\bar{\mathbf{z}} - \mathbf{z}_{k|k}^t\|^2\} / \sigma_d^2\right) \quad (24)$$

The distance  $\|\bar{\mathbf{z}} - \mathbf{z}_{k|k}^t\|$  denotes the Euclidean distance between the centroid and the expected measurement of track  $t$ . From all available tracks  $n_T$ , the minimum distance is selected to evaluate  $p_{\text{birth}}^{\text{sp}}$ . With the variance  $\sigma_d^2$ , the desired distance from the centroid of an existing track to a new track candidate can be specified.

Each new track is initialized at the center of the selected cluster using the cluster spread as initial extension estimate. If a cluster contains only very few detections, some minimal size should be used to ensure numerical stability. The track is deleted when at any time the existence probability falls below some small threshold.

#### IV. SIMULATION RESULTS

The performance of multiple detection JIPDA using the Random matrix update schemes given by [5] are evaluated in two sets of Monte Carlo simulations first. One set of simulation is designed with respect to the desired real data application: A high-resolution automotive radar that is mounted on an unmanned surface vessel or a smaller recreational craft. The targets to be tracked are vessels with an overall length below 10 m. For clarity, during maneuvers, in contrast to the true behavior of a vessel, it is assumed that the major axis of the extension ellipse is always aligned with the direction of motion of the object. The second set of simulations is designed to evaluate the capabilities of the MD-JIPDA when a higher number of detections per object is created, and a complete evaluation of the association tree is infeasible.

In all simulations, the coordinated turn model with  $\mathbf{x}_k = [x, v_x, y, v_y, \omega]^T$  is used as motion model, with Cartesian positions  $x, y$ , the corresponding velocities  $v_x, v_y$  and the turn rate  $\omega$  around the vertical axis.

For the evaluation of each scenario, the modified version of the optimal sub-pattern assignment (OSPA) metric, as introduced in [13], is applied. This modification enhances the OSPA to incorporate also the estimated target size  $\mathbf{X}_{k|k}$  and measurement cardinality  $\gamma_{k|k}$  of an object.

##### A. Low Detection Count Scenarios

To evaluate the joint data association, one scenario with four vessels and one scenario with two vessels

are considered. For all scenarios, the measurements are assumed to be uniformly distributed over the vessel's extension, and the cardinality is Poisson distributed with constant mean. The sensor reports its measurements in polar coordinates, where the accuracy of a point target is  $\sigma_R = 1.0$  m in range and  $\sigma_\phi = 0.1^\circ$  for the bearing angle. The observation area is set to  $200 \text{ m} \times 200 \text{ m}$ , with the sensor located in the center and its sample time  $T = 1/15$  s. Two different clutter rates are considered: A lower case with a mean of 8 false alarms per scan, ( $\lambda_8 = 2 \cdot 10^{-4}/\text{m}^2$ ) and a high clutter case with 80 false positives ( $\lambda_{80} = 2 \cdot 10^{-3}/\text{m}^2$ ). As a further challenge, a different probability of detection is considered in both cases: for the medium case, it is set to 95% and for the high clutter case it is reduced to  $p_D = 80\%$ .

For tracking and filtering the following parameters are used: The motion model process noise is set to  $\sigma_v = 0.1 \text{ m/s}^2$  and  $\sigma_\omega = 1.0^\circ/\text{s}$ . The sensor noise to extension noise ratio in (18) is set to  $c = 1/4$ , and the time constant for the capability of changes in the extension is set to 5 s. The decay constant for track existence is set to  $\tau_E = 10$  s, and the forgetting factor for target measurement rate parameters  $\alpha_{k|k}, \beta_{k|k}$  is set to  $\kappa_\gamma = 1.25$  (see prediction step in the appendix).

For the unassigned measurements, the DBSCAN clustering is performed with a distance threshold of 5 m. A new track is created at the cluster's centroid position with an initial existence likelihood of  $\beta_{\text{birth}} = 0.01$ , a track distance  $\sigma_d = 20$  m and a Poisson distribution with a mean of three for  $p_c^{tref}$ . The initial position uncertainty is set to  $P_0 = \text{diag}\{(3 \text{ m})^2, (1/2 \text{ m})^2, (3 \text{ m})^2, (1 \text{ m/s})^2, (1^\circ)^2\}$ , and the initial extension is a circle with radius 2 m. If the existence probability falls below the level of  $10^{-5}$ , the track is considered dead, and if the probability exceeds 50%, the track is treated as valid.

The modified OSPA metric uses the cut-off values  $c_x = 3$  m,  $c_X = 30 \text{ m}^2$  and  $c_\gamma = 2$  with weights  $w_x = 0.8$ ,  $w_X = 0.1$  and  $w_\gamma = 0.1$ . The norm  $p$  is just set to one.

1) Scenario A.1: The first scenario is a typical multi-target scenario to test the general capabilities of the proposed MOT: A total of four different objects are on a straight line trajectory with a nearly constant velocity of about 3 m/s, see Figure 2 and Table I for details. All objects meet at the same area but keep the distance to their centroids mutually of at least four meters. The scenario has a total of 600 samples (40 s). Each object is created and deleted at a different time step. In this scenario, only the suboptimal MD-JIPDA version is applied with limiting the maximum number of joint hypotheses to  $1e3$ .

The average results for 1000 Monte Carlo Simulation are shown in Figure 3. The tracking cardinality is computed by taking the sum of existence probability over all created tracks at time  $k$ :  $\sum_t p(\chi_k^t)$ . Overall, for both clutter rates and both cardinality models, an acceptable and almost identical performance was achieved. As could be expected, with higher clutter rates and smaller

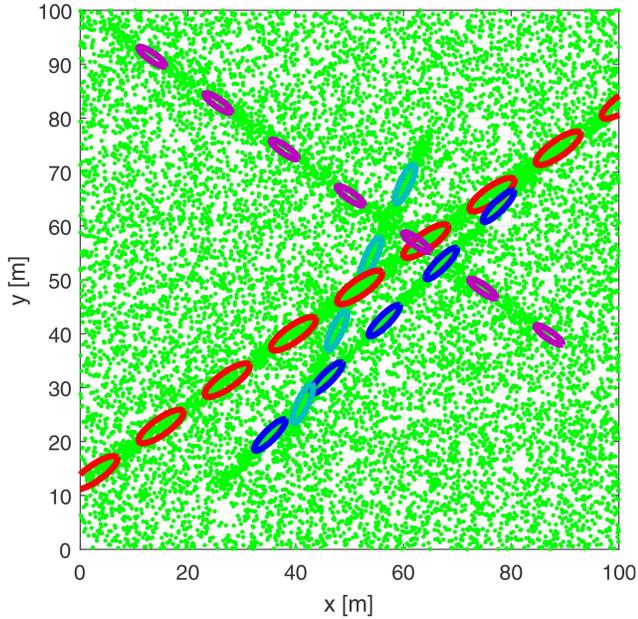


Fig. 2. Targets trajectories and received measurements. Ellipses plotted every five seconds.

TABLE I

Object characteristics for Scenario A. Colors corresponding to Figure 2. The variables  $A$  and  $a$  denote the length of semi-major and semi-minor axis, respectively.

	Blue	Red	Magenta	Gray
$A$	4 m	5 m	4 m	3 m
$a$	1 m	1.5 m	1 m	0.75 m
$\gamma$	3	4	3	2
$t_{\text{birth}}$	8 s	0 s	2 s	6 s
$t_{\text{death}}$	34 s	40 s	36 s	28 s

detection probability, the confirmation of a new track takes longer but is still handled fairly well for all objects.

Eye-catching is, of course, the overestimation of objects in case of low clutter with adaptive measurement cardinality. When clutter measurements occur close to a new born target, the measurement data is ambiguous: It could stem from two smaller objects or one large and clutter. However, clutter measurements are in total rare, so the algorithm prefers to keep both tracks alive for a little longer time. Thus, depending on the current distribution of measurements, two tracks moving behind each other are computed as the most likely event. This effect becomes less dominating when the clutter rate increases, since the general track confirmation takes significantly longer. From the OSPA, it is seen that it still actually performs slightly worse than the fixed cardinality model. This is just due to the same effect: Instead of one larger object, two small objects are built. While the fixed model prefers larger objects, and after an initial phase only one object survives, the adaptive cardinality permits two objects, each with an expected cardinality of one or two detections only.

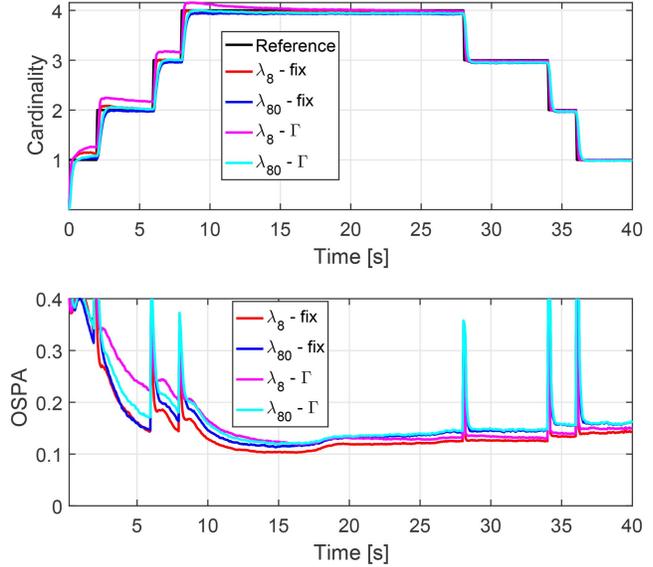


Fig. 3. Track cardinality and OSPA error for different clutter rates. The suffix ‘ $\Gamma$ ’ indicates MD-JIPDA with estimated measurement cardinality, and fix with the constant model.

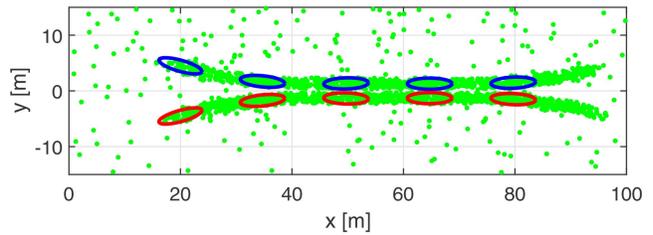


Fig. 4. Trajectories and measurement data for Scenario B. Ellipses plotted every five seconds. The space between both targets is about 22 cm.

2) Scenario B.1: In the second scenario, two objects move towards each other, proceed in parallel for approx. 15 s and then split up again (Figure 4). Both objects have the same characteristics as object ‘Blue’ from the scenario above. To keep the total number of hypotheses feasible for the MD-JIPDA, a limit of  $n_{\text{max}} = 6$  detections per object is set. With this, the scenario enables a comparison between the optimal and its suboptimal version.

The results for 1000 Monte Carlo runs are shown in Figures 5 for the low clutter case, and in 6 for the high clutter case. In these, the additional prefix ‘Opt’ denotes the results using full MD-JIPDA, ‘Sub’ denotes the hypothesis limited approach.

Like in the first scenario, an overshoot in the cardinality can be seen when using an adaptive measurement rate  $\Gamma$ . This is again due to the fact that in some runs, two tracks are initialized for one object. Besides this, in all cases, accurate tracks were created for both objects. The fact that the cardinality remains below the true cardinality is due to the limited detection probability  $P_D$ . In both OSPA also an increase of the OSPA at about half time can be observed. This is due to another well-known problem of the JIPDA: The track coalescence. Since

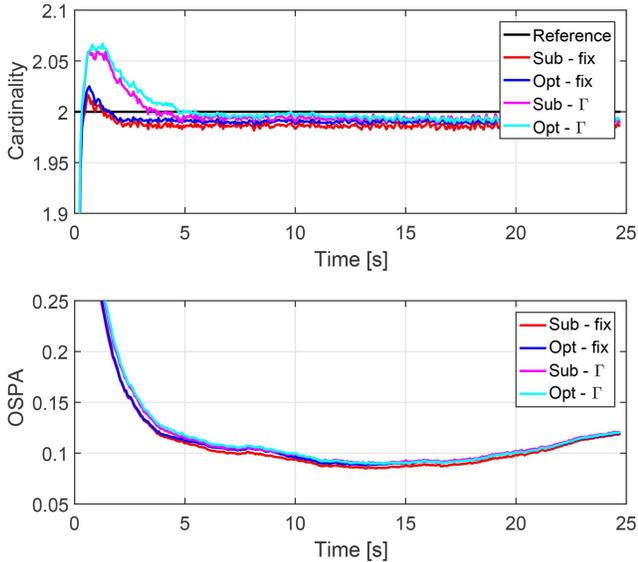


Fig. 5. Track cardinality and OSPA error for medium clutter rate  $\lambda_8$ .

measurements from the other object are also taken into account, the tracks move closer together. This leads to an overlapping of the extension ellipses in this scenario but with still well-separated target centroids.

These figures also reveal that there is no significant difference between the optimal and the suboptimal approach. Although some slight edge for the optimal approach in cardinality can be seen, it actually also suffers more from track coalescence. Nevertheless, the differences are very small, except for the computation time. These are shown in Table II.<sup>1</sup> In a few runs, over one hour is required to compute the full MD-JIPDA for this scenario. This excluded the algorithm from being used in real-time<sup>2</sup> applications. In contrast, the suboptimal MD-JIPDA with a limited hypotheses count is always within reasonable computation time. Especially with the low clutter scenario, each update step can be handled easily within the update time  $T$ . It also has only a small spread between the minimum and maximum required time. The high clutter scenario takes actually more time than the scenario duration. This is due to a large number of tracks that are created and deleted in each update step.

### B. High Detection Count

In most cases of the simulations above, the need of clustering or hypotheses limitation for real-time computation is rare. Thus, in this subsection, some of the scenarios above are simulated again using different settings for the measurement model, e.g. an object can create at least 10 detections per step. For this type of sensors, it is even for a single track computationally too expensive to

<sup>1</sup>The values are obtained using single core Matlab simulation on a 3.7 GHz PC.

<sup>2</sup>Here, a tracker is considered real-time applicable when a complete measurement update step can be computed within the sensor's sample time  $T$ .

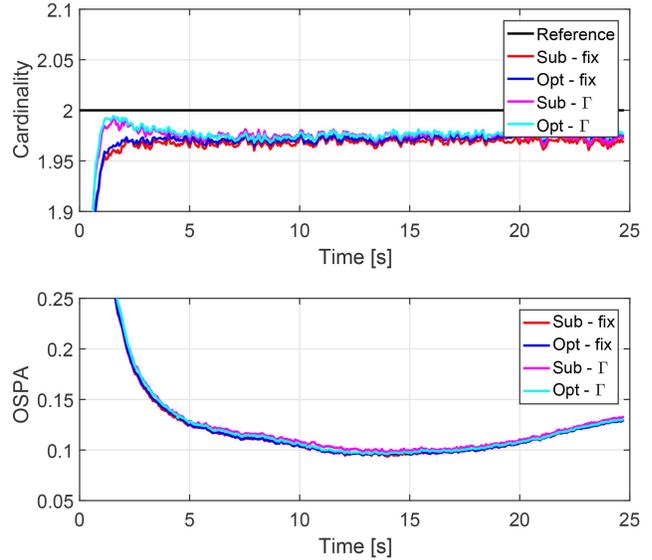


Fig. 6. Track cardinality and OSPA error for high clutter rate  $\lambda_{80}$ .

TABLE II

Computation Time for the complete sequence of Scenario B.1 for the full MD-JIPDA and the hypothesis reduced suboptimal version. The first value is the mean time over 1000 runs, the value in parentheses the maximum occurred time.

	$\lambda_8$ Fix	$\lambda_8$ $\Gamma$	$\lambda_{80}$ Fix	$\lambda_{80}$ $\Gamma$
optimal	42 s (4441 s)	33 s (4331 s)	48 s (3278 s)	50 s (1121 s)
suboptimal	13 s (16 s)	13 s (16 s)	37 s (44 s)	40 s (48 s)

evaluate all possible association hypotheses. However, since e.g. for laser scanner such a measurement count is quite normal, it is important to analyze how the suboptimal approach will perform, and if it can be applied to such problems. Since for this type of sensors the polar sensor noise is low, it is set to zero for these simulations.

1) Scenario A.2: In general, the settings are identical to the previous scenario A.1, except for the target count: Each expected number of measurements in Table I is multiplied with a factor of five, so the expected number is between 10 and 20. In contrast to the scenarios above, with these measurement rates, it is at no time possible to compute a full set of joint association events. The results using the suboptimal with a maximum of  $10^4$  hypotheses are shown in Figure 7. In these figures, again a comparison with a fixed cardinality model with  $\gamma = 15$  and the adaptive scheme are given. As could be expected, only the adaptive scheme handles the scenario correctly, while in the fixed case, for the large red object in many cases, two tracks are established. This effect is again reduced when the clutter rate is higher. For the adaptive scheme, in all runs, the objects are tracked accurately and of course with also better results than in scenario A.1 due to the higher measurement count.

2) Scenario B.2: As a final scenario, the special high clutter scenario as given in [25] is also applied. On the

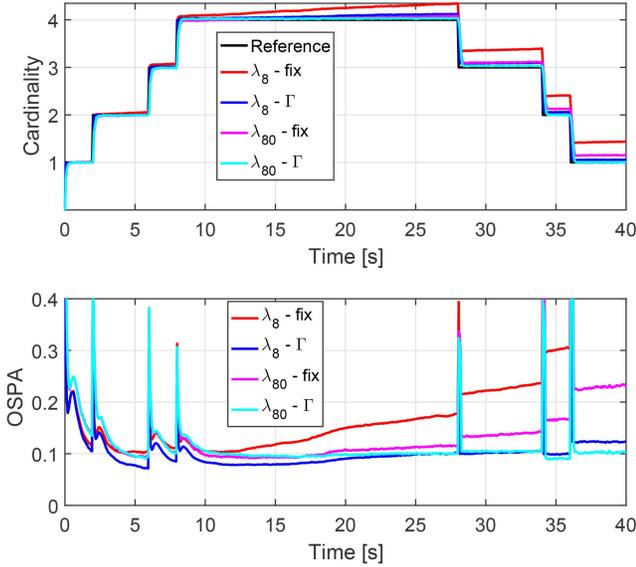


Fig. 7. Track cardinality and OSPA error for different clutter rates for A.2. The suffix ‘ $\Gamma$ ’ indicates MD-JIPDA with estimated measurement cardinality, and ‘fix’ with the constant model.

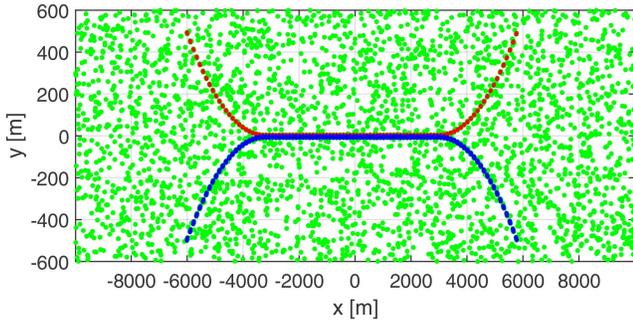


Fig. 8. Target trajectories over the clutter measurement data from *one* scan in the very high clutter case.

first glance, this scenario is similar to the scenario B with a starting turn and a parallel phase (Figure 8). However, the conditions are completely different. Each target is moving with a constant speed of 120 m/s with a distance of less than one meter to the outline of the object during the parallel phase. The first target starts at the south-east, has a size of 40 m  $\times$  20 m and generates about  $\gamma = 20$  measurements per scan. The second target’s dimensions are 20 m  $\times$  10 m with an average measurement count of 10.

The parameters for the suboptimal assignment are set to 5 clusters per track and a total maximum of  $10^4$  hypotheses is chosen.

For this dynamics, the parameters for track initialization and maintenance are therefore changed as follows: The initial position uncertainty is increased to 50 m, and the velocity uncertainty to 80 m/s. The initial extension matrix  $\mathbf{X}_0 = 10^2 \times \mathbf{I}_2$  with DOF  $\nu_0 = 10$ . The cardinality model is initialized with  $\alpha_0 = 10$  and  $\beta_0 = 1$ . The process noise for the velocity is increased to  $\sigma_v = 3 \text{ m/s}^2$ . The remaining track parameters and decay constants are the same as in the scenarios above.

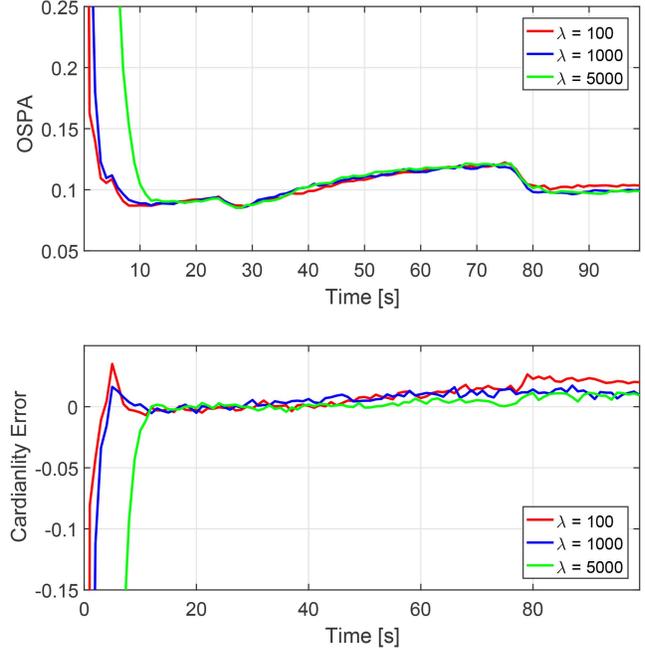


Fig. 9. OSPA and track cardinality error for different clutter rates for scenario B.2.

The scenario is evaluated with three different clutter rates: A low clutter case of 100 false alarms per scan, a high case with 1000 and a very high case with 5000 clutter measurements on a surveillance area of  $2 \cdot 10^3 \text{ m} \times 2 \cdot 10^4 \text{ m}$ . The detection probability is in all cases  $p_D = 98\%$ .

From Figure 9, it can be seen that the proposed algorithm can handle this type of scenarios also very well. As can be seen, the OSPA error is slowly increasing during the parallel target movement. The problem of track coalescence occurs once more. Due to the measurement clustering by the k-means, this effect is stronger here than in scenario B.1. This also affects the estimated size of the targets, which becomes gradually overestimated, however with a rather slow increase. The issue is resolved, as soon the targets split off again.

## V. EXPERIMENTAL RESULTS

For the experimental tests, an automotive radar sensor was mounted on a small vessel. In contrast to the simulation, the radar has an opening angle of only  $\pm 26^\circ$  for a 60 m short range mode and  $\pm 9^\circ$  for larger distances. Due to the limited field of view, the objects under observation are allowed to just perform small maneuvers, as they have to remain inside the field of view (FoV).

The test setup consists of two vessels moving in front of the host vessel (Figure 10). For all three vessels, the GPS traces are recorded. The vessels perform small approaching and bear off maneuvers while slowly



Fig. 10. Target vessels used for data acquisition. The left one (blue) is also used for the single object scenario and has a size of  $8.5 \text{ m} \times 2.5 \text{ m}$ . The dimensions of the right one (red) are  $6.9 \text{ m} \times 2.47 \text{ m}$ .

increasing the distance to the host. The recorded trajectory relative to the host and the received measurements are shown in Figure 11.

The tracking algorithm is executed in the body-fixed coordinates of the host vehicle. This requires the compensation of the motion of the host vehicle for each track before performing the update step. Since this is done using the velocity measurement from the GPS and the yaw rate measurement from a low-cost gyro, additional uncertainty is induced into the estimate. To take

this into account, for the prediction step, the method proposed by [26] is used to rotate the ellipses according to the motion of the host vessel. However, estimates of the target’s yaw rate are not used for extension prediction. The tracking parameters are identical to the values given in the first simulation set for low clutter tracking.

For this scenario, the MD-JIPDA is tested with adaptive measurement rate and a constant rate  $\gamma_k = 4$ . The cardinality estimates and OSPA results for both schemes are shown in Figure 13, and Figure 12 shows the estimate for the variable approach in a local coordinate frame. From the cardinality plot, it can be seen that in the beginning of the sequence, the fixed  $\gamma$  performs slightly better, but at the end, when the red object moves out of the sensor FoV, it performs significantly worse. This is due to the fact that the target reduces to a point target, which, in combination of a low detection rate, leads to several low estimates of the existence probability. This is compensated by the adaptive version. The OSPA metric is identical for both schemes.

With increasing distance, the number of received detections is decreasing, and so the overshoot, as seen in the simulation, does not occur.

## VI. CONCLUSION

This paper presents generalized versions of the JIPDA filter to assign more than one measurement to a

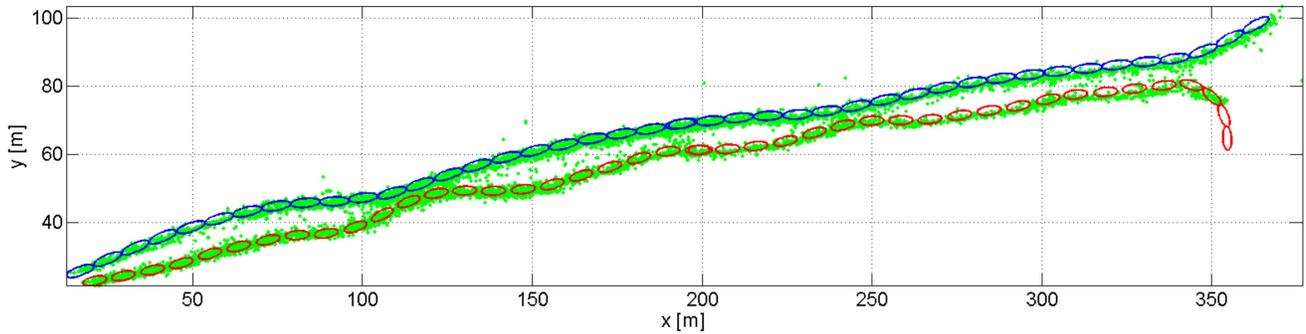


Fig. 11. Reference trajectory of targets (red and blue ellipse) from GPS and received detections (green) in local coordinate frame. The distance to the host vessel during the sequence varies from 20 m to 120 m. The ellipses are plotted in time intervals of 6 s. The scenario starts in the lower left corner, when both ships enter the FoV after overtaking the host vessel. The vessels perform three “draw near and bear away” maneuvers. After about two minutes, the red vessel turns starboard and leaves the FoV (upper right).

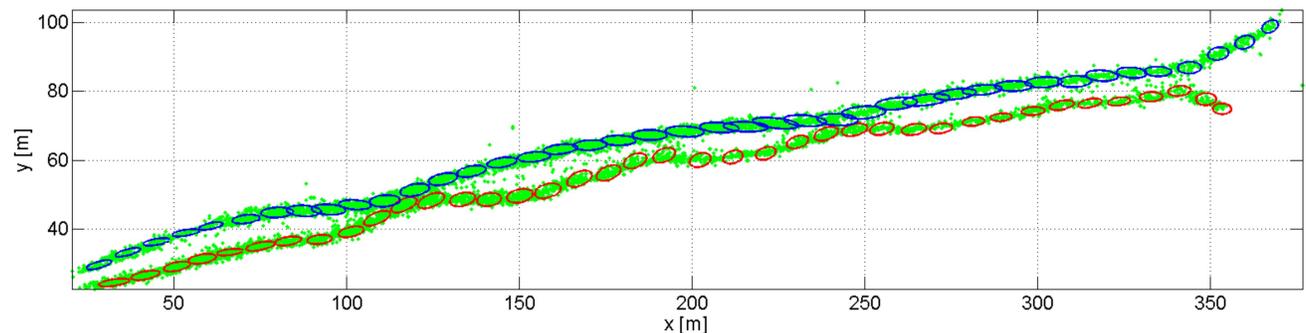


Fig. 12. Estimated trajectory of targets (red and blue ellipse) from MD-JIPDA filter and received detections (green) in local coordinate frame.

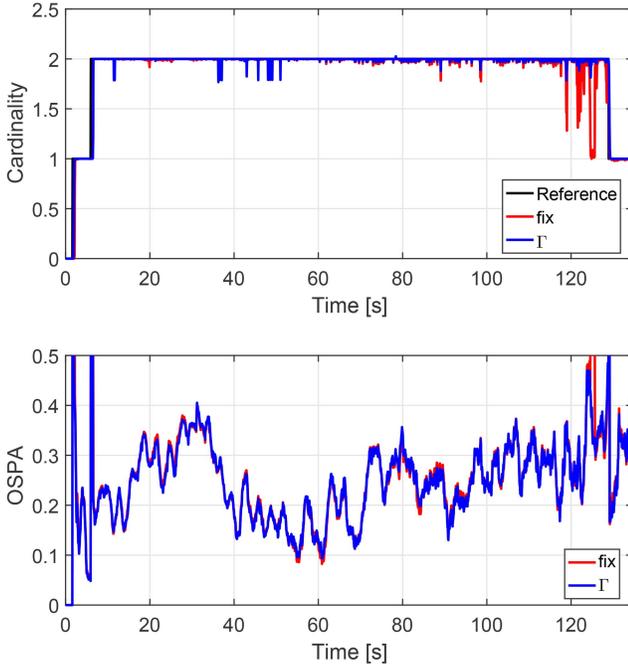


Fig. 13. OSPA metric and number of tracks validated tracks for the experimental data sets.

track. The so-called MD-JIPDA is connected to the random matrix framework to track extended objects with an elliptical shape. To overcome the problem of exponential increase of association hypotheses, a simple cluster and sample technique is applied.

It was shown in simulation and real data scenarios that the proposed MOT algorithm is capable of resolving extended targets, which are moving in close proximity. The results with more complex scenarios indicate that the MD-JIPDA can achieve quite similar results as the RFS approaches. However, like any JPDA, it suffers from track coalescence. Instead, it is the author's opinion, that the MD-JIPDA comes with a reduced complexity e.g. when compared to the recently published LMB. Of course, for a real comparison, the according studies are yet to be made. Another nice feature of the MD-JIPDA is that it can make use of well-known techniques from target tracking, like e.g. gating.

The proposed hypotheses reduction is a rather coarse and intuitive scheme, which offers space for further improvement. An interesting alternative that has to be investigated, is the use of an iterative approach as it was given for the JIPDA in [27]. Starting from the GPDA solution, the association hypothesis tree could be successively expanded up to the desired resolution level. Also, the problem of track coalescence must be addressed by e.g. checking if the techniques from the JPDA can be adopted.

#### APPENDIX A FILTER STEPS FOR TRACK PREDICTION AND UPDATE

The prediction equations for sample time T are given in Table III and the update equations for track

TABLE III  
State filter prediction steps

Kinematic:

$$\mathbf{x}_{k|k-1} = \mathbf{F}_k \mathbf{x}_{k-1|k-1}$$

$$\mathbf{P}_{k|k-1} = \mathbf{P}_{k-1|k-1} + \mathbf{F}_k \mathbf{Q}_k \mathbf{F}_k^T$$

Extension:

$$\mathbf{X}_{k|k-1} = \mathbf{X}_{k-1|k-1}$$

$$\delta_{k|k-1} = \delta_{k-1|k-1} \cdot e^{-T/\tau_\delta}$$

Measurement rate:

$$\alpha_{k|k-1} = \frac{1}{\kappa_\gamma} \alpha_{k-1|k-1}$$

$$\beta_{k|k-1} = \frac{1}{\kappa_\gamma} \beta_{k-1|k-1}$$

TABLE IV  
State filter update steps

Kinematic:

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_{k|k-1} (\bar{\mathbf{z}}_k - \mathbf{H} \mathbf{x}_{k|k-1})$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} + \mathbf{K}_{k|k-1} \mathbf{H} \mathbf{P}_{k|k-1}^T$$

$$\mathbf{S}_{k|k-1} = \mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^T + \frac{1}{n_k} \mathbf{Y}_{k|k-1}$$

$$\mathbf{K}_{k|k-1} = \mathbf{P}_{k|k-1} \mathbf{H}^T \mathbf{S}_{k|k-1}^{-1}$$

$$\mathbf{Y}_{k|k-1} = c \mathbf{X}_{k-1|k-1} + \mathbf{R}_k$$

Extension:

$$\mathbf{X}_{k|k} = \frac{1}{\delta_{k|k}} (\delta_{k|k-1} \mathbf{X}_{k|k-1} + \hat{\mathbf{N}}_{k|k-1} + \hat{\mathbf{Y}}_{k|k-1})$$

$$\delta_{k|k} = \delta_{k|k-1} + n_k$$

$$\hat{\mathbf{N}}_{k|k-1} = (\bar{\mathbf{z}}_k - \mathbf{H} \mathbf{x}_{k|k-1}) (\bar{\mathbf{z}}_k - \mathbf{H} \mathbf{x}_{k|k-1})^T$$

$$\hat{\mathbf{N}}_{k|k-1} = \mathbf{X}_{k|k-1}^{1/2} \mathbf{S}_{k|k-1}^{-1/2} \mathbf{N}_{k|k-1}^{1/2} (\mathbf{S}_{k|k-1}^{-1/2})^T (\mathbf{X}_{k|k-1})^T$$

$$\hat{\mathbf{Y}}_{k|k-1} = \mathbf{X}_{k|k-1}^{1/2} \mathbf{Y}_{k|k-1}^{-1/2} \bar{\mathbf{Z}}_k^{1/2} (\mathbf{Y}_{k|k-1}^{-1/2})^T (\mathbf{X}_{k|k-1})^T$$

Measurement rate:

$$\alpha_{k|k} = \alpha_{k|k-1} + n_k$$

$$\beta_{k|k} = \beta_{k|k-1} + 1$$

state estimate for an assigned measurement set with  $n_k$  detections, centroid  $\bar{\mathbf{z}}_k$  and spread  $\bar{\mathbf{Z}}_k$  in Table IV.

#### REFERENCES

- [1] M. Blaich, S. Köhler, M. Schuster, J. Reuter, and T. Tietz "Mission integrated collision avoidance for USVs using laser ranger," in *OCEANS 2015 MTS/IEEE*, Genova, 2015.
- [2] M. J. Waxman and O. E. Drummond "A bibliography of cluster (group) tracking," *Proc. SPIE*, vol. 5428, pp. 551–560, 2004. [Online]. Available: <http://dx.doi.org/10.1117/12.548357>.

- [3] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, and S. Godsill  
“Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking,”  
*Digital Signal Processing*, vol. 25, no. 0, pp. 1–16, 2014.
- [4] J. Koch  
“Bayesian approach to extended object and cluster tracking using random matrices,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, no. 3, pp. 1042–1059, 2008.
- [5] M. Feldmann, D. Franken, and W. Koch  
“Tracking of extended objects and group targets using random matrices,”  
*Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1409–1420, April 2011.
- [6] U. Orguner  
“A variational measurement update for extended target tracking with random matrices,”  
*Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3827–3834, July 2012.
- [7] J. Lan and X. Li  
“Tracking of extended object or target group using random matrix—Part I: New model and approach,”  
in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 2177–2184.
- [8] ———  
“Tracking of extended object or target group using random matrix—Part II: Irregular object,”  
in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 2185–2192.
- [9] M. Baum and U. Hanebeck  
“Random hypersurface models for extended object tracking,”  
in *Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on*, Dec 2009, pp. 178–183.
- [10] M. Wieneke and W. Koch  
“A PMHT approach for extended objects and object groups,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 3, pp. 2349–2370, JULY 2012.
- [11] M. Wieneke and S. Davey  
“Histogram-PMHT for extended targets and target groups in images,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 50, no. 3, pp. 2199–2217, July 2014.
- [12] K. Granström and U. Orguner  
“A PHD filter for tracking multiple extended targets using random matrices,”  
*Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5657–5671, Nov 2012.
- [13] C. Lundquist, K. Granström, and U. Orguner  
“An extended target CPHD filter and a gamma Gaussian inverse Wishart implementation,”  
*IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 472–483, June 2013.
- [14] M. Beard, S. Reuter, K. Granström, B. T. Vo, B. N. Vo, and A. Scheel  
“Multiple extended target tracking with labeled random finite sets,”  
*IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1638–1653, April 2016.
- [15] Y. Bar-Shalom, T. Kirubarajan, and X. Lin  
“Probabilistic data association techniques for target tracking with applications to sonar, radar and EO sensors,”  
*Aerospace and Electronic Systems Magazine, IEEE*, vol. 20, no. 8, pp. 37–56, 2005.
- [16] D. Musicki, R. Evans, and S. Stankovic  
“Integrated probabilistic data association,”  
*Automatic Control, IEEE Transactions on*, vol. 39, no. 6, pp. 1237–1241, 1994.
- [17] D. Musicki and R. Evans  
“Joint integrated probabilistic data association: JIPDA,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 40, no. 3, pp. 1093–1099, July 2004.
- [18] H. Zhu, T. Ma, S. Chen, and W. Jiang  
“A random matrix based method for tracking multiple extended targets,”  
in *Information Fusion (FUSION), 2014 17th International Conference on*, July 2014, pp. 1–8.
- [19] B. Habtemariam, R. Tharmarasa, T. Kirubarajan, D. Grimmett, and C. Wakayama  
“Multiple detection probabilistic data association filter for multistatic target tracking,”  
in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, July 2011, pp. 1–6.
- [20] B. Habtemariam, R. Tharmarasa, T. Thayaparan, M. Mallick, and T. Kirubarajan  
“A multiple-detection joint probabilistic data association filter,”  
*Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 3, pp. 461–471, June 2013.
- [21] R. Schubert, C. Adam, E. Richter, S. Bauer, H. Lietz, and G. Wanielik  
“Generalized probabilistic data association for vehicle tracking under clutter,”  
in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, 2012, pp. 962–968.
- [22] S. Challa, M. R. Morelande, D. Musicki, and R. J. Evans  
*Fundamentals of Object Tracking*.  
Cambridge University Press, 2011, Cambridge Books Online. [Online]. Available: <http://dx.doi.org/10.1017/CBO9780511975837>.
- [23] R. P. Samuel Blackman  
*Design and Analysis of Modern Tracking Systems*.  
Artech House Boston, 1999.
- [24] D. Lerro and Y. Bar-Shalom  
“Tracking with debiased consistent converted measurements versus EKF,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 29, no. 3, pp. 1015–1022, Jul 1993.
- [25] S. Reuter, M. Beady, K. Granström, and K. Dietmayer  
“Tracking extended targets in high clutter using a GGIW-LMB filter,”  
in *Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015*, Oct 2015, pp. 1–6.
- [26] K. Granström and U. Orguner  
“New prediction for extended targets with random matrices,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1577–1589, 2014.
- [27] T. L. Song, H. W. Kim, and D. Musicki  
“Iterative joint integrated probabilistic data association for multitarget tracking,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 51, no. 1, pp. 642–653, January 2015.



**Michael Schuster** received his master degree in electrical engineering in 2010 from the Konstanz University of Applied Science. He is working as research assistant towards his Ph.D. at Chemnitz University of Technology in cooperation with Konstanz University of Applied Sciences. His research interests include data fusion, target tracking, and advanced driver assistance systems.



**Johannes Reuter** received Dipl. Ing. and Ph.D. degrees in 1995 and 2000, both from Technische Universität Berlin. From 2000 to 2001 he worked as a development engineer for IAV GmbH Berlin and from 2001 to 2004 as a development engineer and department manager for IAV Automotive Engineering Inc. in Ann Arbor MI, USA. From 2004 to 2007 he was with the Innovation Center of EATON Corporation in Southfield MI, USA. Since 2007 he has been a Professor for Automatic Control Systems at University of Applied Sciences Konstanz. His current research interests are related to multi sensor data fusion, autonomous maritime systems, observer-based control of magnetic actuators and renewable energy systems.



**Gerd Wanielik** received his Ph.D. in 1988 from the Technical University of Karlsruhe (Germany) in electrical engineering. From 1979 he has worked as a scientist at the AEG-Telefunken Research Institute and later at the Daimler Chrysler Research Center. Since 1999 he has been Professor of Communication Engineering at the Chemnitz University of Technology (Germany). His research interests include multi sensor processing and systems, polarimetry, communication and navigation systems.

# Converted Measurements Bayesian Extended Target Tracking Applied to X-band Marine Radar Data

GEMINE VIVONE  
PAOLO BRACA  
KARL GRANSTRÖM  
ANTONIO NATALE  
JOCELYN CHANUSSOT

**X-band marine radar systems are flexible and low-cost tools for monitoring multiple targets in a surveillance area. They can provide high resolution measurements both in space and time. Such features offer the opportunity to get accurate information not only about the target kinematics, as other conventional sensors, but also about the target size.**

**In this paper we exploit the random matrix framework to track extended targets. Proper measurement models to deal with the radar's measurement noise and its conversion into Cartesian coordinates are presented here. Benefits of the proposed extended target tracking using converted measurements can be mainly related to the problem of the targets' size estimation, while advantages on estimation of the targets' kinematic features can be considered negligible. The validity of the proposed approach has been demonstrated by using both simulated and real data. Gains up to 70% for the targets' width estimation accuracy and around 65% for the length are observed on real data. The integration of the proposed model into the gamma Gaussian inverse Wishart probability hypothesis density tracker is also provided and tested on real data.**

Manuscript received December 18, 2015; revised July 8, 2016; released for publication July 22, 2016.

Refereeing of this contribution was handled by Dr. Marcus Baum.

Authors' addresses: G. Vivone and P. Braca are with the North Atlantic Treaty Organization (NATO) Science and Technology Organization (STO) Centre for Maritime Research and Experimentation, 19126 La Spezia, Italy (E-mail: {gemine.vivone,paolo.braca}@cmre.nato.int). K. Granström is with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, 412 96, Sweden (E-mail: karl.granstrom@chalmers.se). A. Natale is with IREA-CNR, Napoli, Italy (E-mail: natale.a@irea.cnr.it). J. Chanussot is with the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab), Grenoble Institute of Technology, 38000 Grenoble, France, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 107 Reykjavík, Iceland. (E-mail: jocelyn.chanussot@gipsa-lab.grenoble-inp.fr).

The authors declare no conflict-of-interest.

1557-6418/17/\$17.00 © 2017 JAIF

## I. INTRODUCTION

Securing the waterways is of critical importance, and surveillance activities take on a central role. Ship traffic monitoring and port protection represent big challenges (e.g. in terms of law enforcement, search and rescue, environmental protection, and resource management) and, in the last years, it has stimulated intensive research activities, e.g. [9], [22], [31], [41].

Radars are widely exploited technologies. Among these, X-band marine radar systems represent flexible and low-cost tools for tracking of multiple targets. Features, such as high resolution in both space and time, make these kinds of systems very appealing because, if compared to conventional radars, they are able to provide indications about the targets' size and not only about targets' kinematics. This additional information can be very helpful for subsequent signal processing phases, e.g. target classification.

The tracking literature is mostly focused on approaches that make the hypothesis of at most one detection per target for each frame, see for instance [3], [7], [28], [31], [41], which is no longer valid for the data considered in this paper. We refer to this problem as *extended target tracking* (ETT). Several approaches can be found in the literature to address the ETT problem.

Bar-Shalom, et al. [3] propose to segment the acquired image. Clustering and centroid extraction phases are subsequently used to provide data for feeding the probabilistic data association (PDA) algorithm. A technique for data association using a multi-assignment approach to track a large number of closely spaced (and overlapping) targets is also presented in [24]. In [14], the authors propose an approach for ETT under the assumption that the number of received measurements is a Poisson distribution. The algorithm is illustrated with point targets, which may generate more than one measurement and have a 1-D extension. A sequential Monte Carlo method is also proposed in [13], where sensor measurements are modeled as a Poisson process with a spatially dependent intensity parameter, which leads to the representation of physical extent as an intensity distribution that avoids the evaluation of explicit data association hypotheses. A similar approach is taken in [8] where track-before-detect theory is used to track a point target with a 1-D extent. The application of track-before-detect theory together with particle filters on X-band marine radar data has also been investigated in [11]. An interacting multiple model data augmentation algorithm and a modified version of the mixture Kalman filter are proposed for extended target tracking in [2]. Two models, based on support functions for smooth object shapes and extended Gaussian image in the non-smooth object case, are proposed and used for extended target tracking in [39]. In [34] the problem of group structure inference and joint detection and tracking for group and individual targets within a

Bayesian filtering framework is addressed. Group dynamical models from a continuous time setting, the interaction models for closely spaced targets, and a group structure transition model are proposed. Baum, et al. introduce in [5], [6] the random hypersurface model for estimating both kinematic and shape parameters of extended targets. Specific estimators are derived for elliptic and star-convex shapes. In [30], Mahler proposes an expansion to extended targets of his probability hypothesis density filter [29] to manage the multi-target tracking problem. Unfortunately, the proposed filter requires processing of all possible measurement set partitions, which is generally unfeasible to implement. An approach for limiting the number of considered partitions is proposed and discussed in [20]. A sequential Monte Carlo multi-target Bayes filter based on finite set statistics is exploited for pedestrian tracking in [35]. Furthermore, a nonlinear Bayesian methodology for image sequences incorporating the statistical models for the background clutter, target motion, and target aspect change is proposed in [10].

A popular and computationally efficient framework to handle this issue, under the hypothesis of elliptical spread of the target, is provided by Koch in [25] where an approximate Bayesian solution to the target tracking problem is proposed. Random matrices are exploited to model the ellipsoidal object extensions, which are treated as additional state variables to be estimated or tracked. The target kinematic states are modeled using a Gaussian distribution, while the ellipsoidal target extension is modeled using an inverse Wishart distribution. Random matrices are used to model extended targets under kinematic constraints [26]. In [45], [46] and [19], the integration of random matrices into the probabilistic multi-hypothesis tracking and the probability hypothesis density filter, respectively, address the multi-target tracking problem. Furthermore, a new approach is derived in [12] to overcome some of the weaknesses in [25]. Indeed, in [25] sensor inaccuracies are neglected and, if they are large in comparison to target size, the lack of modeling may lead to an overestimation of target size, see [11]. New measurement and time updates for [12] are proposed in [33] and [21], respectively. An extension of random matrices for non-ellipsoidal group and extended target tracking based on a combination of multiple ellipsoidal sub-objects, each represented by a random matrix, is discussed in [27]. A comparison between random matrices and the random hypersurface model [5] under a single target assumption is given in [4]. An interesting application using real-world radar data, acquired during the recovery operations of the Costa Concordia wreckage in October 2013, and the random matrices framework is reported in [16], [17].

An overview of the state-of-art for group and extended target tracking techniques is given in [32]. Sequential Monte Carlo methods and their variants are mainly discussed. An overview including Markov chain

Monte Carlo (MCMC) methods, random matrices approaches, and random finite set statistics techniques is also provided.

In radar signal processing a crucial point is given by the data conversion. The measurement of the target's position is usually reported in polar coordinates, while the target position and dynamic are usually modeled in Cartesian coordinates. The effects of data conversion have to be properly taken into consideration.

In this paper, we propose to investigate further the conversion between polar and Cartesian coordinates into the approach presented in [12]. An extended target tracking algorithm is presented here and two measurement models using two kinds of coordinate conversions (i.e. the standard and the unbiased ones) are illustrated and integrated into the random matrix framework. Sects. II-C and III-B show the proposed models and how it is possible to integrate them into the random matrix framework. Furthermore, we derive that the update equations are similar to the ones in [12]. Estimations for both kinematic parameters (i.e. positions and velocities) and sizes are performed. The performance of the proposed models are assessed on both simulated data (reproducing three different scenarios) and real data acquired by an X-band marine radar installed in the Gulf of La Spezia, Italy. Comparisons with the measurement model that neglects the sensors' noise effects (e.g. [25]) and the one with a constant covariance matrix [12] are provided. Automatic identification system (AIS) static and kinematic reports are exploited as ground truth in order to assess the performance. The simulation results are confirmed by real data. Ten different target datasets exploiting different kinds of targets, with over  $10^3$  frames of acquisition, are used to obtain a significant statistical analysis. It is demonstrated that the main advantage is the improvement in the estimation of target size, while comparable performance can be shown on the estimation of kinematic parameters. More specifically, gains up to 70% for the targets' width estimation accuracy and 65% for the length are observed by exploiting the proposed models. The integration of the proposed model into the gamma Gaussian inverse Wishart probability hypothesis density tracker [17] to address real scenarios with clutter and expected multiple extended targets is also provided and tested on real data. To the best of the authors' knowledge, this represents the first attempt to integrate the coordinate conversion into the random matrix framework and to quantitatively evaluate the effects of the sensors' noise and data conversion in ETT using both simulated and X-band marine radar data. Indeed, even if other few performance assessments relied upon these data can be found in the literature, see [11], [17], a serious analysis on the effects of the sensors' noise has not been accounted for.

Furthermore, the problem of dealing with nonlinearities is of great interest in radar signal processing

especially because these strategies risk to fail in specific cases and ad hoc improvements need to be implemented, see e.g. [37]. This is also the case of the proposed paper, in which the non-linearity problem (conversion of data) is completely neglected for the ETT literature, which has attracted great interest in the recent years. The impact in neglecting the noise and its conversion between polar and Cartesian coordinates is particularly clear, see Sect. V, on the target's size estimation. Nowadays, many papers propose its own extended target tracking approach and these techniques are increasing interest thanks to the high resolution features of several new radar systems. However, all these approaches neglect the polar-Cartesian conversion issue causing a bias in the estimation of the targets' extension. Thus, new researches can arise from this paper integrating the conversion in these approaches and by evaluating the benefits on real radar data. The contribute of the paper is twofold:

- **Theoretical.** We derive first the *converted measurement extended target tracking* (CM-ETT), which represents the extension of the converted measurement Kalman filter to the context of extended targets, see Sects. II-C and III. Furthermore, the similarities with respect to the work of Feldmann et al. [12] are remarked. Similar update equations can be derived but exploiting a state-dependent covariance noise matrix.
- **Experimental.** We show that the CM-ETT significantly outperforms the ETT state-of-the-art strategies by validating it in extensive experiments, that is hardly to find in this literature. Several simulated scenarios have been tested quantifying the benefits in using the proposed model. Furthermore, the use of 10 real datasets including data from several kinds of targets, acquired by our X-band marine radar in the Gulf of La Spezia, together with the AIS information enables us to further corroborate the simulated outcomes on real data. The integration with the multiple extended target strategy in [17] is also provided and validated on a challenging real data set.

The work presented in this paper is an extension of previously reported progress on ETT applied to X-band marine radar data [40]. A broader experimental analysis, a more detailed analysis of the literature, the introduction of the unbiased coordinate conversion, and the extension to the multiple ETT case integrating the proposed converted measurement model into the gamma Gaussian inverse Wishart probability hypothesis density tracker [17] validating it on real data, can be considered the main novelties of this paper with respect to the conference version.

The paper is organized as follows. Sect. II describes the Bayesian extended target modeling, including a coordinate conversion model approach. Sect. III presents the filtering equations that the modeling lead to. Sect. IV is devoted to the integration of the proposed approach into the gamma Gaussian inverse Wishart probability

hypothesis density tracker. The experimental results using both simulated and real data are shown in Sect. V. Results in the multiple extended target tracking case are presented in Sect. VI. Finally, conclusions and future developments are drawn in Sect. VII.

## II. BAYESIAN EXTENDED TARGET MODELING

This section is devoted to the description of the proposed measurement model using converted measurements and its integration into the Bayesian extended target tracking framework presented first in [25], and later improved in [12] with the consideration of the sensors' measurement errors. It is worth pointing out that the above-mentioned papers concentrate attention on the track filtering. Estimations under observation-to-track association uncertainty with possible presence of missed detections and false alarms are out-of-scope. The same assumption is made in this paper. Readers who are interested in this topic are instead encouraged to see [17] and Sect. IV in order to get deeper insights about the problem of the extended multi-target tracking for X-band marine radar data.

### A. State Model

The extended target kinematics (position and velocity) are defined in 2D Cartesian coordinates and modeled by the vector  $\mathbf{x}_k \triangleq [x_k, \dot{x}_k, y_k, \dot{y}_k]^T$ , where  $x_k, y_k$  and  $\dot{x}_k, \dot{y}_k$  are the position and velocity components along the  $X, Y$  directions, respectively, and  $[\cdot]^T$  is the transpose operator. The extended target's extent (shape and size) is assumed elliptic and is modeled by the positive definite matrix  $\mathbf{X}_k$ .

Let  $\mathbf{Z}^k = \{\mathbf{Z}_m\}_{m=0}^k$  denote all the measurement sets up to and including frame  $k$ . The extended target state, i.e.  $\mathbf{x}_k$  and  $\mathbf{X}_k$ , is Gaussian inverse Wishart distributed,

$$p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \mathcal{IW}(\mathbf{X}_k; \alpha_{k|k}, \hat{\mathbf{X}}_{k|k}) \quad (1)$$

where  $\hat{\mathbf{x}}_{k|k}$  and  $\mathbf{P}_{k|k}$  are the expected value and covariance of the Gaussian distribution, and  $\hat{\mathbf{X}}_{k|k}$  and  $\alpha_{k|k}$  are the expected value and degrees of freedom of the inverse Wishart distribution.

### B. Dynamic Model

The target's motion is described by a nearly constant velocity model [3]. The state-update equation is as follows

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{\Gamma}\mathbf{w}_k \quad (2)$$

where  $\mathbf{F} = \tilde{\mathbf{F}} \otimes \mathbf{I}_d$ ,  $\mathbf{I}_d$  is the identity matrix with dimension  $d \times d$  (i.e.  $2 \times 2$  in our case),  $\otimes$  denotes the Kronecker product,

$$\tilde{\mathbf{F}} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, \quad (3)$$

$T_s$  is the sampling time,  $\mathbf{\Gamma} = \tilde{\mathbf{\Gamma}} \otimes \mathbf{I}_d$ ,

$$\tilde{\mathbf{\Gamma}} = \sigma_{pos} \cdot \begin{bmatrix} T_s^2/2 \\ T_s \end{bmatrix}, \quad (4)$$

and  $\sigma_{pos}$  represents the process noise (equal in both  $X$  and  $Y$  directions). The process noise  $\mathbf{w}_k$  takes into account the target acceleration and the unmodeled dynamics and it is assumed to be Gaussian with zero-mean and identity covariance matrix.

The time evolution of the extent  $\mathbf{X}_k$  is modeled as approximately constant over time. This model is accurate for targets that can be assumed to move linearly, i.e. targets that do not turn significantly (a turn causes the extension to rotate). For the scenarios considered in this paper this assumption is true. Motion models for turning targets can be found in related literature, see e.g. [21].

### C. Measurement Model Using Converted Measurements

Measurements of the target's positions are usually provided in polar coordinates (i.e. in range and azimuth) for data acquired by radar systems. However, the target motion is typically modeled in Cartesian coordinates. Hence, a conventional linear Kalman filter can be exploited only after the measurements have been converted from polar to Cartesian coordinates. It is important for the tracking results that the effects of this conversion are properly taken into consideration.

The components of the  $j$ th measurement vector at frame  $k$  are defined as  $\zeta_k^j \triangleq [r_k^j, \theta_k^j]^T$ , where  $r_k^j$  and  $\theta_k^j$  are the  $j$ th range and azimuth radar measurements at frame  $k$ , respectively. These measurements are modeled as the true range and azimuth values, plus measurement errors that are zero-mean Gaussian distributed with standard deviations equal to  $\sigma_r$  and  $\sigma_\theta$ , respectively. To convert measurements from polar to Cartesian coordinates we employ the standard coordinate conversion,

$$\mathbf{z}_k^{L,j} \triangleq [x_k^{L,j}, y_k^{L,j}]^T = [r_k^j \cos \theta_k^j, r_k^j \sin \theta_k^j]^T \quad (5)$$

where the superscript  $L$  stands for linearization.

Taking the first order terms of the Taylor series expansion of the standard coordinate conversion, i.e. using linearization, we obtain the Cartesian coordinate errors, which have zero-mean and covariance matrix [3]

$$\mathbf{R}^L(\zeta_k^j) = \mathbf{J}(\zeta_k^j) \text{diag}([\sigma_r^2, \sigma_\theta^2]) \mathbf{J}^T(\zeta_k^j), \quad (6)$$

where

$$\mathbf{J}(\zeta_k^j) = \begin{bmatrix} \cos \theta_k^j & -r_k^j \sin \theta_k^j \\ \sin \theta_k^j & r_k^j \cos \theta_k^j \end{bmatrix} \quad (7)$$

is the Jacobian matrix, and  $\text{diag}(\cdot)$  indicates a diagonal matrix.

A remark is related to the validity of the standard coordinate conversion. A rule of thumb is provided in [3]. When it is not valid, the unbiased conversion [3] can be exploited to deal with the problem of converting measurements from polar to Cartesian coordinates. In

this case, we have that for the  $j$ th measurement at frame  $k$ ,  $\mathbf{z}_k^{U,j} \triangleq [x_k^{U,j}, y_k^{U,j}]^T$ , where

$$x_k^{U,j} = x_k^{L,j} b^{-1}, \quad (8)$$

$$y_k^{U,j} = y_k^{L,j} b^{-1}, \quad (9)$$

and  $b = \exp(-\sigma_\theta^2/2)$  assuming that the noise in the polar domain is Gaussian distributed.

The covariance matrix is as follows:

$$\mathbf{R}^U(\zeta_k^j) = \begin{bmatrix} \mathbf{R}_{11}^U(\zeta_k^j) & \mathbf{R}_{12}^U(\zeta_k^j) \\ \mathbf{R}_{21}^U(\zeta_k^j) & \mathbf{R}_{22}^U(\zeta_k^j) \end{bmatrix}, \quad (10)$$

where its elements are defined as [3]

$$\begin{aligned} \mathbf{R}_{11}^U(\zeta_k^j) &= (b^{-2} - 2)(r_k^j)^2 \cos^2 \theta_k^j \\ &\quad + [(r_k^j)^2 + \sigma_r^2][1 + b^4 \cos 2\theta_k^j]/2, \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{R}_{22}^U(\zeta_k^j) &= (b^{-2} - 2)(r_k^j)^2 \sin^2 \theta_k^j \\ &\quad + [(r_k^j)^2 + \sigma_r^2][1 - b^4 \cos 2\theta_k^j]/2, \end{aligned} \quad (12)$$

$$\begin{aligned} \mathbf{R}_{21}^U(\zeta_k^j) &= \mathbf{R}_{12}^U(\zeta_k^j) = b^{-2}(r_k^j)^2/2 \sin 2\theta_k^j \\ &\quad + [(r_k^j)^2 + \sigma_r^2]b^4/2 \sin 2\theta_k^j - (r_k^j)^2 \sin 2\theta_k^j. \end{aligned} \quad (13)$$

For the radar data used in this paper the standard coordinate conversion was empirically found to be sufficient, see Sect. V for further details.

We assume, as done in [12], [25], that at each frame there is a set of  $n_k$  independent Cartesian position measurements, denoted  $\mathbf{Z}_k = \{\mathbf{z}_k^j\}$  (i.e. either  $\mathbf{z}_k^{L,j}$  or  $\mathbf{z}_k^{U,j}$ ). The detection set likelihood is

$$p(\mathbf{Z}_k | n_k, \mathbf{x}_k, \mathbf{X}_k) = \prod_{j=1}^{n_k} p(\mathbf{z}_k^j | \mathbf{x}_k, \mathbf{X}_k). \quad (14)$$

Each detection  $\mathbf{z}_k^j$  is modeled as a noisy measurement of a reflection point  $\mathbf{y}_k^j$  located somewhere on the extended target. Further, each reflection point is modeled as a point randomly sampled from the target's extension. The detection likelihood is thus

$$p(\mathbf{z}_k^j | \mathbf{x}_k, \mathbf{X}_k) = \int p(\mathbf{z}_k^j | \mathbf{y}_k^j, \mathbf{x}_k, \mathbf{X}_k) p(\mathbf{y}_k^j | \mathbf{x}_k, \mathbf{X}_k) d\mathbf{y}_k^j \quad (15)$$

In other words, the detection likelihood (15) is the marginalization of the reflection point  $\mathbf{y}$  out of the estimation problem.

For the type of radar systems considered here the measurement noise is accurately modeled as zero mean Gaussian,

$$p(\mathbf{z}_k^j | \mathbf{y}_k^j, \mathbf{x}_k, \mathbf{X}_k) = \mathcal{N}(\mathbf{z}_k^j; \mathbf{y}_k^j, \mathbf{R}(\mathbf{y}_k^j)), \quad (16)$$

where  $\mathbf{R}(\mathbf{y})$  is the covariance matrix (i.e. either  $\mathbf{R}^L(\mathbf{y})$  using (6) or  $\mathbf{R}^U(\mathbf{y})$  using (10)) obtained when converting polar radar detections to Cartesian coordinates. Further,

the reflection points are accurately modeled as uniform samples from the target shape,

$$p(\mathbf{y}_k^j | \mathbf{x}_k, \mathbf{X}_k) = \mathcal{U}(\mathbf{y}_k^j; \mathbf{x}_k, \mathbf{X}_k). \quad (17)$$

As suggested by Feldmann et al. [12], for an elliptically shaped target the uniform distribution (17) is approximated by the following Gaussian distribution

$$p(\mathbf{y}_k^j | \mathbf{x}_k, \mathbf{X}_k) = \mathcal{N}(\mathbf{y}_k^j; \mathbf{H}\mathbf{x}_k, \rho\mathbf{X}_k) \quad (18)$$

where  $\rho$  is a scaling factor. Here  $\mathbf{H}$  is a measurement model that selects the position components in the state vector (i.e.  $\mathbf{H} = [\mathbf{I}_d, \mathbf{0}_d]$  where  $\mathbf{0}_d$  indicates the null matrix with  $d = 2$  in our case). In a simulation study Feldmann et al. showed that  $\rho = 1/4$  is a good parameter setting. In the result section we will address what is an appropriate parameter setting when using real radar data.

By combining equations (15), (16) and (18), the likelihood is

$$p(\mathbf{z}_k^j | \mathbf{x}_k, \mathbf{X}_k) = \int \mathcal{N}(\mathbf{z}_k^j; \mathbf{y}_k^j, \mathbf{R}(\mathbf{y}_k^j)) \mathcal{N}(\mathbf{y}_k^j; \mathbf{H}\mathbf{x}_k, \rho\mathbf{X}_k) d\mathbf{y}_k^j. \quad (19)$$

The marginalization (19) is analytically intractable. To achieve a computationally efficient measurement update, two assumptions are made. First, assume that in (16) the measurement noise covariance can be approximated as  $\mathbf{R}(\mathbf{y}_k^j) \approx \mathbf{R}(\mathbf{H}\mathbf{x}_k)$ , i.e.

$$p(\mathbf{z}_k^j | \mathbf{y}_k^j, \mathbf{x}_k, \mathbf{X}_k) \approx \mathcal{N}(\mathbf{z}_k^j; \mathbf{y}_k^j, \mathbf{R}(\mathbf{H}\mathbf{x}_k)). \quad (20)$$

REMARK In general, this approximation is less accurate the larger the distance is between the reflection point  $\mathbf{y}$  and the target's position, as given by  $\mathbf{H}\mathbf{x}_k$ . This implies that the approximation is less accurate the larger the target is, since a large target means that the distance between the reflection point and position may be large. For the radar sensors and the targets that are considered in this paper, we have empirically found that the approximation is sufficiently accurate.

Following the assumption in (20), applying it in (15), considering (18), and exploiting the product formula for two multivariate Gaussian distributions, we have that

$$p(\mathbf{z}_k^j | \mathbf{x}_k, \mathbf{X}_k) \approx \mathcal{N}(\mathbf{z}_k^j; \mathbf{H}\mathbf{x}_k, \rho\mathbf{X}_k + \mathbf{R}(\mathbf{H}\mathbf{x}_k)). \quad (21)$$

Considering that the prior target distribution is Gaussian inverse Wishart, i.e.

$$p(\mathbf{x}_k, \mathbf{X}_k | \mathbf{Z}^{k-1}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \times \mathcal{IW}(\mathbf{X}_k; \alpha_{k|k-1}, \hat{\mathbf{X}}_{k|k-1}), \quad (22)$$

we assume that the following approximation holds, i.e.

$$p(\mathbf{z}_k^j | \mathbf{x}_k, \mathbf{X}_k) \approx \mathcal{N}(\mathbf{z}_k^j; \mathbf{H}\mathbf{x}_k, \rho\mathbf{X}_k + \mathbf{R}(\mathbf{H}\hat{\mathbf{x}}_{k|k-1})), \quad (23)$$

namely the measurement noise covariance can be approximated by replacing  $\mathbf{x}_k$  with its predicted expected value  $\hat{\mathbf{x}}_{k|k-1}$ .

REMARK This approximation is trivially satisfied when  $\mathbf{R}(\cdot)$  is a constant matrix. In general the assumption holds approximately when  $\mathbf{R}(\cdot)$  does not vary too much in the uncertainty region for the extended target. Empirically we have found that, for the sensors and targets considered here, the signal to noise ratio is high enough to make the uncertainty region small enough.

Under the two assumptions above, the detection likelihood  $p(\mathbf{z}_k^j | \mathbf{x}_k, \mathbf{X}_k)$  assumes the same form as in [12] replacing the covariance noise matrix  $\mathbf{R}$  with its state-dependent version  $\mathbf{R}(\mathbf{H}\hat{\mathbf{x}}_{k|k-1})$ . Thus, the measurement update results analogous to the measurement update proposed in [12] by substituting  $\mathbf{R}$  with  $\mathbf{R}(\mathbf{H}\hat{\mathbf{x}}_{k|k-1})$ . This approach is here called *converted measurement extended target tracking* (CM-ETT) and its time and measurement updates are presented in the next section.

### III. CONVERTED MEASUREMENTS EXTENDED TARGET FILTERING

In this section we show the time update and measurement update for the models presented in the previous section.

#### A. Time Update

With the assumed independence between the estimates for centroid kinematics and extension and further assuming independent dynamic models for both of them, the standard Kalman filter prediction equations can be exploited [3], [12]:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}, \quad (24)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \mathbf{\Gamma}. \quad (25)$$

The prediction of the target's extension comes directly from the hypothesis that the extension does not tend to change over time, i.e.

$$\hat{\mathbf{X}}_{k|k-1} = \hat{\mathbf{X}}_{k-1|k-1}. \quad (26)$$

Finally, the prediction of the degrees of freedom parameter  $\alpha_{k|k-1}$  is given as [12]

$$\alpha_{k|k-1} = 2 + \exp(-T_s/\tau)(\alpha_{k-1|k-1} - 2), \quad (27)$$

where  $\tau$  is a time constant related to the agility with which the target may change its extension over time.

#### B. Measurement Update

The measurement updated expected value and covariance of the state vector estimate are obtained by a Kalman filter update [12]

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_{k|k-1}(\bar{\mathbf{z}}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}), \quad (28)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_{k|k-1}\mathbf{S}_{k|k-1}\mathbf{K}_{k|k-1}^T, \quad (29)$$

where

$$\mathbf{S}_{k|k-1} = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \frac{\mathbf{Z}_{k|k-1}}{n_k}, \quad (30)$$

$$\mathbf{K}_{k|k-1} = \mathbf{P}_{k|k-1}\mathbf{H}^T\mathbf{S}_{k|k-1}^{-1} \quad (31)$$

are the innovation covariance and the gain, and

$$\mathbf{Z}_{k|k-1} = \rho\hat{\mathbf{X}}_{k|k-1} + \mathbf{R}_k(\mathbf{H}\hat{\mathbf{x}}_{k|k-1}) \quad (32)$$

indicates the predicted covariance of a single measurement. Note that  $\mathbf{R}_k(\mathbf{H}\hat{\mathbf{x}}_{k|k-1})$  depends on the predicted expected value  $\hat{\mathbf{x}}_{k|k-1}$  (differently from [12] where it is constant) and the posterior of the kinematic state conditioned on  $\mathbf{x}_k$  is again assumed to be close to a normal distribution.

The updated expected value and degrees of freedom of the extension estimate  $\hat{\mathbf{X}}_{k|k}$  are obtained as follows [12]

$$\hat{\mathbf{X}}_{k|k} = \frac{\alpha_{k|k-1}\hat{\mathbf{X}}_{k|k-1} + \hat{\mathbf{N}}_{k|k-1} + \hat{\mathbf{Z}}_{k|k-1}}{\alpha_{k|k}}, \quad (33)$$

$$\alpha_{k|k} = \alpha_{k|k-1} + n_k, \quad (34)$$

where

$$\hat{\mathbf{N}}_{k|k-1} = \hat{\mathbf{X}}_{k|k-1}^{1/2}\mathbf{S}_{k|k-1}^{-1/2}\mathbf{N}_{k|k-1}(\mathbf{S}_{k|k-1}^{-1/2})^T(\hat{\mathbf{X}}_{k|k-1}^{1/2})^T, \quad (35)$$

$$\hat{\mathbf{Z}}_{k|k-1} = \hat{\mathbf{X}}_{k|k-1}^{1/2}\mathbf{Z}_{k|k-1}^{-1/2}\bar{\mathbf{Z}}_k(\mathbf{Z}_{k|k-1}^{-1/2})^T(\hat{\mathbf{X}}_{k|k-1}^{1/2})^T, \quad (36)$$

$$\mathbf{N}_{k|k-1} = (\bar{\mathbf{z}}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})(\bar{\mathbf{z}}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})^T, \quad (37)$$

and

$$\bar{\mathbf{z}}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \mathbf{z}_k^j, \quad (38)$$

$$\bar{\mathbf{Z}}_k = \sum_{j=1}^{n_k} (\mathbf{z}_k^j - \bar{\mathbf{z}}_k)(\mathbf{z}_k^j - \bar{\mathbf{z}}_k)^T \quad (39)$$

are the centroid measurement and the measurement spread. Note that the marginalized prior density of the target extension is assumed to be an inverse Wishart density [25]. This implies that the posterior is again of the same form.

#### IV. GAMMA GAUSSIAN INVERSE WISHART PROBABILITY HYPOTHESIS DENSITY

A multiple extended target tracker is briefly described in this section. The converted measurements model in Sect. II is integrated in the measurement update exploiting the results in Sect. III to obtain an improvement in the targets' size estimation. The core of the tracker is a probability hypothesis density (PHD) filter which provides for each radar frame the kinematics, the size and the shape, as well as the expected number of detections relevant to each target occurring in the surveillance area. The filter is fed with the measurements provided by the detector described in Sect. V-D.

The extended target state  $\xi_k$  can be redefined as

$$\xi_k \triangleq (\gamma_k, \mathbf{x}_k, \mathbf{X}_k), \quad (40)$$

where the random vector  $\mathbf{x}_k$  and  $\mathbf{X}_k$  represent again the kinematic and extension states and  $\gamma_k > 0$  is the measurement rate that describes how many measurements the target, on average, generates per frame. The number of target generated measurements is assumed to be Poisson distributed, and  $\gamma_k$  is in this case the Poisson rate [13], [14].

Conditioned on a history of previous measurement sets,  $\mathbf{Z}^k$ ,  $\xi_k$  is modeled as a gamma-Gaussian-inverse Wishart (GGIW) distribution [12], [18],

$$p(\xi_k | \mathbf{Z}^k) = p(\gamma_k | \mathbf{Z}^k)p(\mathbf{x}_k | \mathbf{Z}^k)p(\mathbf{X}_k | \mathbf{Z}^k) \quad (41a)$$

$$= \mathcal{G}(\gamma_k; \alpha_{k|k}^g, \beta_{k|k})\mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \quad (41b)$$

$$\times \mathcal{IW}(\mathbf{X}_k; \alpha_{k|k}, \hat{\mathbf{X}}_{k|k})$$

$$= \mathcal{GGIW}(\xi_k; \zeta_{k|k}), \quad (41c)$$

where  $\zeta_{k|k} = \{\alpha_{k|k}^g, \beta_{k|k}, \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \alpha_{k|k}, \hat{\mathbf{X}}_{k|k}\}$  is the set of GGIW density parameters.

The probability hypothesis density (PHD)  $D_{k|k}(\cdot)$  is an intensity function whose integral is the expected value of the number of targets, and whose peaks correspond to likely target locations, see [15], [29]. The PHD intensity is typically approximated either using Sequential Monte Carlo methods, see [43], or using distribution mixtures, see [20], [42]. In this case, the PHD intensity  $D_{k|k}(\cdot)$  at frame  $k$ , given the measurement sets up to and including frame  $k$ , is approximated by a mixture of GGIW distributions,

$$D_{k|k}(\xi_k) = \sum_{j=1}^{J_{k|k}} w_{k|k}^{(j)} \mathcal{GGIW}(\xi_k; \zeta_{k|k}^{(j)}), \quad (42)$$

where  $J_{k|k}$  is the number of components,  $w_{k|k}^{(j)}$  is the weight of the  $j$ th component, and  $\zeta_{k|k}^{(j)}$  is the density parameter of the  $j$ th component. On behalf of brevity, no detail for the GGIW tracker is provided. The interested reader can refer to [17] to get all the information related to the time and measurement update equations and the post-processing step (i.e. mixture reduction, track extraction, and track estimation).

#### V. EXPERIMENTAL RESULTS

The validity of the proposed approach is here demonstrated by exploiting both simulated and real data. The latter are acquired by an X-band marine radar located in La Spezia, Italy. Further to the aim of validating our approach, as already proposed in [31], we exploit, for tracking assessment, the automatic identification system (AIS) [1] static/kinematic messages.

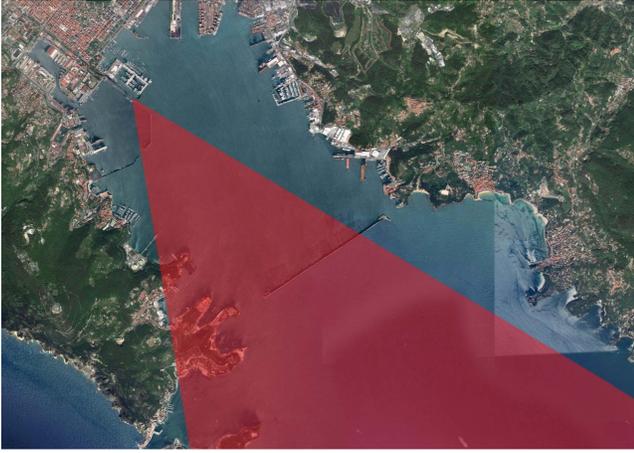


Fig. 1. The X-band marine radar’s field of view (in red).

The proposed Bayesian ETT method is compared to two other approaches: one random matrix-based tracking algorithm without a model accounting for the sensors’ errors, i.e. with  $\mathbf{R}_k = \mathbf{0}$  [25], and another random matrix-based approach that exploits a constant covariance matrix  $\mathbf{R}$  [12]. The proposed method (both in the case of the standard and the unbiased conversion) is here named converted measurements-ETT (CM-ETT). For a constant covariance matrix  $\mathbf{R}$  three different possibilities are tested. They are calculated using (6) by setting  $\theta_k^j$  to the azimuth mean value on the surveillance area, and letting  $r_k^j$  assume one of three values. This gives three different matrices:  $\mathbf{R}_1$  calculated for targets that move close to the sensor around range 0.5 km;  $\mathbf{R}_2$  calculated for range 2 km, corresponding to the middle of the considered surveillance area; and  $\mathbf{R}_3$  hypothesizes that the target sails in a longer range area, around 4 km range.

This section is organized as follows. A description of the X-band marine radar experiment is provided to the reader in Sect. V-A. The real datasets used for the validation of the approach and the AIS message format are briefly illustrated in Sects. V-B and V-C, respectively. The detection strategy is presented in Sect. V-D, while, Sects. V-E and V-F are related to the analysis of the results on simulated and real data, respectively.

#### A. X-band Marine Radar Experiment

The X-band marine radar is a coherent linear frequency modulated continuous wave radar [38]. It is a compact and lightweight system, still maintaining a high performance with relatively simple electronics, since the transmitted power is low and constant.

The radar is installed in the Gulf of La Spezia (Italy) (see the radar’s field of view in Fig. 1). The use of pulse compression [44] and a small transmitted power make it a compact, quickly deployable, and scalable system, used for research in the areas of extended target detection and tracking, with application to surveillance

TABLE I  
Marine Radar Specifications

Parameter	Specification
Frequency	9.6 GHz
Bandwidth	Adjustable up to 150 MHz
Range resolution	$\Delta r = 1$ m
Antenna type	Rotating slotted waveguide
Antenna angular resolution	$\Delta\theta = 1^\circ$
Antenna angular aperture elevation	$20^\circ$
Gain	32 dBi
Azimuth antenna speed	0 (stopped) up to 40 revolutions per minute
Polarization	Linear horizontal
Transmitted power	Adjustable 50 mW–5 W (17–37 dBm)
Pulse repetition frequency	Adjustable 350 Hz–10 KHz

of small craft at short to medium ranges (maximum 5–6 km) for harbor protection and coastal surveillance.

The marine radar has an antenna mounted on a rotor with variable speed of rotation and the possibility to lock and hold the position towards a specific direction with  $0.1^\circ$  accuracy. The main radar parameters are shown in Tab. I. The radiating system for this node consists of two slotted waveguide antennas, one for transmitting and another for receiving, both using linear horizontal polarization. Nevertheless, cross polar vertical-horizontal signatures can be collected in bistatic mode. The high directivity of the slotted waveguide allows a precise determination of the angular position of a target, also allowing the acquisition of targets at long distance with small power.

#### B. Datasets

Ten datasets have been acquired by the X-band marine radar located in the Gulf of La Spezia, Italy. They have been used for the performance assessment of our approach. The 10 AIS tracks, one for each dataset, are depicted in Fig. 2. They have been generated by 8 different ships. The main features of these ships are briefly outlined below:

- *Grand Holiday* is a Bahamian passenger (cruise) ship with Maritime Mobile Service Identity (MMSI) equal to 255803790. The size of this ship is 222 m  $\times$  32 m. The gross tonnage is 46052 t.
- *Palinuro* is a three-masted, iron-hulled barquentine, active as a sails training vessel for the Italian Navy. The MMSI is equal to 247939000. The size of this ship is 59 m  $\times$  10 m.
- *Fabio Duo* is an Italian cargo with MMSI equal to 247241500. The size of this ship is 80 m  $\times$  16 m. The gross tonnage is 2080 t.
- *Euro* is an Italian passenger ship with MMSI equal to 247030500. The size of this ship is 28 m  $\times$  6 m.
- *Monokini* is a pleasure craft with MMSI equal to 6904672. The size of this ship is 45 m  $\times$  8 m.

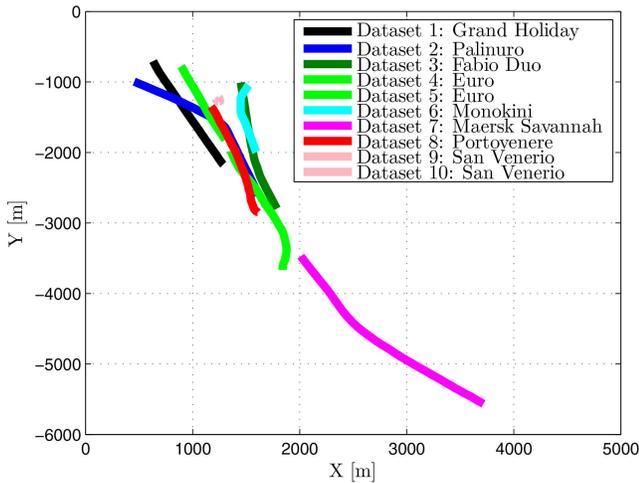


Fig. 2. AIS tracks for all the analyzed datasets.

- *Maersk Savannah* is a Danish container ship with MMSI equal to 219231000. The size of this ship is 334 m × 45 m. The gross tonnage is 92293 t.
- *Portovenere* is an Italian tug with MMSI equal to 247076200. The size of this ship is 28 m × 10 m. The gross tonnage is 279 t.
- *San Venerio* is an Italian tug with MMSI equal to 247841000. The size of this ship is 31 m × 10 m. The gross tonnage is 307 t.

### C. AIS Data

Ships and vessels exceeding a given gross tonnage<sup>1</sup> are equipped with AIS transponders for position-reporting, as established by the SOLAS Convention [1]. Ships repeatedly broadcast their name, position, and other details for automatic display on nearby ships. While this allows ships to be aware and keep track of other ships in their immediate vicinity, coastal states will also be able to receive, plot, and log the data by means of base stations along the coast. AIS reports contain both dynamic information (e.g. latitude, longitude, course-over-ground (COG), speed-over-ground (SOG), time) and static information (e.g. vessel type, size information).

To allow their proper use as ground truth for our applications, AIS ship reports are checked in order to remove possible outliers, missing position reports, and unreliable data. An interpolation phase is also required to align in time radar data and AIS contacts.

### D. Detection Strategy

Each radar image is processed by a detector to obtain a cloud of detections that represents the input for the Bayesian extended target tracking approaches. In

<sup>1</sup>The AIS is required for all the ships exceeding 300 gross tonnage and engaged on international voyages, for all cargo ships of 500 gross tonnage, not engaged on international voyages, and all passenger ships. On average, a gross weight of 300 t corresponds to a length of about 25 m.

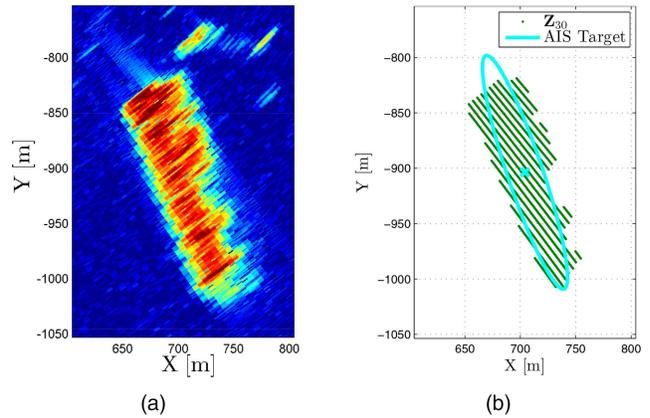


Fig. 3. (a) Amplitude of X-band marine radar data and (b) detections with AIS information for Frame 30 on the *Grand Holiday* dataset.

this paper, we exploit a maximum likelihood detector, which represents a good trade-off between computational burden and performance. Generally speaking, other more computationally demanding approaches are possible, i.e. an ordered statistic-based constant false alarm rate (CFAR) detector [36], but their comparison is here considered out of scope.

Two assumptions are made in this phase. First, the conditional independence among nearby pixels is assumed. Furthermore, the distributions of the power images under the *target* and *non-target* hypotheses are considered exponential. The rate parameters  $\lambda_t > 0$  (i.e. under *target* hypothesis) and  $\lambda_{nt} > 0$  (i.e. under *non-target* hypothesis), which characterize the whole exponential distributions, are estimated using the *k-means* clustering algorithm [23]. An example of detections for Frame 30 on the *Grand Holiday* dataset is depicted in Fig. 3.

### E. Simulated Results

The analysis of results reached by the compared algorithms on simulated data is here performed. Three kinds of simulations are exploited in order to understand the capabilities of the approach to work in similar conditions with those expected in the real-world. First, a ship of 80 m × 30 m has been simulated sailing on a straight line from 1 km to 4 km along the range direction using a nearly constant velocity model [3] with zero-mean Gaussian noise described by the parameter  $\sigma_{pos}$ , see Eq. (2). The spread of the detections is Gaussian distributed in polar coordinates according to the used model with scaling parameter  $\rho$  equal to 1. The simulator parameters are shown in Tab. II.

Fig. 4 shows the comparison between the standard, see Eq. (5), and the unbiased, see Eqs. (8) and (9), conversions. We can easily see that the outcomes provided by both the models are equivalent (for both kinematic and size estimations). This further corroborates the validation limit rule of thumb in [3], which claims the equivalence between the two models for the considered radar and surveillance area extension. Indeed, following

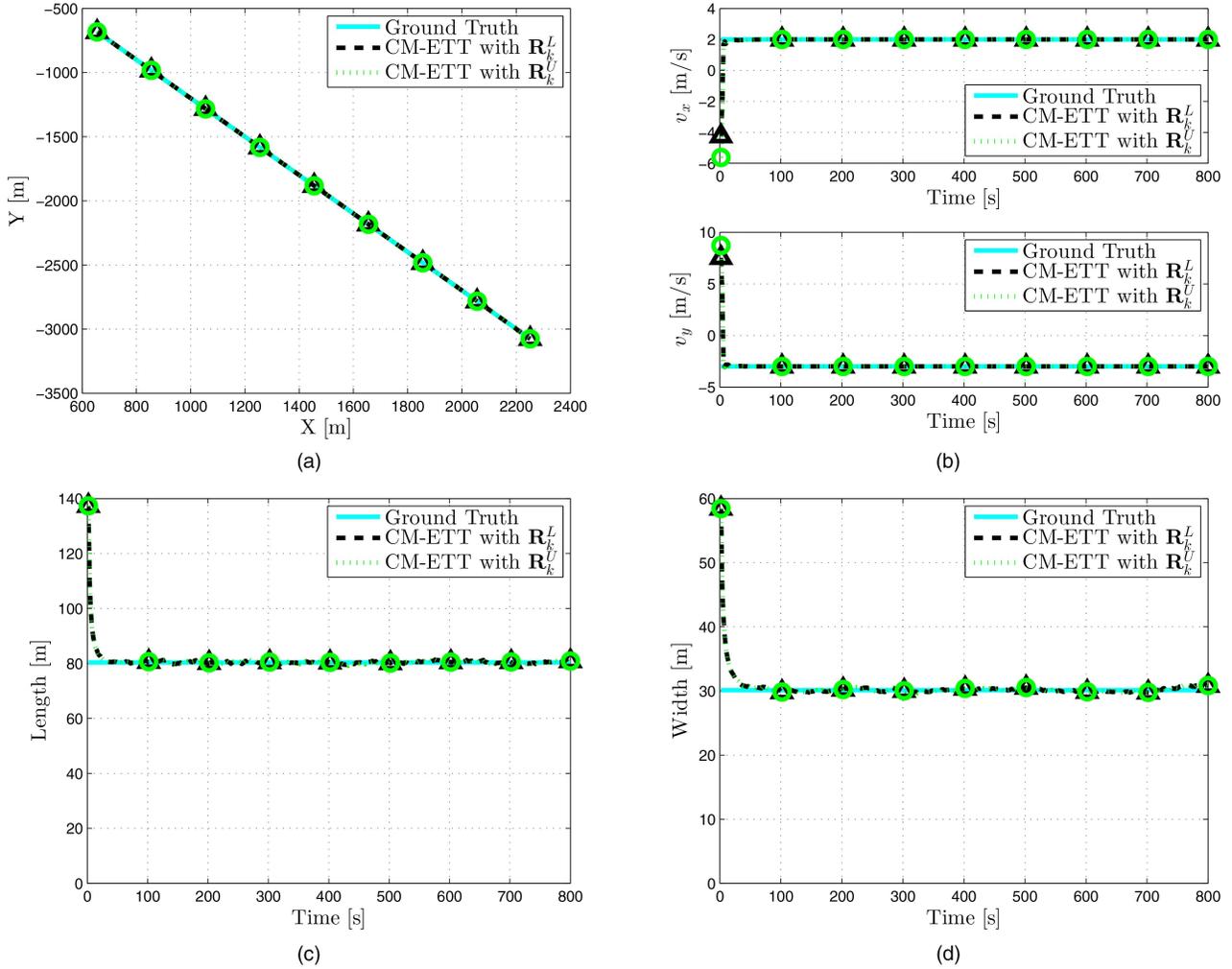


Fig. 4. (a) Position, (b) velocity, (c) length, and (d) width estimations for the proposed CM-ETT approaches on simulated data. Triangular markers are used for indicating the approach using the standard coordinate conversion model, while, circles are exploited for the unbiased conversion.

TABLE II  
Parameter Setting Simulator

Parameter	Value	Specification
$T_s$	2 s	Sampling time
$\sigma_{pos}$	$10^{-4}\text{ms}^{-2}$	Std. process noise
$\Sigma$	$\begin{bmatrix} 0.6653 & -0.6453 \\ -0.6453 & 1.1764 \end{bmatrix} \cdot 10^3$	Cov. spread target
$k_{\max}$	400	Number of frames
$N_d$	2000	Num. detects. frame
$\sigma_r^s$	0.5 m	Std. noise range
$\sigma_\theta^s$	$0.5^\circ$	Std. noise azimuth

the above-mentioned rule, the maximum range  $r_{\max}$  that allows proper use of the standard coordinate conversion is defined as

$$r_{\max} = 0.4 \frac{\sigma_r}{\sigma_\theta}; \quad (43)$$

that is, by substituting the values in Tab. IV, about 3 km, in agreement with the defined surveillance area. Thus,

in order to ease the reading of the following results, from hereon we will omit the unbiased conversion.

**REMARK** The rule of thumb in (43) results satisfied in the most of radar systems, but this is not the case in sonar systems or in long range radar with small  $\sigma_r$  and relatively large  $\sigma_\theta$ . When the above-mentioned condition is not verified, the unbiasedness property is not valid and the unbiased conversion has to be exploited instead of the simpler linearization.

The estimations provided by the compared approaches are depicted in Fig. 5. The outcomes confirm the ability of the proposed models to properly take into account the measurement noise. Because the analyzed case shows a ship that sails radially with respect to the radar position and the inaccuracy in range is less than the one in azimuth (as we will see after for the first real case), no difference is perceptible along the range direction (i.e. the target's length). The advantages are instead obvious along the azimuth direction (i.e. the target's width). More specifically, comparing the proposed approach with the one using  $\mathbf{R}_k = \mathbf{0}$  or the one

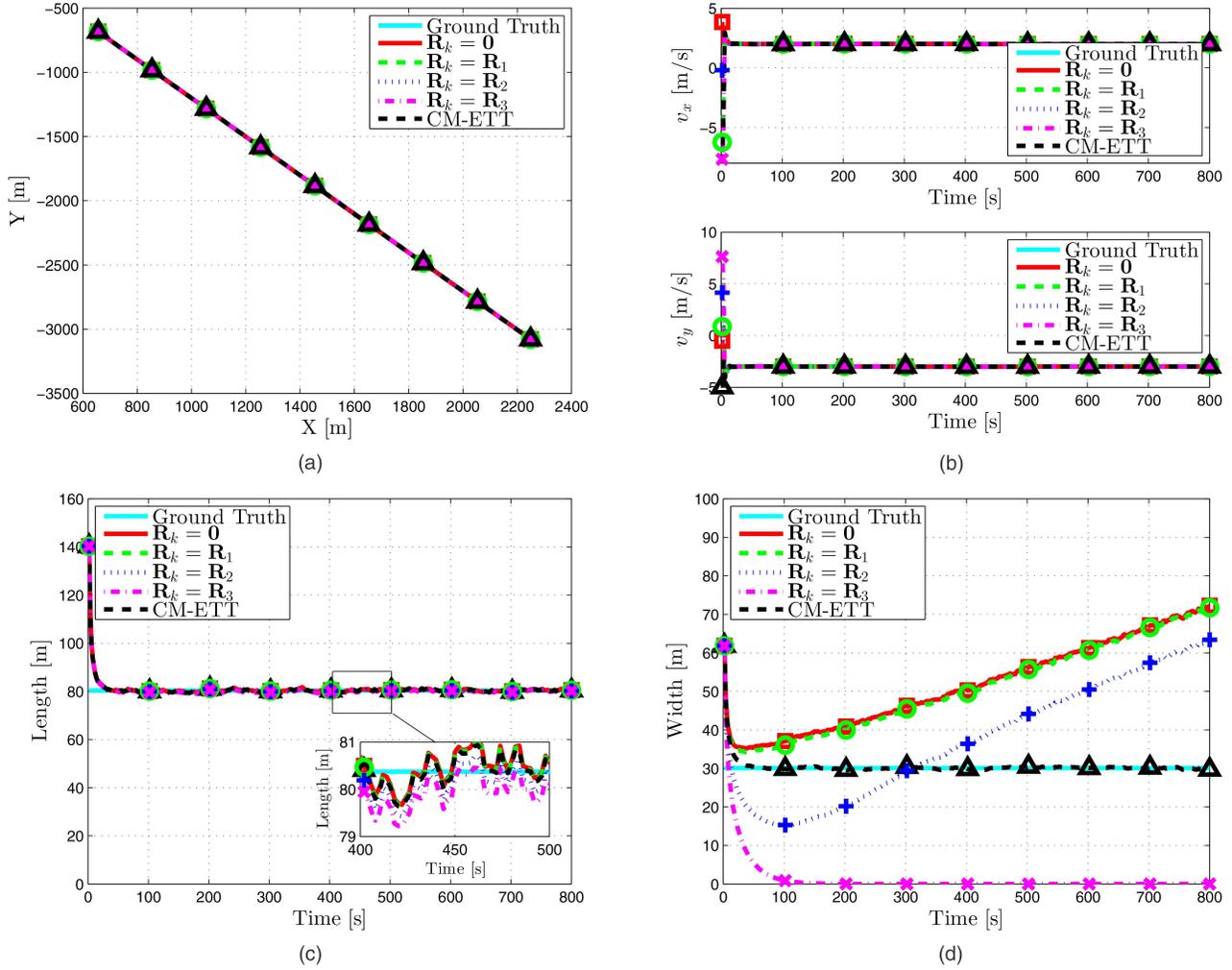


Fig. 5. (a) Position, (b) velocity, (c) length, and (d) width estimations for the compared approaches on simulated data for a target that follows a radial track.

using  $R_1$  it is clear that the more the target obtains large values of range, the greater the advantages are (see the differences for high time values). Both  $R_k = 0$  and  $R_1$  result in the width being overestimated. Using  $R_3$  gives the opposite behavior, i.e. the width is underestimated. Finally,  $R_2$  gives a performance that is in-between the results of the  $R_1$  and the  $R_3$  algorithms.

The second test case simulates the target sailing in an almost constant range track. Because we simulate an almost constant range track the opposite width/length estimation results are expected with respect to the previous test case. Fig. 6 shows the kinematic and size estimations provided by the compared approaches. Again, the CM-ETT shows its ability to properly estimate both the length and width parameters. As expected, considerable advantages are shown for the cross-range size estimation (length), while comparable performance can be pointed out for the estimation of the width parameter (along-range size).

The third test case in Fig. 7, where a non-along constant range and non-radial track is simulated, corroborates that the proposed CM-ETT gives improved performance both for the cross-range size and the along-range

size. Again, comparing the proposed approach with the one using  $R_k = 0$  or the one using  $R_1$  it is clear that the more the target obtains large values of range, the greater the advantages are. Finally,  $R_2$  represents again a good compromise among the compared approaches.

A final note is related to the estimation of the kinematic parameters (i.e. position and velocity). All the algorithms perform well and the results reached by them can be considered comparable, see Figs. 5(a)–(b), Figs. 6(a)–(b), and Figs. 7(a)–(b). Indeed, for this application, the sensors' inaccuracies mainly impact the estimation of the ship sizes instead of the kinematic parameters.

Finally, Tab. III shows the performance assessment for all the three simulated test cases. Best results are in boldface. The root mean square errors (RMSEs) in position (i.e.  $\epsilon^{pos}$ ), velocity (i.e.  $\epsilon^{vel}$ ), width (i.e.  $\epsilon^{wid}$ ), and length (i.e.  $\epsilon^{len}$ ) are calculated. Furthermore, the RMS Frobenius error (FE) (i.e.  $\epsilon^{FE}$ ) between the simulated (reference) matrix  $S$ , which describes the ellipsoidal simulated target, and the estimations provided by the 5 compared approaches  $\hat{X}$  is shown in Tab. III. This is defined as  $\|\hat{X} - S\|_F$ , where  $\|\cdot\|_F$  is the Frobenius

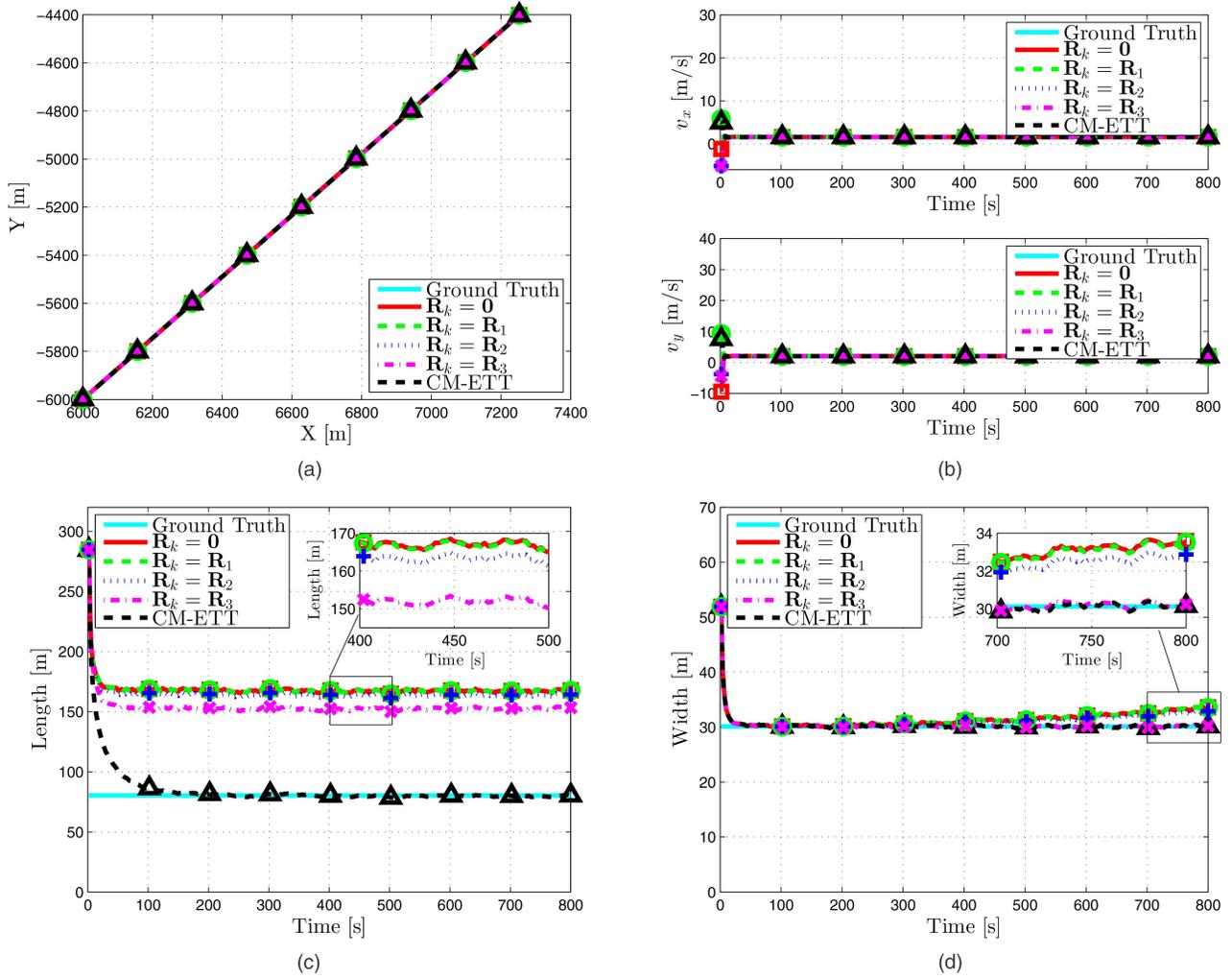


Fig. 6. (a) Position, (b) velocity, (c) length, and (d) width estimations for the compared approaches on simulated data for a target that follows an almost along constant range track.

TABLE III  
Performance Assessment on Simulated Data

Test Cases	Methods	$e^{pos}$	$e^{vel}$	$e^{wid}$	$e^{len}$	$e^{FE}$
Radial Track	$\mathbf{R}_k = \mathbf{0}$	<b>0.242</b>	<b>0.003</b>	25.8	<b>0.4</b>	595.0
	$\mathbf{R}_k = \mathbf{R}_1$	<b>0.242</b>	<b>0.003</b>	25.2	<b>0.4</b>	578.3
	$\mathbf{R}_k = \mathbf{R}_2$	<b>0.242</b>	<b>0.003</b>	17.3	0.5	361.5
	$\mathbf{R}_k = \mathbf{R}_3$	0.247	<b>0.003</b>	30.1	0.6	249.8
	CM-ETT	<b>0.242</b>	<b>0.003</b>	<b>0.2</b>	<b>0.4</b>	<b>18.9</b>
Almost Constant Range Track	$\mathbf{R}_k = \mathbf{0}$	0.441	<b>0.004</b>	1.6	87.1	5414.6
	$\mathbf{R}_k = \mathbf{R}_1$	0.441	<b>0.004</b>	1.6	86.9	5396.0
	$\mathbf{R}_k = \mathbf{R}_2$	0.441	<b>0.004</b>	1.3	83.5	5117.4
	$\mathbf{R}_k = \mathbf{R}_3$	0.441	<b>0.004</b>	<b>0.2</b>	72.1	4232.3
	CM-ETT	<b>0.439</b>	<b>0.004</b>	<b>0.2</b>	<b>1.4</b>	<b>56.8</b>
Non-along Constant Range/Non-radial Track	$\mathbf{R}_k = \mathbf{0}$	<b>1.545</b>	<b>0.004</b>	8.1	12.3	677.9
	$\mathbf{R}_k = \mathbf{R}_1$	<b>1.545</b>	<b>0.004</b>	7.9	12.0	659.8
	$\mathbf{R}_k = \mathbf{R}_2$	1.546	<b>0.004</b>	5.1	7.4	401.2
	$\mathbf{R}_k = \mathbf{R}_3$	1.548	<b>0.004</b>	29.8	8.1	584.2
	CM-ETT	1.546	<b>0.004</b>	<b>0.2</b>	<b>0.4</b>	<b>18.2</b>

norm. This further metric is also able to capture all the differences between estimated and reference matrices, e.g. due to target rotations. The first samples are left out

in the calculation of the errors because of the random initializations of the compared approaches. Conclusions as above can be drawn starting from the analysis of the

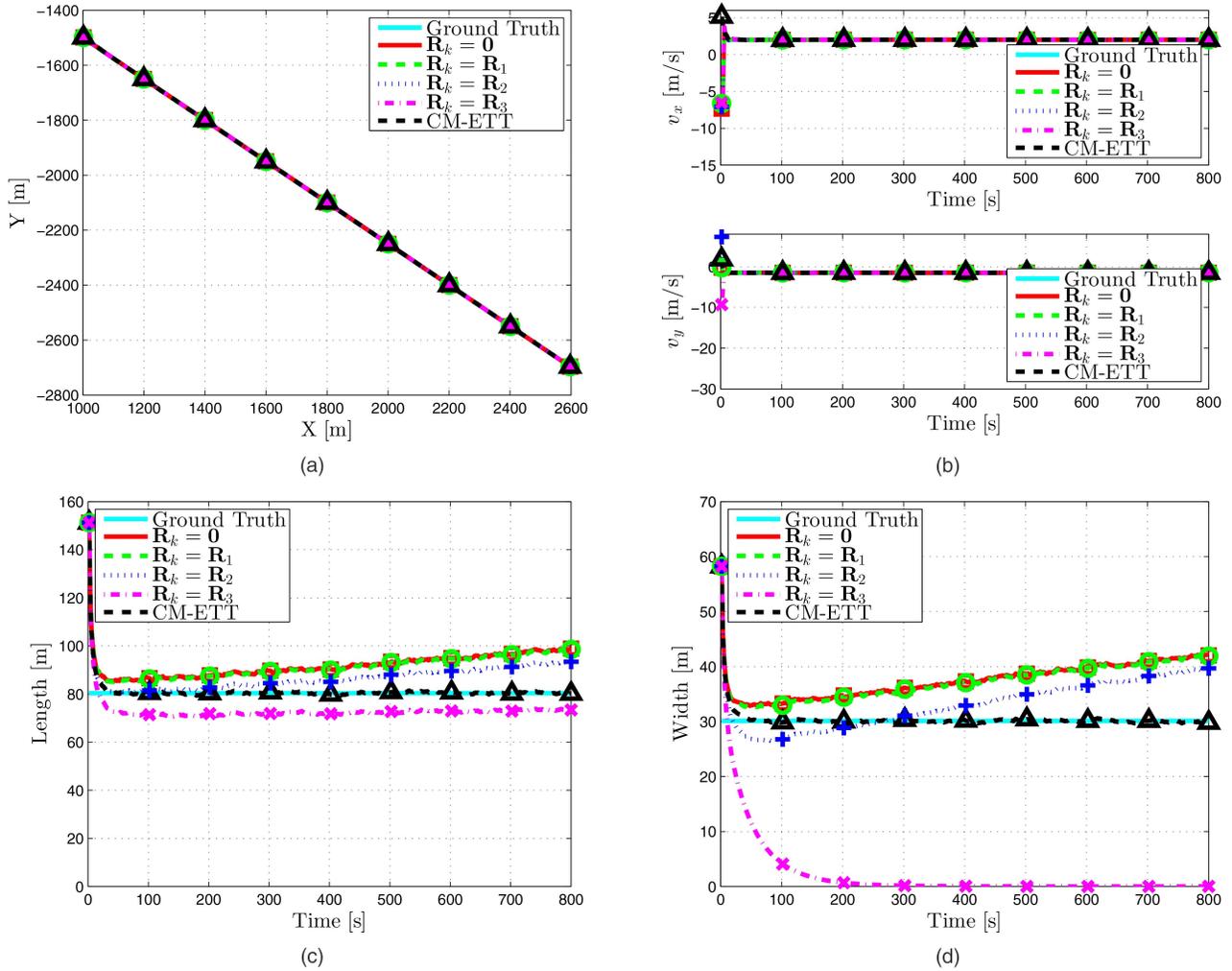


Fig. 7. (a) Position, (b) velocity, (c) length, and (d) width estimations for the compared approaches on simulated data for a target that follows a non-radial and non-constant range track.

outcomes in the table. These results will be further corroborated by the analysis of the real data provided in the next subsection.

#### F. Real X-band Marine Radar Data

The description of the outcomes on real data acquired by the X-band marine radar described in Sect. V-A is here provided. The main features of the 8 ships that generate the 10 datasets are shown in Sect. V-B.

Initially, we determine an appropriate value for the parameter  $\rho$  (cf. Eq. (18)). Two values of  $\rho$  are tested using the CM-ETT approach. In a simulation study presented by Feldmann et al. [12], it is suggested to use  $\rho = 1$  to model a Gaussian spread of the detections, while  $\rho = 1/4$  models a uniform distribution. Fig. 8 clearly shows that the CM-ETT with  $\rho = 1/4$  performs better obtaining a closer match with the AIS ship information. This experimental analysis confirms that data with a uniform detection spread is best modeled by  $\rho = 1/4$ . In the remainder of the paper, the compared extended target filters are implemented with  $\rho = 1/4$ . The other tracking parameters used in the experiments

are shown in Tab. IV. The sampling time  $T_s$  is indicated ranging from 2 s to 5 s. The reason why we have a range instead of a fixed value is that the azimuth antenna speed to acquire the 10 real datasets is different from a dataset to the other, see Tab. I for the radar specifications. Hence, we have different parameters' configuration in order to obtain a trade-off between the sampling time and the number of samples acquired along the azimuth direction, which, for instance, can have an impact on the aliasing. Therefore, the  $T_s$  parameter used in the tracking approach can be directly derived by the selected azimuth antenna speed value  $\theta$  (i.e.  $T_s = 60/\theta$ ). The  $\tau$  parameter is instead adjusted according to the used  $T_s$  value.  $\tau$  is related to the agility with which the target may change its extension over time. Thus, the datasets with lower  $T_s$  values tend to have a more static (i.e. less variable along frames) target extensions (i.e. higher  $\tau$  parameters are advisable). This is due to the fact that targets sail for a higher number of frames in the same zone and, thus the acquisition system tends to have a same target representation (i.e. less variable target extensions are expected). However, the tuning of

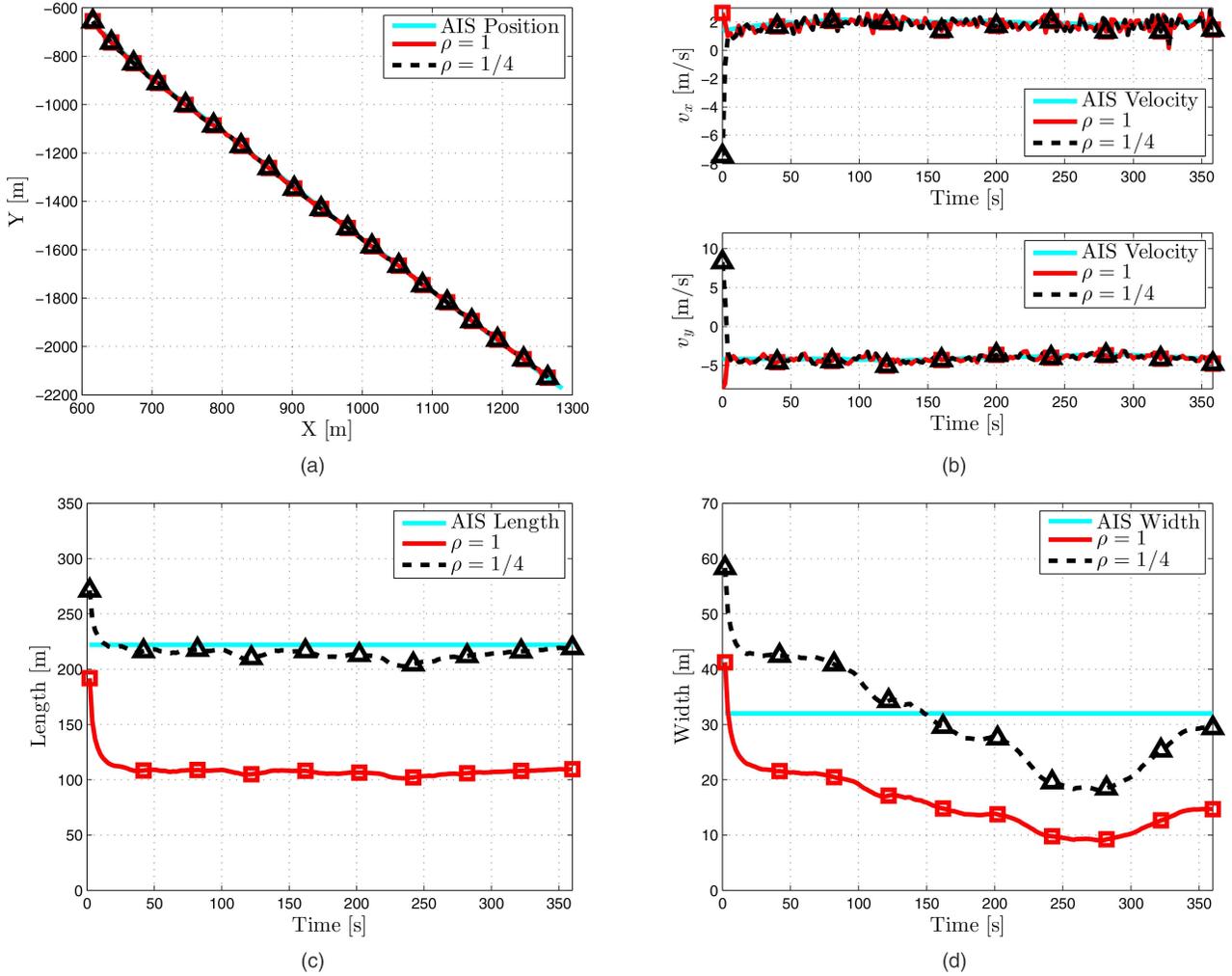


Fig. 8. (a) Position, (b) velocity, (c) length, and (d) width estimation for the CM-ETT using  $\rho = 1$  (red solid line) and  $\rho = 1/4$  (black dashed line) on the *Grand Holiday* dataset.

TABLE IV  
Parameter Setting Real Cases

Parameter	Value	Specification
$T_s$	2 s–5 s	Sampling time
$\tau$	10/5	Agility object size
$\sigma_{pos}$	0.2 ms <sup>-2</sup>	Std. process noise
$v_{max}$	10 ms <sup>-1</sup>	Maximum velocity
$\sigma_r$	$\Delta r/2$ (see Tab. I)	Std. noise range
$\sigma_\theta$	$\Delta\theta/2$ (see Tab. I)	Std. noise azimuth

this parameter as the other parameters in Tab. IV cannot be considered critical, e.g.  $\sigma_{pos}$  and  $v_{max}$  have simply been tuned according to the kinds of targets and the area under test (i.e. medium and large ships that sail in the near coastal area).

A graphic representation of the gains in estimating the size by properly accounting for the measurement noise is provided in Fig. 9 on the *Grand Holiday* dataset. The target, in this case, is moving toward higher range values in an almost radial direction, as can be seen in Fig. 2. Six frames are depicted in Fig. 9 starting

from Frame 30 to Frame 180 with temporal resolution equal to 60 s (i.e. an image every 30 frames). We only compare results using the proposed approach and using  $\mathbf{R}_k = \mathbf{0}$ , because showing all results makes the figures too cluttered. The more the ship sails toward high range values (i.e. the higher the frame number), the greater the spread of the detections. This behavior is mainly due to the polar geometry of the acquisition of the radar.

A first remark is related to Fig. 9(a). Indeed, it is simple to see that in the radar’s short range operating region, the proposed model that compensates the radar’s noise effects does not gain advantage with respect to the  $\mathbf{R}_k = \mathbf{0}$  model. The advantages between the proposed approach and the  $\mathbf{R}_k = \mathbf{0}$  model become more evident with increasing range. Note that due to the almost radial track, the target width parameter shows the greatest performance gain. A better match between the algorithm that runs with the proposed model and the ground-truth is straightforward. These outcomes confirm the simulations and they are shown in Figs. 10(c) and (d). Kinematic features are well captured by all the

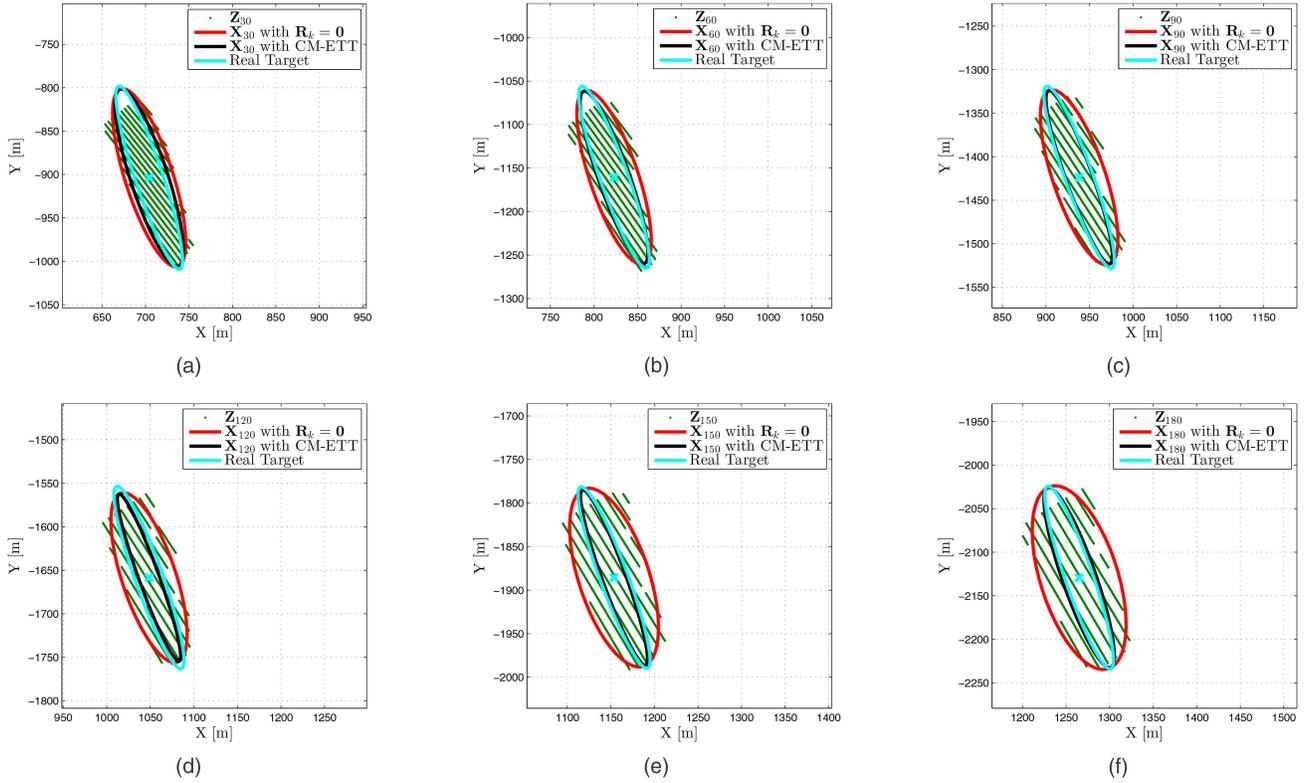


Fig. 9. Estimated ellipsoids provided by the  $\mathbf{R}_k = \mathbf{0}$ , the CM-ETT, and the ground-truth on the *Grand Holiday* dataset. (a) Frame 30. (b) Frame 60. (c) Frame 90. (d) Frame 120. (e) Frame 150. (f) Frame 180.

TABLE V  
Performance Assessment on Real Data

Methods	$\epsilon^{pos}$ [m]	$\epsilon^{vel}$ [ $\text{ms}^{-1}$ ]	$\epsilon^{wid}$ [m]	$\epsilon^{len}$ [m]
$\mathbf{R}_k = \mathbf{0}$	<b>35.4</b>	0.85	38.4	41.0
$\mathbf{R}_k = \mathbf{R}_1$	<b>35.4</b>	0.85	37.5	39.6
$\mathbf{R}_k = \mathbf{R}_2$	<b>35.4</b>	0.84	23.5	26.8
$\mathbf{R}_k = \mathbf{R}_3$	<b>35.4</b>	<b>0.80</b>	18.0	21.6
CM-ETT	<b>35.4</b>	0.84	<b>11.8</b>	<b>14.3</b>

approaches, see Figs. 10(a) and (b). For the error in position, a small displacement between the AIS reports and the estimations provided by the presented algorithms can be pointed out. This is due to the fact that the AIS reports the position of the AIS transponder, while the algorithms estimate the position of the center of the target, which generally speaking, can differ from the AIS transponder's position.

A further test case on the *Portovenere* dataset is also detailed in Fig. 11. This dataset is composed of 130 frames. The AIS track is depicted in Fig. 2. Advantages in the size estimations for the proposed model can be easily pointed out, see Figs. 11(c) and (d). Again, no gain can be seen in the estimation of the kinematic parameters, see Figs. 11(a) and (b). In this case, due to the non-radial track direction, these benefits can be appreciated on both length and width size estimations.

To further corroborate the validity of the proposed approach in providing an improved method of estimat-

ing the targets' size and to have a more significant statistical analysis, 8 further real test cases have been performed (the total amount of frames analyzed by the presented algorithms is about  $10^3$ ). On behalf of brevity, the results are summarized in Tab. V, where the RMSEs in position  $\epsilon^{pos}$ , velocity  $\epsilon^{vel}$ , width  $\epsilon^{wid}$ , and length  $\epsilon^{len}$  averaged on all the datasets are shown. The Frobenius error is not available for the real test cases because of the lack of the targets' real orientation in the AIS information. Best results are in boldface. Due to the random initialization of the algorithms, the first frames are neglected to evaluate the RMSEs. No gain can be pointed out for the kinematic parameters' estimation and the outcomes can be considered good in the light of the above-mentioned considerations with regard to the AIS information. The advantages are clear and the errors in both width and length are significantly reduced by properly considering the radar's measurement noise. The RMSEs in width are 11.8 m for the CM-ETT algorithm, 18.0 m for the  $\mathbf{R}_3$  method, and 38.4 m in the case of the  $\mathbf{R}_k = \mathbf{0}$  approach. Whereas, the RMSEs in length are 14.3 m, 21.6 m, and 41.0 m, respectively. A histogram representation of the absolute errors in width and length for the compared algorithms is depicted in Fig. 12. The results for  $\mathbf{R}_1$  and  $\mathbf{R}_2$  are worse than the results for  $\mathbf{R}_3$ , therefore we only compare the proposed CM-ETT filter to  $\mathbf{R}_3$  and  $\mathbf{R}_k = \mathbf{0}$ . Same conclusions as in Tab. V can be drawn.

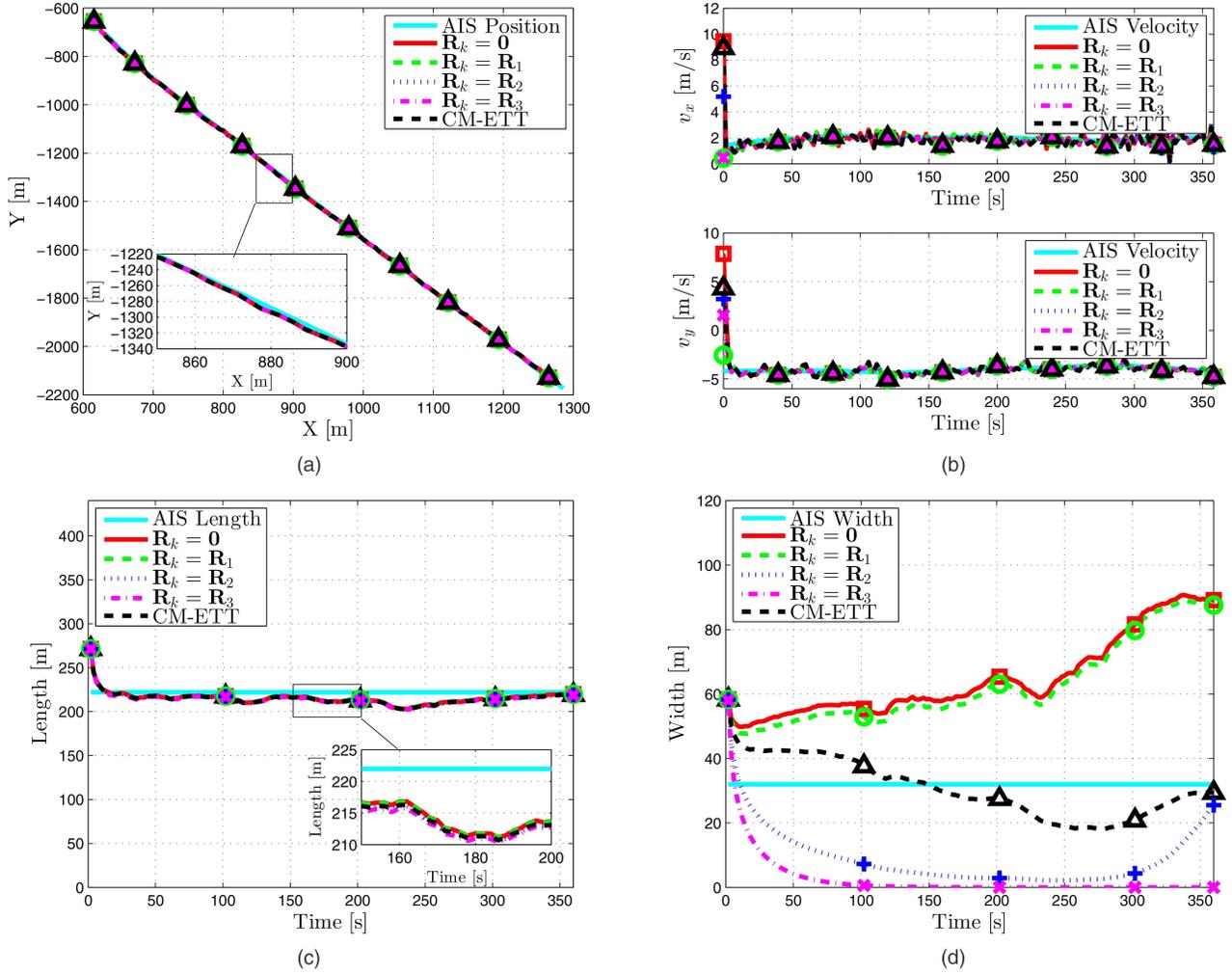


Fig. 10. (a) Position, (b) velocity, (c) length, and (d) width estimations for the compared approaches on the *Grand Holiday* dataset.

It is evident that by properly modeling the polar measurement noise the errors in both width and length are significantly reduced. For the data used here, the average gains of the proposed CM-ETT approach are 70% compared to the  $\mathbf{R}_k = \mathbf{0}$  approach and 35% with respect to  $\mathbf{R}_3$  for the targets' width estimation accuracy, while advantages of 65% and 34%, respectively, can be observed for the targets' length estimation accuracy.

## VI. MULTI-TARGET EXPERIMENTAL RESULTS

In this section the experimental results for the case of multiple extended targets in a cluttered environment are shown. The performance metrics used to assess the quality of the approach are presented first. Afterwards, the experimental results on a real dataset acquired by the X-band marine radar located in the Gulf of La Spezia, Italy, are described.

### A. Performance Metrics

This subsection is devoted to the description of the performance metrics, already introduced in [31], suitable for performance assessment in a multiple target and cluttered environment. They are briefly listed, below:

- The **time-on-target (ToT)** is defined as the ratio between the time during which the tracker follows the target and the whole time duration of the true target trajectory. Its ideal value is 1.
- The **false alarm rate (FAR)** is defined as the number of false track contacts normalized with the recording interval and the area of the surveyed region. Its ideal value is 0 that indicates no false alarm.
- The **track fragmentation ( $N^{TF}$ )** is calculated by summing the number of radar tracks associated with a unique AIS track. It provides a measurement of the track fragmentation (TF). The ideal value is 1.
- The **tracker accuracy (TA)** is evaluated using the errors in position ( $\epsilon^{pos}$ ), velocity ( $\epsilon^{vel}$ ), length ( $\epsilon^{len}$ ), and width ( $\epsilon^{wid}$ ). Average values along frames are provided as overall indexes. The ideal values are 0.

### B. Experimental Results

The assessment is conducted on real data acquired by the X-band marine radar. The main parameters used for the GGIW-PHD approach are summarized in Tab. VI. The dataset consists of 260 frames. AIS data

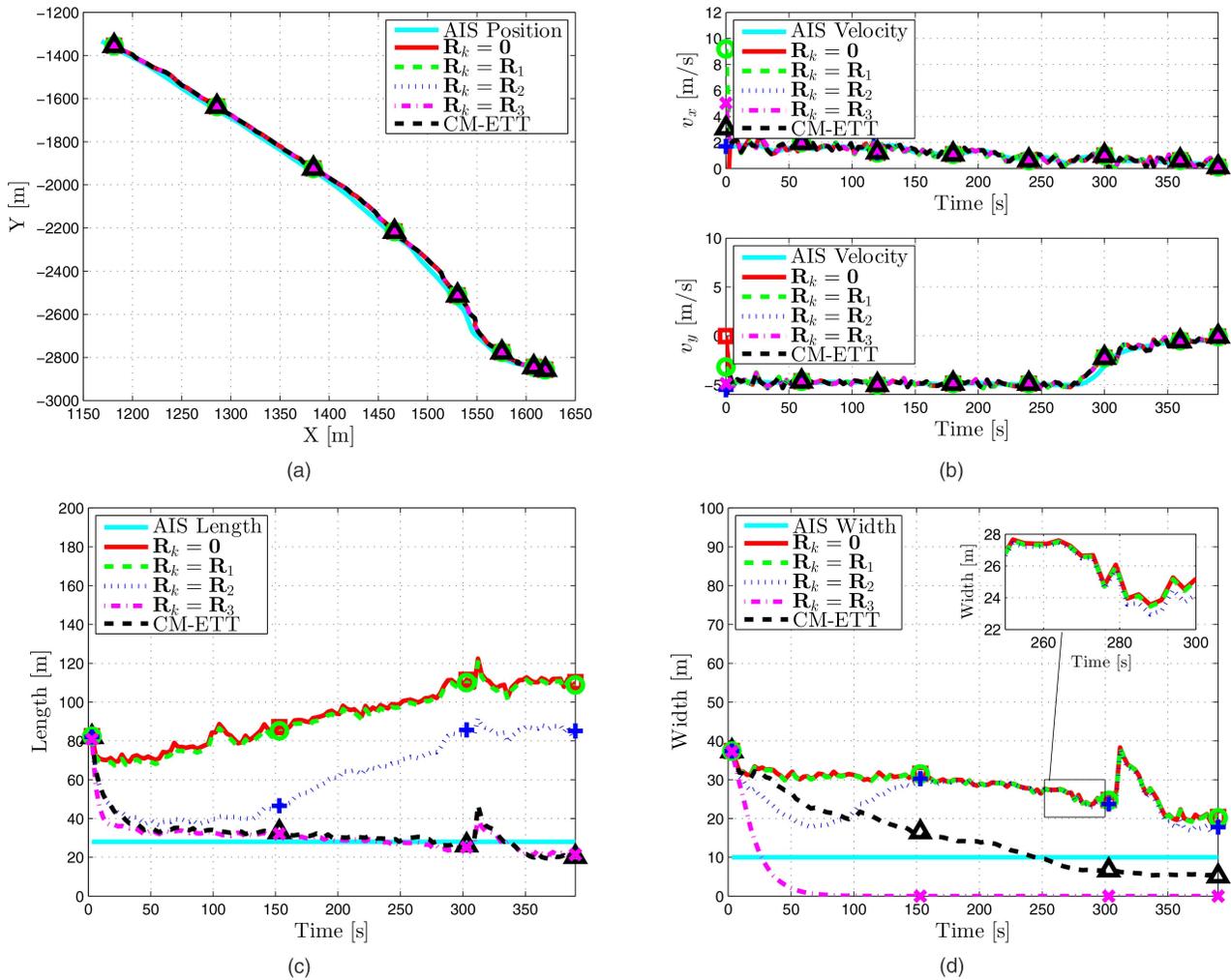


Fig. 11. (a) Position, (b) velocity, (c) length, and (d) width estimations for the compared approaches on the *Portovenere* dataset.

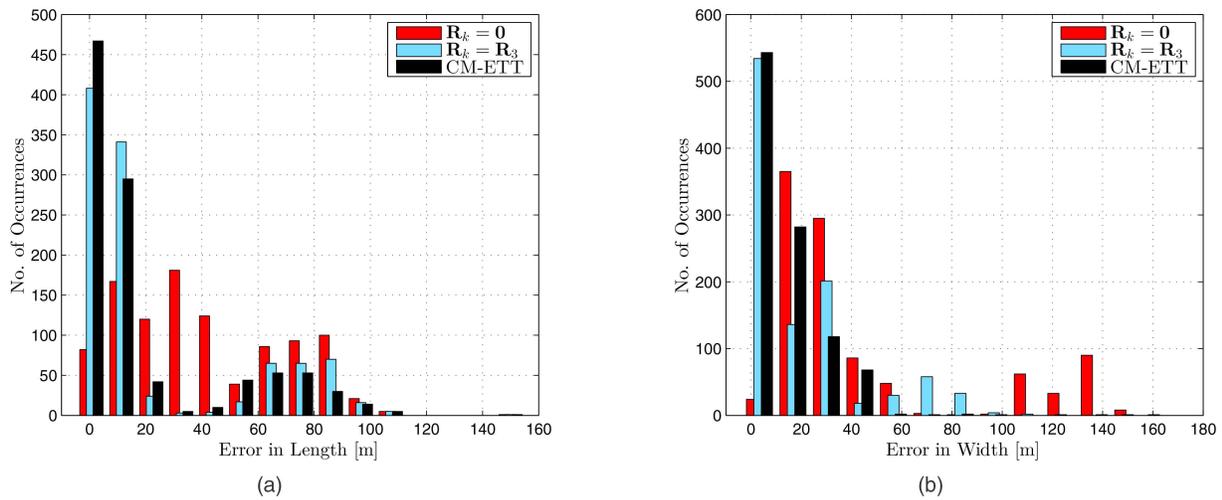


Fig. 12. Histograms of absolute errors in (a) length and (b) width for the  $R_k = 0$ , the  $R_k = R_3$ , and the CM-ETT approach calculated on all the datasets.

are exploited as ground-truth. An example of the outcomes of the GGIW tracker with converted measurements model in the case of three very closely spaced targets is shown in Fig. 13. The tracking results are de-

picted in Fig. 14. They show the estimations for both kinematic and size parameters for all the targets in the scenario. Solid lines denote the values provided by the AIS, while dashed lines represent the estimations pro-

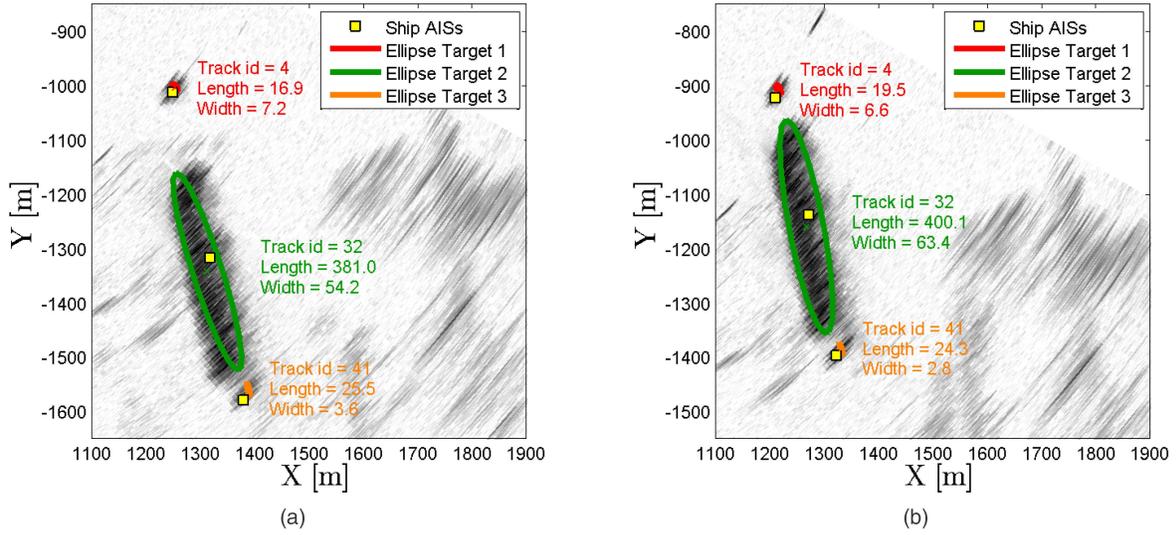


Fig. 13. Tracking of three spatially close targets that are part of the multiple extended target data set: (a) Frame 70; (b) Frame 90. AIS contacts are depicted with yellow square markers. The outcomes of the GGIW tracker with converted measurements model are indicated with colored ellipses.

TABLE VI  
Parameter Settings GGIW-PHD

Parameter	Value	Specification
$T_s$	2 s	Sampling time
$\eta_k$	1.04	Forgetting factor
$\sigma_a, \sigma_w$	$0.1 \text{ ms}^{-2}, \frac{0.1\pi}{180T_s} \text{ ms}^{-2}$	Kinematic noises
$n$	100	Extension uncertainty
$\tau$	99	Temporal decay
$\rho$	1/4	Scaling parameter
$\sigma_r, \sigma_\theta$	$\Delta r/2, \Delta\theta/2$ (see Tab. I)	Measurement noises
$\beta_{FA}$	$100/V(A) \text{ m}^{-2}$	Clutter density
$P_D$	0.99	Detection probability
$P_S$	0.99	Survival probability
$w_k^{(b)}$	$10^{-2}$	Birth weight
$m_k^{(b)}$	$\mathbf{0}_{5 \times 1}$	Birth mean
$P_k^{(b)}$	$\text{diag} \left( 1, 1, 1, 1, \frac{0.01\pi}{180T_s} \right)$	Birth covariance
$\alpha_k^{(b)}, \beta_k^{(b)}$	0.04, 0.008	Birth rate
$v_k^{(b)}, V_k^{(b)}$	120, $0.01\mathbf{I}_d$	Birth extension
$\bar{w}_0$	0.5	Extraction threshold
$T$	$10^{-3}$	Pruning threshold
$U$	25	Merging threshold
$\bar{w}_1, \bar{w}_2, \bar{w}_3$	1.1, 1, 0.8	Weight thresholds

vided by the GGIW tracker. The proposed approach reaches overall good performance. More specifically, we can appreciate only a small displacement between AIS information and the tracker's position estimation due to the fact that our approach estimates the center of the ellipse that represents the target (i.e. the ship), while the AIS returns the position of the transponder located on-board (usually not the ellipse's center). A further remark is related to the size estimation. Thanks to the usage of the proposed model, the tracker is able to compensate the usual bias in the size estimation with respect to the AIS values mainly due to the non-idealities of the

acquisition system (i.e. the width of the radar antenna pattern's main lobe).

Regarding to the performance metrics, the TA indexes confirm the previous analysis. Indeed, Tab. VII shows limited errors. Average errors are 30.7 m and  $0.69 \text{ ms}^{-1}$  that are due to the discrepancy between the information provided by the radar and the one that the AIS is able to provide. Average errors for the size estimation are 19.7 m in length and 6.9 m in width and can be considered limited for the ships under test (we have ships hundreds meter long and an obscuration phenomena that affects the size estimation of the ship with maritime mobile service identity (MMSI) equal to 351361000 and increases the  $\epsilon^{len}$  and the  $\epsilon^{wid}$ ).

The ToT is always very high (except for the ship with MMSI = 247031200, which is on the border of the surveillance area and could be not properly detected for some frames and the ship with MMSI = 247222500 that is a pilot boat and is often shadowed by or merged with the container ship with MMSI = 351361000). The overall ToT is 85%, see Tab. VII. Furthermore, the TF is almost ideal (with average values equal to 1.20). The small reduction of this index is only due to the obscuration phenomena for the container ship with MMSI = 351361000. Indeed, for about 40 frames, the passengers ship with MMSI equal to 255803790 interposed between the container ship and the radar causing a fragmentation of the container ship track for few frames and a reduction of the size estimation for that period. Finally, the FAR index is about  $6.7 \cdot 10^{-7} \text{ s}^{-1} \text{ m}^{-2}$ . Furthermore, it is worth pointing out that the most of the false alarms are due to signal leakages in the electronics, buoys, and ghosts (coming from the radar antenna pattern's secondary lobes).

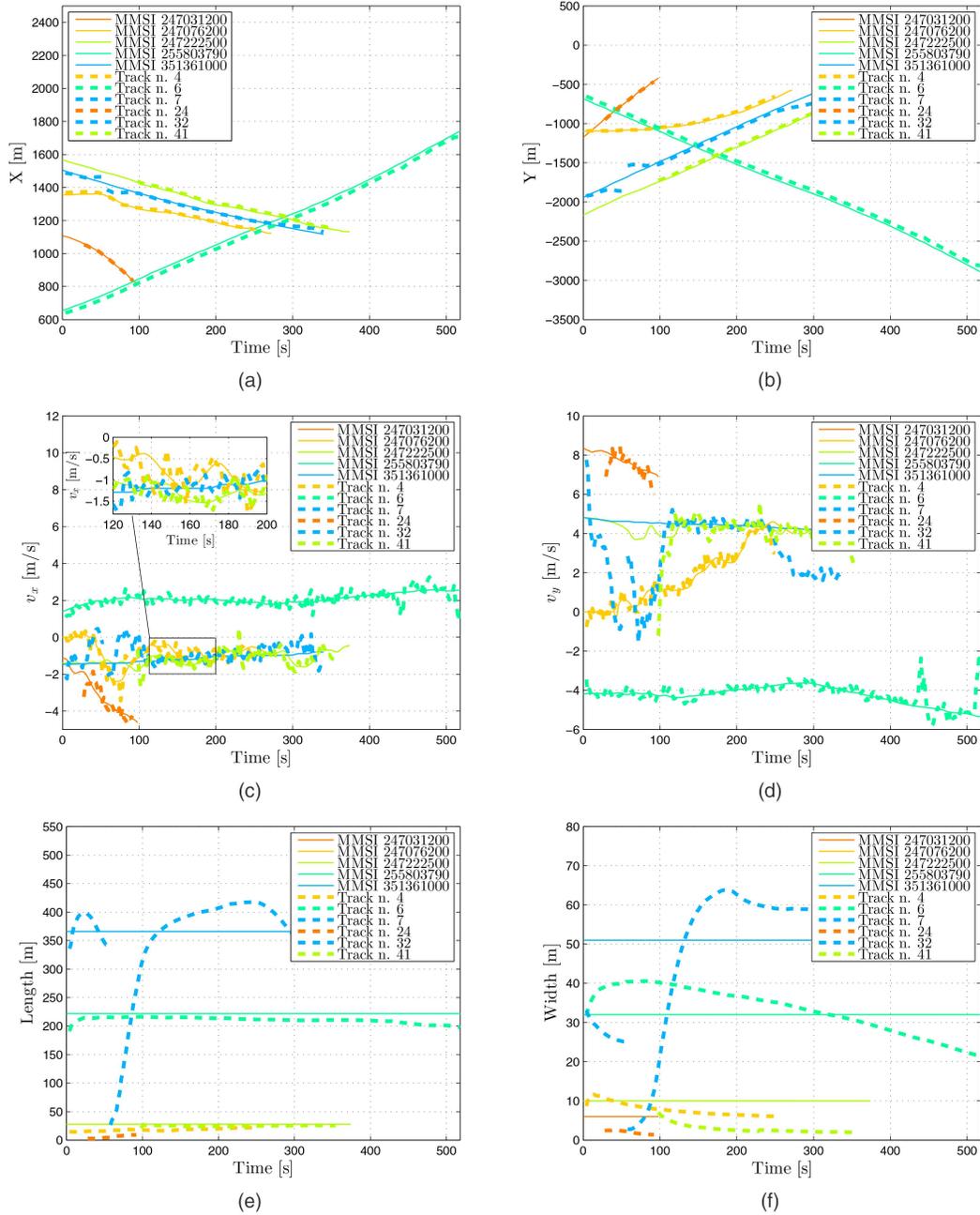


Fig. 14. Dashed lines represent tracks estimated by the GGIW tracker, while solid lines are used to depict AIS contacts. The association between GGIW estimated tracks and AIS tracks is indicated with different colors.

TABLE VII  
Tracking Metrics

MMSI Ship	247031200	247076200	247222500	255803790	351361000	Average Results
$\epsilon^{pos}$ [m]	5.05	14.3	20.1	51.6	62.7	30.7
$\epsilon^{vel}$ [ $\text{ms}^{-1}$ ]	0.52	0.41	0.49	0.33	1.70	0.69
$\epsilon^{len}$ [m]	22.2	9.8	2.9	11.2	52.4	19.7
$\epsilon^{wid}$ [m]	4.0	2.4	7.2	5.2	15.8	6.9
ToT	0.66	0.91	0.69	0.99	0.98	0.85
$N^{TF}$	1.00	1.00	1.00	1.00	2.00	1.20
FAR [ $\text{m}^{-2}\text{s}^{-1}$ ]	$6.7 \cdot 10^{-7}$					

## VII. CONCLUSIONS AND FUTURE DEVELOPMENTS

A maritime-surveillance system based on an X-band marine radar has been presented. The ability to estimate the targets' positions, velocities, and sizes using extended target tracking approaches relying upon random matrices has been evaluated. Standard and unbiased models have been proposed to properly take into account the radar's measurement noise. The validation has been conducted on both simulated and real data acquired by an X-band marine radar node installed in the Gulf of La Spezia, Italy. The integration of the proposed model into the gamma Gaussian inverse Wishart probability hypothesis density tracker has also been provided and tested on real data. Although comparable performance on the estimation of kinematic parameters has been pointed out, the experimental analysis confirms the ability of the proposed approach to better estimate the targets' size with respect to the approaches in [25] and [12], thanks to the use of proper models of the radar's measurement noise. Average gains up to 70% for the targets' width estimation accuracy and 65% for the length are observed on real data.

Future developments are devoted to the integration of the converted measurements model in other extended target tracking frameworks, such as the random hypersurface one, enabling a fair comparison among the different frameworks to track extended targets using polar/Cartesian coordinate conversion.

## REFERENCES

- [1] *Safety of Life at Sea (SOLAS) Convention*. Chapter V, Regulation 19.
- [2] D. Angelova and L. Mihaylova  
Extended object tracking using Monte Carlo methods,  
*IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 825–832, Feb. 2008.
- [3] Y. Bar-Shalom, P. Willett, and X. Tian  
*Tracking and Data Fusion: A Handbook of Algorithms*.  
Storrs, CT: YBS Publishing, Apr. 2011.
- [4] M. Baum, M. Feldmann, D. Fränken, U. D. Hanebeck, and J. W. Koch  
Extended object and group tracking: A comparison of random matrices and random hypersurface models,  
in *Proceedings of IEEE International Society of Information Fusion 7th Workshop Sensor Data Fusion*, Leipzig, Germany, Oct. 2010.
- [5] M. Baum and U. D. Hanebeck  
Extended object tracking with random hypersurface models,  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 149–159, Jan. 2014.
- [6] M. Baum and U. Hanebeck  
Shape tracking of extended objects and group targets with star-convex RHMs,  
in *14th International Conference on Information Fusion (FUSION)*, Jul. 2011.
- [7] S. S. Blackman  
Multiple hypothesis tracking for multiple target tracking,  
*IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, Jan. 2004.
- [8] Y. Boers, H. Driessen, J. Torstensson, M. Trieb, R. Karlsson, and F. Gustafsson  
Track before Detect algorithm for tracking extended targets,  
*IEEE Proceedings Radar, Sonar and Navigation*, vol. 153, no. 4, pp. 345–351, Aug. 2006.
- [9] L. Bruno, P. Braca, J. Horstmann, and M. Vespe  
Experimental evaluation of the range-Doppler coupling on HF surface wave radars,  
*IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 850–854, Jul. 2013.
- [10] M. Bruno  
Bayesian methods for multiaspect target tracking in image sequences,  
*IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1848–1861, Jul. 2004.
- [11] B. Errasti-Alcala and P. Braca  
Track before Detect algorithm for tracking extended targets applied to real-world data of X-band marine radar,  
in *17th International Conference on Information Fusion (FUSION)*, Salamanca, Spain, Jul. 2014.
- [12] M. Feldmann, D. Franken, and J. W. Koch  
Tracking of extended objects and group targets using random matrices,  
*IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, Apr. 2011.
- [13] K. Gilholm, S. Godsill, S. Maskell, and D. Salmond  
Poisson models for extended target and group tracking,  
in *Proceedings SPIE, Signal and Data Processing of Small Targets*, San Diego, CA, Aug. 2005, pp. 230–241.
- [14] K. Gilholm and D. Salmond  
Spatial distribution model for tracking extended objects,  
*IEEE Proceedings Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 364–371, Oct. 2005.
- [15] K. Granström, C. Lundquist, F. Gustafsson, and U. Orguner  
Random set methods: Estimation of multiple extended objects,  
*IEEE Robotics and Automation Magazine*, vol. 21, no. 2, pp. 73–82, Jun. 2014.
- [16] K. Granström, A. Natale, P. Braca, G. Ludeno, and F. Serafino  
PHD extended target tracking using an incoherent X-band radar: Preliminary real-world experimental results,  
in *17th International Conference on Information Fusion (FUSION)*, Salamanca, Spain, Jul. 2014.
- [17] ———  
Gamma Gaussian inverse Wishart probability hypothesis density for extended target tracking using X-band marine radar data,  
*IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 12, pp. 6617–6631, Jul. 2015.
- [18] K. Granström and U. Orguner  
Estimation and maintenance of measurement rates for multiple extended target tracking,  
in *15th International Conference on Information Fusion (FUSION)*, Singapore, Jul. 2012.
- [19] ———  
A PHD filter for tracking multiple extended targets using random matrices,  
*IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5657–5671, Nov. 2012.
- [20] K. Granström, C. Lundquist, and O. Orguner  
Extended target tracking using a Gaussian-mixture PHD filter,  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.
- [21] K. Granström and O. Orguner  
New prediction for extended targets with random matrices,  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1577–1589, Apr. 2014.

- [22] S. Grosdidier, A. Baussard, and A. Khenchaf  
HFSW radar model: Simulation and measurement,  
*IEEE Transactions on Geoscience and Remote Sensing*,  
vol. 48, no. 9, pp. 3539–3549, Sep. 2010.
- [23] A. K. Jain, M. N. Murty, and P. J. Flynn  
Data clustering: a review,  
*Association for Computing Machinery Computing Surveys*,  
vol. 31, no. 3, pp. 264–323, Sep. 1999.
- [24] T. Kirubarajan, Y. Bar-Shalom, and K. R. Pattipati  
Multiassignment for tracking a large number of overlapping  
objects,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 37, no. 1, pp. 2–21, Jan. 2001.
- [25] J. W. Koch  
Bayesian approach to extended object and cluster tracking  
using random matrices,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 44, no. 3, pp. 1042–1059, Apr. 2008.
- [26] J. W. Koch and M. Feldmann  
Cluster tracking under kinematical constraints using ran-  
dom matrices,  
*Robotics and Autonomous Systems*, vol. 57, no. 3, pp. 296–  
309, Mar. 2009.
- [27] J. Lan and X. R. Li  
Tracking of maneuvering non-ellipsoidal extended object  
or target group using random matrix,  
*IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp.  
2450–2463, May 2014.
- [28] R. Mahler  
Multitarget Bayes filtering via first-order multitarget mo-  
ments,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 39, no. 4, pp. 1152–1178, Jan. 2003.
- [29] ———  
*Statistical Multisource-Multitarget Information Fusion*.  
Artech House, 2007.
- [30] ———  
PHD filters for nonstandard targets, I: Extended targets,  
in *12th International Conference on Information Fusion (FU-  
SION)*, Seattle, WA, Jul. 2009, pp. 915–921.
- [31] S. Maresca, P. Braca, J. Horstmann, and R. Grasso  
Maritime surveillance using multiple high-frequency  
surface-wave radars,  
*IEEE Transactions on Geoscience and Remote Sensing*,  
vol. 52, no. 8, pp. 5056–5071, Aug. 2014.
- [32] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang,  
and S. Godsill  
Overview of Bayesian sequential Monte Carlo methods for  
group and extended object tracking,  
*Digital Signal Processing*, vol. 25, pp. 1–16, Feb. 2014.
- [33] U. Orguner  
A variational measurement update for extended target track-  
ing with random matrices,  
*IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp.  
3827–3834, Jul. 2012.
- [34] S. K. Pang, J. Li, and S. Godsill  
Detection and tracking of coordinated groups,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 47, no. 1, pp. 472–502, Jan. 2011.
- [35] S. Reuter and K. Dietmayer  
Pedestrian tracking using random finite sets,  
in *14th International Conference on Information Fusion (FU-  
SION)*, Jul. 2011.
- [36] H. Rohling  
Radar CFAR thresholding in clutter and multiple target  
situations,  
*IEEE Transactions on Aerospace and Electronic Systems*, vol.  
AES-19, no. 4, pp. 608–621, Jul. 1983.
- [37] K. Romeo, P. Willett, and Y. Bar-Shalom  
Particle filter tracking for banana and contact lens prob-  
lems,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 51, no. 2, pp. 1098–1110, Apr. 2015.
- [38] M. I. Skolnik  
*Radar handbook*.  
McGraw-Hill, Incorporated, 1970.
- [39] L. Sun, X. R. Li, and L. J.  
Modeling of extended objects based on support functions  
and extended Gaussian images for target tracking,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 50, no. 4, pp. 3021–3035, Oct. 2014.
- [40] G. Vivone, P. Braca, K. Granström, A. Natale, and J. Chanus-  
sot  
Converted measurements random matrix approach to ex-  
tended target tracking using X-band marine radar data,  
in *18th International Conference on Information Fusion (FU-  
SION)*, 2015.
- [41] G. Vivone, P. Braca, and J. Horstmann  
Knowledge-based multi-target ship tracking for HF surface  
wave radar systems,  
*IEEE Transactions on Geoscience and Remote Sensing*,  
vol. 53, no. 7, pp. 3931–3949, Jul. 2015.
- [42] B.-N. Vo and W.-K. Ma  
The Gaussian mixture probability hypothesis density filter,  
*IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp.  
4091–4104, Nov. 2006.
- [43] B.-N. Vo, S. Singh, and A. Doucet  
Sequential Monte Carlo methods for multitarget filtering  
with random finite sets,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 41, no. 4, pp. 1224–1245, Oct. 2005.
- [44] D. Wehner  
*High resolution radar*.  
Artech House, 1987.
- [45] M. Wieneke and J. W. Koch  
Probabilistic tracking of multiple extended targets using  
random matrices,  
in *Proceedings SPIE, Signal and Data Processing of Small  
Targets*, Orlando, FL, Apr. 2010.
- [46] ———  
A PMHT approach for extended objects and object groups,  
*IEEE Transactions on Aerospace and Electronic Systems*,  
vol. 48, no. 3, pp. 2349–2370, Jul. 2012.



**Gemine Vivone** received the B.Sc. (summa cum laude), the M.Sc. (summa cum laude), and the Ph.D. (highest rank) degrees in information engineering from the University of Salerno, Salerno, Italy, in 2008, 2011, and 2014, respectively. He is currently a Scientist at the North Atlantic Treaty Organization (NATO) Science & Technology Organization (STO) Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy. In 2014, he joined the NATO STO CMRE, La Spezia, Italy as a Research Fellow. In 2013, he was as a Visiting Scholar with Grenoble Institute of Technology (INPG), Grenoble, France, conducting his research at the Laboratoire Grenoblois de l'Image, de la Parole, du Signal et de l'Automatique GIPSA-Lab. In 2012, he was a Visiting Researcher with the NATO Undersea Research Centre, La Spezia, Italy. His main research interests focus on statistical signal processing, detection of remotely sensed images, data fusion, and tracking algorithms. Dr. Vivone serves as a Referee for several journals, such as *IEEE Transactions on Geoscience and Remote Sensing*, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, and *IEEE Geoscience and Remote Sensing Letters*. Dr. Vivone was the recipient of the Symposium Best Paper Award at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015.

**Paolo Braca** received the Laurea degree (summa cum laude) in electronic engineering, and the Ph.D. degree (highest rank) in information engineering from the University of Salerno, Italy, in 2006 and 2010, respectively. In 2009, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, USA. In 2010–2011, he was a Postdoctoral Associate with the University of Salerno, Italy.

In 2011, he joined the NATO Science & Technology Organization Centre for Maritime Research and Experimentation (CMRE) as a Scientist with the Research Department.

Dr. Braca conducts research in the general area of statistical signal processing with emphasis on detection and estimation theory, wireless sensor network, multi-agent algorithms, target tracking and data fusion, adaptation and learning over graphs, distributed radar (sonar) processing.

He is coauthor of more than 70 publications in international scientific journals and conference proceedings. Dr. Braca serves as an Associate Editor of *IEEE Transactions on Signal Processing*, *IEEE Transactions on Aerospace and Electronic Systems*, *IEEE Signal Processing Magazine* (E-Newsletter), *ISIF Journal of Advances in Information Fusion*, and *EURASIP Journal on Advances in Signal Processing*. He is in the Technical Committee of the major international conferences in the field of signal processing and data fusion. He was the recipient of the Best Student Paper Award (first runner-up) at the 12th International Conference on Information Fusion in 2009.



**Karl Granström** (M'08) is a postdoctoral research fellow at the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden. He received the M.Sc. degree in Applied Physics and Electrical Engineering in May 2008, and the Ph.D. degree in Automatic Control in November 2012, both from Linköping University, Sweden. He previously held postdoctoral positions at the Department of Electrical and Computer Engineering at University of Connecticut, USA, from September 2014 to August 2015, and at the Department of Electrical Engineering of Linköping University from December 2012 to August 2014. His research interests include estimation theory, multiple model estimation, sensor fusion and target tracking, especially for extended targets. He has received paper awards at the Fusion 2011 and Fusion 2012 conferences.





**Antonio Natale** was born in Naples, Italy, on July 3, 1982. He received the B.S. and M.S. degrees (both cum laude) in telecommunication engineering and the Ph.D. degree from the University of Naples Federico II, Naples, in 2005, 2008, and 2012, respectively. In 2011, he was a Visiting Scientist with the Surrey Space Centre, University of Surrey, Surrey, U.K. In that period, he was also the Principal Researcher for the project “Applications for S-band SAR,” which was funded by EADS Astrium Ltd. He is currently a Research Fellow with the Institute for Electromagnetic Sensing of the Environment (IREA) of the Italian National Research Council (CNR), Naples. In 2013, he spent a period at the NATO Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy, as a Visiting Scientist to develop target detection and tracking strategies from high-resolution radar data. His research interests include remote sensing, the modeling of electromagnetic scattering from natural surfaces, signal processing, and the estimation of parameters from radar data. Dr. Natale received the 2009 IEEE Geoscience and Remote Sensing South Italy Chapter Prize for the best Italian thesis in remote sensing discussed in 2008. Moreover, he was the recipient of the 2009 S. A. Schelkunoff Transactions Prize Paper Award from the IEEE Antennas and Propagation Society, for the best paper published in 2008 on the *IEEE Transactions on Antennas and Propagation*.



**Jocelyn Chanussot** (M’04–SM’04–F’12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from Savoie University, Annecy, France, in 1998. In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l’Armement. Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of Signal and Image Processing. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing. He has been a Visiting Scholar with Stanford University (USA), KTH (Sweden), and NUS (Singapore). Since 2013, he has been an Adjunct Professor with the University of Iceland. From 2015 to 2017, he is a Visiting Professor at the University of California, Los Angeles. He was the Founding President of the IEEE Geoscience and Remote Sensing French Chapter (2007–2010), which received the 2010 IEEE GRS-S Chapter Excellence Award. He was a co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award. He was a member of the IEEE Geoscience and Remote Sensing Society AdCom (2009–2010), in charge of membership development. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing. He was the Chair (2009–2011) and Co-Chair of the GRS Data Fusion Technical Committee (2005–2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006–2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing, (2009). He was an Associate Editor of the *IEEE Geoscience and Remote Sensing Letters* (2005–2007) and *Pattern Recognition* (2006–2008). Since 2007, he has been an Associate Editor of the *IEEE Transactions on Geoscience and Remote Sensing*. He was the Editor-in-Chief of the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2011–2015). In 2013, he was a Guest Editor of the *Proceedings of the IEEE* and the *IEEE Signal Processing Magazine* in 2014. He is a member of the Institut Universitaire de France (2012–2017).

# Priority-Based Tracking of Extended Objects

KEVIN WYFFELS  
MARK CAMPBELL

Inspired by human perception, a novel framework for dynamically allocating algorithmic and computational resources to achieve variable precision tracking of extended objects is presented. Probabilistic object relevancy metrics reflect the priority of each tracked object to the consumer of the tracking output, and are leveraged to trigger mode transitions in a hybrid system implementation of the proposed priority-based framework. In this way, the bulk of the algorithmic and computational resources are reserved for tracking objects of highest priority with high-precision methods, while low priority objects are tracked with inexpensive, qualitative methods. An example implementation of the proposed framework is provided for an autonomous driving application, in which the consumer of the tracking output is an anticipatory path planner. Simulation results demonstrate the ability of the framework to automatically trade computational complexity for tracking precision as a function of an object's priority to the tracking consumer.

Manuscript received October 31, 2015; revised February 18, 2016; released for publication May 10, 2016.

Refereeing of this contribution was handled by Karl Granstrom.

Authors' address: Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14850 USA (E-mail: {klw94, mc288}@cornell.edu).

---

1557-6418/17/\$17.00 © 2017 JAIF

## 1. INTRODUCTION

Humans consistently outperform robots in perceptual tasks despite certain hardware advantages favoring robots over humans; this suggests that human cognition is superior to analogous robotic algorithms in these areas, and potentially worth emulating. For instance, sensors providing metric information over a wide field-of-view (FOV) are readily available for robots, while human sensors provide ordinal information, at best, over a limited FOV. In fact, empirical studies have concluded that human vision provides ordinal information via a variety of visual cues, such as occlusion, binocular disparities, and motion parallax, which the human brain fuses into a single cohesive belief of the perceptual space [12], [13], [25], [27]. As distance from the observer decreases, the number, type, and quality of available ordinal cues increases, and the human belief quickly converges from an imprecise ordinal representation to a precise metrical one, despite purely ordinal sensor information.

As with many other biological phenomena, the characteristic of human perception described above serves as a complement to most human ventures, in that humans typically only require detailed, metrical representations of the nearby scene in which they are currently an active participant; therefore, anything more than a general ordinal awareness of objects at greater distances is, at the very least, a misallocation of limited cognitive resources, and a precursor for distraction.

Given that computers/robots with finite computational resources are often employed to perform human tasks such as navigation or driving, many of the requirements of human perception discussed above apply equally well to computer/robotic perception. However, analogs to the complementary human perception characteristics are largely absent from robotic algorithms. Therefore, inspired by human cognition, this work proposes a priority-based framework for allocating algorithmic and computational resources as a function of *priority* in extended object tracking (EOT); the automatic allocation of computational resources as a function of priority is a novel contribution to the EOT field.

Object tracking is a perception application in which the states (e.g. kinematics) of objects present in the local environment are estimated from sensor data. Extended objects are defined as objects of non-negligible size relative to the sensor resolution, such that they cannot be accurately modeled as a mathematical point. EOT differs from traditional object tracking in that it violates the foundational assumption that each object can return, at most, a single measurement per sensor query. Further, extended objects cast shadows in sensor data, known as occlusion shadows, which result in incomplete or missing measurements of the objects of interest—this includes self-occlusion, in which the object surface nearest the sensor occludes its remaining self. While the proposed priority-based framework is general enough to

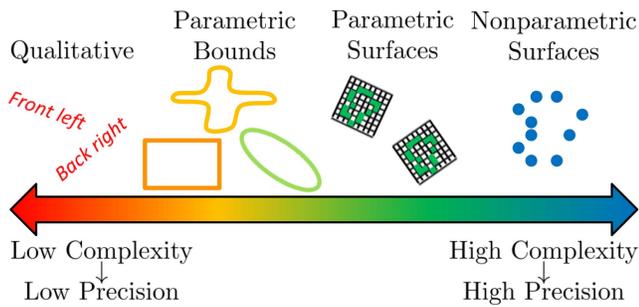


Fig. 1. Distilled spectrum of extent model complexity and tracking precision, which tend to be directly correlated.

consider any EOT methods, the authors are particularly interested in the general case in which the extended object size and shape is unknown a priori. Therefore, tracking approaches that rely on this information, such as spatial distribution models [14], are not included in the following discussion of prior work.

Various successful approaches to extended object tracking in the absence of a priori shape information have been proposed, spanning a spectrum of computational complexity and tracking precision. Two central and related factors determining an EOT algorithm's position on this spectrum, are the detail and accuracy of the object shape/extent model. Specifically, detailed and accurate knowledge of object shape/extent enables high fidelity sensor models that offer a precise and detailed interpretation of the sensor data, and its relationship to the object state. In this way, detailed and accurate extent models engender high precision tracking, generally at the expense of increased computational complexity. Fig. 1 depicts a generalization of this trade-off, patently distilled for the following discussion of prior work.

At the low complexity end of the spectrum, qualitative/topological object tracking approaches exist, where the object state definition itself is imprecise by nature. For example, the state could be defined as a single discrete random variable representing the region of space, or topological node, that a dynamic object occupies over time. Qualitative state representations have gained interest in robotic applications due to their efficiency, scalability, and natural synergy with inexpensive ordinal sensor information, such as that provided by monocular cameras or human input [28]. In a conceptually related approach, traditional static occupancy grid mapping concepts have been extended to characterize dynamic scenes [2], [32]. These approaches tend to be extremely efficient, but their qualitative/topological state representation is too imprecise for many applications, such as those requiring agents to safely interact with other dynamic objects.

Moving along the spectrum toward higher precision, simple parametric object shape models are prescribed a priori, and the parameters of the model are jointly estimated as states in the object tracker. Common simple parametric shapes include circular discs [6], ellipses [5], [7], [10], [23], [24], and rectangles [8], [26], [31]. These

simple shape models generally represent a tight enclosing bound or circumscription of the true, more complex, underlying shape, rather than the shape itself; therefore, they are sufficient for tracking a variety of arbitrarily shaped objects without a priori knowledge of the object shape or size. However, the inherent, uncharacterized mismatch between the true underlying object shape and the simple prescribed circumscription restrict the sensor model fidelity, thereby degrading tracking precision. This degradation is mitigated somewhat by increasing the complexity and flexibility of the shape bound, which is the goal of star convex random hypersurface models (RHM) [9]. These models are appropriate when detailed a priori information about object shape and size is unavailable, and computational efficiency is at a premium.

A simple and intuitive method for tracking arbitrarily shaped extended objects involves the use of occupancy grids anchored to an object-centric coordinate frame, dubbed Object Local Grid Maps (OLG) [3]. OLG shape models can be rigorous and flexible, however, the precision and complexity depend on an appropriate choice of grid extent and resolution, which requires some a priori knowledge of the size and shape complexity of objects to be tracked.

Finally, at the high precision end of the spectrum, very detailed, non-parametric point cloud models are employed [18]–[20], [29], [30], [34]–[37]. These models are extremely flexible in providing rich, 2 or 3-dimensional (3D) renderings of the true underlying object surfaces for arbitrary shapes, and often do not require a priori information about the object shape and size. These detailed surface representations enable high precision sensor models, which, in turn, contribute to high precision kinematic state estimates; all at the expense of high computational complexity. Therefore, these methods are appropriate when tracking precision is at a premium, a priori information about object shape and size is unavailable, and computational resources are abundant.

Akin to human perception, within a given EOT application the *relevance* of each object to the consumer of the tracking output, henceforth referred to as *the consumer*, may vary from object-to-object or from instant-to-instant. For example, in a surveillance application, objects exhibiting anomalous behavior may be more relevant than those exhibiting benign behavior; in navigation, nearby objects may be more relevant than distant objects; or in a pursuit application, the lead object may be more relevant than other followers. In these cases, the EOT requirements may also vary with object relevance, deeming a single appropriate EOT method difficult to identify.

This work addresses this issue by proposing a *priority*-based tracking framework for extended objects, where *priority* refers to the object's relevance to the consumer. The proposed framework is implemented via a hybrid system model in which each discrete mode

represents a different EOT method with unique characteristics on the complexity-precision spectrum of Fig. 1. Further, probabilistic object relevancy metrics are designed to reflect the time-varying priority of each object to the consumer, and leveraged to inform the hybrid system switching strategy. In this way, objects most relevant to the consumer are allocated more resources and tracked with higher precision, while objects of peripheral relevance are efficiently accounted for with minimal computational burden. To demonstrate the ability of the framework to prioritize objects by automatically trading computation for tracking precision, an example implementation of the hybrid framework is provided for an autonomous driving application in which the consumer is an anticipatory planner.

Section 2 formally defines the general object tracking problem, Section 3 introduces the proposed hybrid system implementation of the priority-based EOT framework, Sections 4 and 5 provide an example implementation of the hybrid framework for an autonomous driving application, Section 6 presents a discussion of simulation results, and finally 7 provides some concluding remarks.

## 2. OBJECT TRACKING PROBLEM FORMULATION

The goal of general multi-object tracking is to estimate the full latent object state history,  $X_{1:K}$ , of an unknown number of maneuvering objects,  $N_O$ , from a history of noisy observations,  $Z_{1:K}$ , without knowledge of object controls or intent. Probabilistic inference provides a rigorous means for accounting for the many sources of uncertainty in the problem, deeming it a valuable tool for multi-object tracking. Specifically, rather than estimating the latent variables directly, inference methods estimate the joint posterior probability distribution over the latent variables conditioned on the observations,

$$p(X_{1:K}^{1:N_O} | Z_{1:K}) = \prod_{n=1}^{N_O} p(X_{1:K}^n | Z_{1:K}, X_{1:K}^{1:n-1}) \quad (1)$$

from which optimal estimates of state trajectories,  $\hat{X}_{1:K}^{1:N_O}$ , can be computed via existing techniques, such as Minimum Mean Square Error (MMSE) or Maximum-a-Posteriori (MAP).

In most practical applications, it is accurate to model the objects as being mutually independent,  $p(X_{1:K}^n | X_{1:K}^m) = p(X_{1:K}^n) \forall n \neq m$ , which is convenient in that it simplifies the full joint multi-object tracking problem into the product of single-object marginals that can be studied independently in parallel:

$$p(X_{1:K}^{1:N_O} | Z_{1:K}) = \prod_{n=1}^{N_O} p(X_{1:K}^n | Z_{1:K}) \quad (2)$$

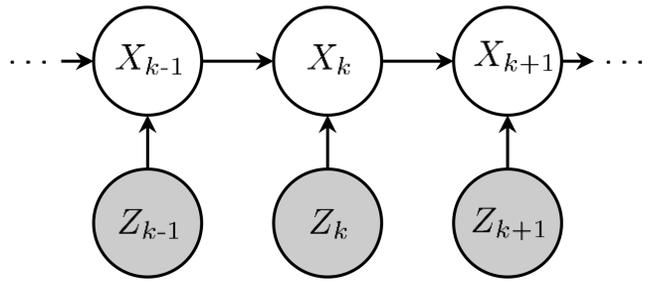


Fig. 2. Graphical representation of the Hidden Markov model (HMM) used to represent the single-object tracking problem. Shaded nodes denote observed variables, and unshaded nodes denote the hidden, i.e. latent, variables to be estimated.

Therefore, theoretical development in object tracking is commonly focused on the single-object tracking problem, i.e. estimating:

$$p(X_{1:K} | Z_{1:K}) \quad (3)$$

Further, computation of the posterior in (3) is made tractable, efficient, and deterministic via the following conditional independence assumptions, which gives rise to the Hidden Markov model (HMM) depicted graphically in Fig. 2:

$$\begin{aligned} p(Z_k | X_k, Z_\ell) &= p(Z_k | X_k) \quad \forall \ell \neq k \\ p(X_k | X_{k-1}, X_{k-\ell}) &= p(X_k | X_{k-1}) \quad \forall \ell > 1 \end{aligned} \quad (4)$$

Lastly, online tracking applications require state estimates in real-time as data is received, and are often principally concerned with the current object state, rather than the full time-history. Therefore, the inference problem is further simplified to estimating:

$$p(X_k | Z_{1:k}) \quad \forall k \in \{1, \dots, K\} \quad (5)$$

That is, the marginal distribution over the states at any time,  $k$ , is conditioned only on the observation history through  $k$ .

Given the HMM of Fig. 2, (5) can be computed recursively at each time step via the following two step process:

- 1) *Prediction step*: the posterior state belief at the previous time step,  $p(X_{k-1} | Z_{1:k-1})$ , is propagated forward in time via the prescribed stochastic object dynamics model represented by the transition density,  $p(X_{k-1}, X_k)$ . The result is the prior state belief at the current time step:  $p(X_k | Z_{1:k-1})$ .
- 2) *Update step*: the prior state belief at the current time step,  $p(X_k | Z_{1:k-1})$ , is updated with the current observation via the prescribed stochastic measurement model represented by the observation density,  $p(X_k, Z_k)$ . The result is the posterior state belief at the current time step:  $p(X_k | Z_{1:k})$ .

Estimating (5) is commonly referred to as *filtering*, which can be supplemented with *smoothing* to estimate the full posterior distribution over the state history in (3).

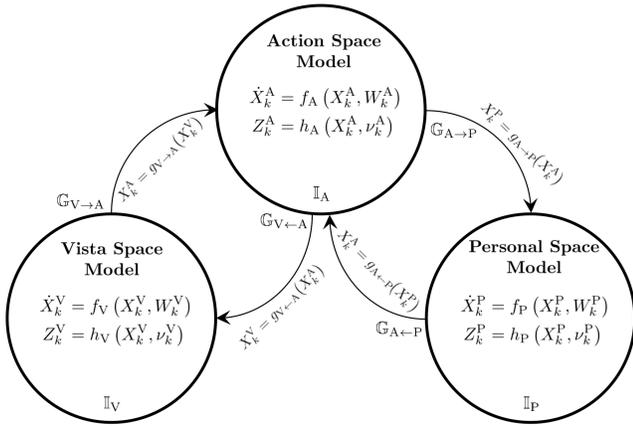


Fig. 3. Hybrid model of the proposed tracking framework, where  $\mathbb{G}$  and  $\mathbb{I}$  denote the guards and invariants governing the discrete mode transitions.

### 3. ESTIMATION FRAMEWORK

As discussed in the introduction, human visual perception degrades from a precise metrical representation to a rough ordinal one as distance from the observer increases; this is a direct consequence of the diminishing quality and availability of ordinal visual cues, and an evolutionary advantage given finite cognitive resources and the relative importance of close objects compared to distant ones. In characterizing this phenomenon, cognitive scientists have discretized perceptual space into three distinct regions defined by distance from the human observer. Specifically, in order of decreasing distance and improved convergence to a metrical representation: *Vista* space, *Action* space, and *Personal* space. In human trials, the distances to the boundaries dividing these regions were found to depend on a variety of variables, including the quality of the observer's vision, and characteristics of the particular scene, e.g. clutter, object familiarity, and scene geometry [13] [12].

Many natural analogies exist between human and robot/computer perception; both operate under resource constraints (cognition vs. computation), and both utilize sensor information that often degrades with distance from the observer, to name only two. Given these analogies, and the fact that computers/robots are often designed to perform human tasks, such as surveillance or navigation, a priority-based tracking framework inspired by the human perception concepts of *attention* and *focus* is proposed here, which automatically trades computational and algorithmic resources for tracking precision as a function of object relevance to the consumer of the EOT output.

Fig. 3 depicts the hybrid system model designed to implement the proposed priority-based EOT framework. Each discrete mode, Vista, Action, and Personal, represents a unique EOT approach chosen from the left, center, and right, respectively, of the complexity-precision spectrum depicted in Fig. 1. Further, probabilistic object relevancy metrics inform the mode switching strategy

such that, as an object becomes increasingly *relevant* to the consumer, the tracker transitions along the path: Vista  $\rightarrow$  Action  $\rightarrow$  Personal, causing the overall tracking framework to transition from left to right on the spectrum depicted in Fig. 1. In this way, objects most relevant to the consumer are allocated more resources and tracked with higher precision, while objects of peripheral relevance are efficiently accounted for with inexpensive EOT methods.

The parameters of the hybrid system model of Fig. 3, i.e. the modal EOT methods and object relevancy metrics, should be chosen to reflect the specific EOT application and goals motivating the use of the proposed priority-based framework. In this way, the *optimal* parameterization of the proposed priority-based framework in Fig. 3 is highly application and consumer dependent, and therefore beyond the scope of this work. However, for demonstration purposes, an example parameterization for an autonomous driving application is provided in the coming sections, coupled with some discussion of equally valid alternatives. For this example, the *consumer* of the tracker output is an anticipatory planning routine tasked with planning control inputs to *safely progress* the vehicle toward its destination. In this vein, the consumer defines object *priority* in terms of its potential contribution to the current plan. Specifically, while all objects in the local environment are considered when planning a future path, those that have potential to violate the planner's *safety* requirement are of the highest priority, i.e. those that pose an immediate risk of collision, followed by those with potential to violate the planner's *liveness* requirement, i.e. those that inhibit the ego-vehicle's progress toward the goal location.

While beyond the scope of this paper, the hybrid system framework depicted in Fig. 3 can also be outfit to address alternative tracking goals. For instance, consider the goal of achieving tracking robustness. An EOT approach robust to occlusion could be selected when driving in a cluttered environment, e.g. [1], [15], [34], [35], [37], while an alternative approach may prove prudent when the clutter subsides. In heavy traffic, cars could be tracked in groups rather than individually, e.g. [24], or EOT methods that account for the inherent correlations in the behavior of the traffic participants could be developed; i.e. by omitting the independence assumption leading to (2). Further, the model can be outfit to transition according to *ability*-based (or other) metrics, rather than object relevance. For instance, general object trackers, such as those discussed in the introduction, can be leveraged at object track initialization when specific static attributes of the object, such as object type or class, are unavailable; then, as estimates of the static object attributes converge, the system can transition to more specific, ad hoc, trackers designed to leverage information inferred from the estimated object attribute. Alternatively, a unique synergy may exist between EOT approaches and available sensor types; for instance, dense 3D colored point cloud approaches, e.g.

TABLE I  
Suggested modal tracking methods

Vista Space	Action Space	Personal Space
<i>Qualitative</i>	<i>Parametric Bound</i>	<i>Nonparametric Surface</i>
<ul style="list-style-type: none"> <li>• Temporal occupancy grid [2]</li> <li>• Markov chain occupancy grid [32]</li> <li>• Topological</li> </ul>	<ul style="list-style-type: none"> <li>• Circular Disk [6]</li> <li>• Ellipse [5], [7], [10], [23], [24]</li> <li>• Rectangle [8], [26], [31]</li> <li>• Star Convex RHM [9]</li> </ul>	<ul style="list-style-type: none"> <li>• 2D Point Cloud [29], [30], [34]–[37]</li> <li>• 3D Point Cloud [18]–[20]</li> <li>• 3D Surface Reconstruction e.g. KinectFusion [21]</li> </ul>

[18]–[20], perform well in regions of space where the field-of-view (fov) of a color camera intersects the fov of one (or more) lidar sensor(s). In cases such as this, transitions can be triggered as objects enter and exit the fov of different sensors, or areas of multi-sensor overlap, leveraging the identified synergistic sensor-tracker pairings.

#### 4. MODAL TRACKING APPROACHES

The focus of this work is to invoke high precision EOT methods for objects that are *relevant* to the consumer, and inexpensive EOT methods for objects that are of peripheral relevance. Therefore, the Vista, Action, and Personal space models are chosen from the left, center, and right of the tracking spectrum presented in Fig. 1, respectively. A partial list of existing EOT approaches appropriate for each mode is provided in Table I, and those chosen for the autonomous driving example are presented in detail in the coming sections.

Note that, while not a requirement of the priority-based EOT framework, all measurement models chosen for the autonomous driving example correspond to sensors providing (potentially multiple) position/distance returns per query, such as lidar, radar, or binocular/RGB-D cameras. Throughout the paper, an unadorned  $Z_k$  denotes the raw position measurement, or set of measurements, at time step  $k$ , while superscripted variables,  $Z_k^{V/A/P}$ , denote a particular interpretation of the raw data (e.g. metadata, or summary statistics) leveraged by the sensor model associated with the hybrid mode identified in the superscript.

##### 4.1. Vista Space Model

Vista mode is reserved for objects with the most peripheral significance to the consumer; a general awareness of objects in Vista space is useful, but the computational resources required for detailed object tracking are better spent elsewhere. Further, similar to human sensors, robot sensor precision/resolution often degrades with distance from the sensor (e.g. the spatial resolution of a spinning lidar); in these cases, the tracking precision also degrades with distance from the sensor regardless of the chosen tracking algorithm, and thus the benefits of ‘high precision’ methods are limited. To this

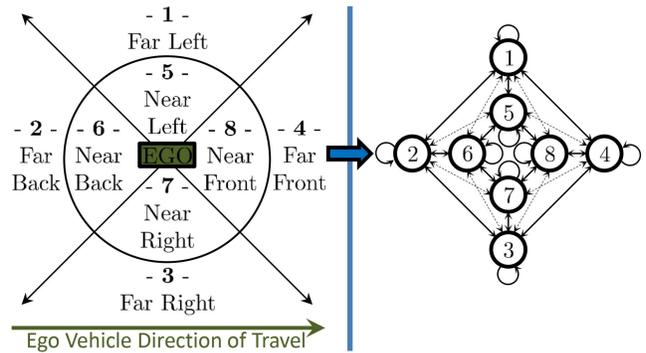


Fig. 4. Qualitative abstraction of the perceptual space of the ego robot. **Left:** Eight qualitative discrete states comprised of two range sets and four bearing quadrants. **Right:** Graphical model representation of the available transitions between the qualitative states of the discrete abstraction. Dashed arrows denote transitions enabled by the discrete time nature of the filter driven by the finite temporal resolution of the sensor.

end, the perceptual space surrounding the ego-vehicle is abstracted into the eight disjoint discrete qualitative states depicted in Fig. 4 (left), the topology of which is encoded in the state transition diagram in Fig. 4 (right). Each qualitative state,  $X_k^V$ , is parameterized by a bearing and a range interval,  $\mathbb{B}_{X_k^V}$  and  $\mathbb{R}_{X_k^V}$ , respectively, defined as:

$$\mathbb{B}_{X_k^V} = \frac{\pi}{2} \cdot \left\{ X_k^V + \left[ -\frac{1}{2}, \frac{1}{2} \right) \right\}$$

$$\mathbb{R}_{X_k^V} = \begin{cases} [0, \bar{\rho}) & \text{if } X_k^V \in \text{Near} \\ [\bar{\rho}, \infty) & \text{if } X_k^V \in \text{Far} \end{cases} \quad (6)$$

where  $\bar{\rho}$  is the user-defined range boundary between ‘Near’ and ‘Far’; i.e. the circle in Fig. 4 (left).

Qualitative state representations have gained interest in robotic/computer applications, such as relational mapping [28], due to their efficiency, scalability, and natural synergy with inexpensive ordinal sensor information, such as that provided by monocular cameras or human input. The qualitative states depicted in Fig. 4 are chosen because of their similarity to human accounting of objects in the Vista space of human perception; common robotic sensors provide this information directly (i.e. bearing and range), eliminating the need for intricate interpretations of the data, such as reasoning about object shapes and surfaces. The object state in Vista space,  $X_k^V$ , is then an integer denoting the qualitative state of the object at time  $k$ , the belief of which,  $p(X_k^V | Z_{1:k})$ , is distributed according to a categorical distribution.

##### 4.1.1. Belief Prediction:

Minding the conditional independence rules defined in (4), the posterior categorical distribution over the object state at time  $k - 1$  is predicted forward to time

$k$  as follows:

$$\begin{aligned} p(X_k^V | Z_{1:k-1}) &= \sum_{X_{k-1}^V} p(X_{k-1}^V, X_k^V | Z_{1:k-1}) \\ &= \sum_{X_{k-1}^V} p(X_k^V | X_{k-1}^V) p(X_{k-1}^V | Z_{1:k-1}) \end{aligned} \quad (7)$$

where  $p(X_{k-1}^V | Z_{1:k-1})$  is the posterior state distribution at time  $k-1$ , and  $p(X_k^V | X_{k-1}^V)$  is the symmetric discrete state transition density defined over the transition graph on the right of Fig. 4:

$$p(X_k^V | X_{k-1}^V) = \frac{\mathcal{L}(X_k^V | X_{k-1}^V)}{\sum_{X_k^V} \mathcal{L}(X_k^V | X_{k-1}^V)} \quad (8)$$

where the conditional likelihood function is defined as:

$$\mathcal{L}(X_k^V = \iota | X_{k-1}^V = J) = \begin{cases} \mathcal{L}_1 & \text{if } \iota = J \\ \mathcal{L}_{\bowtie} & \text{if } \iota \bowtie J \\ \mathcal{L}_{\tilde{\bowtie}} & \text{if } \iota \tilde{\bowtie} J \\ 0 & \text{if } \iota \not\bowtie J \end{cases} \quad (9)$$

and  $\bowtie$ ,  $\not\bowtie$ , and  $\tilde{\bowtie}$  denote adjacency, non-adjacency, and diagonal adjacency of qualitative states (Fig. 4 left), and appear as solid, missing, and dashed edges between graph nodes in Fig. 4 (right), respectively. When applied to the graph in Fig. 4, the conditional distribution in (8) is depicted as the following symmetric, positive definite matrix:

$$p(X_k^V | X_{k-1}^V) = \begin{matrix} & \overbrace{\begin{matrix} x_k^V \\ \begin{matrix} p_1 & p_{\bowtie} & 0 & p_{\bowtie} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} \\ p_{\bowtie} & p_1 & p_{\bowtie} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 \\ 0 & p_{\bowtie} & p_1 & p_{\bowtie} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} \\ p_{\bowtie} & 0 & p_{\bowtie} & p_1 & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} \\ p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_1 & p_{\bowtie} & 0 & p_{\bowtie} \\ p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_1 & p_{\bowtie} & 0 \\ 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_1 & p_{\bowtie} \\ p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\bowtie} & 0 & p_{\bowtie} & p_1 \end{matrix} \end{matrix}} \\ \left. \begin{matrix} x_{k-1}^V \\ \begin{matrix} p_{\bowtie} & p_{\tilde{\bowtie}} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_1 & p_{\bowtie} & 0 \\ 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\tilde{\bowtie}} & 0 & p_{\bowtie} & p_1 & p_{\bowtie} \\ p_{\tilde{\bowtie}} & 0 & p_{\tilde{\bowtie}} & p_{\bowtie} & p_{\bowtie} & 0 & p_{\bowtie} & p_1 \end{matrix} \end{matrix} \right\} \end{matrix} \quad (10)$$

where:

$$p_{(\cdot)} = \frac{\mathcal{L}_{(\cdot)}}{\mathcal{L}_1 + 3\mathcal{L}_{\bowtie} + 2\mathcal{L}_{\tilde{\bowtie}}} \quad (11)$$

Conceptually, the likelihoods,  $\mathcal{L}_{(\cdot)}$ , can be set according to the relative area of the boundary associated with each type of transition, giving:  $\mathcal{L}_1 > \mathcal{L}_{\bowtie} > \mathcal{L}_{\tilde{\bowtie}}$ .

#### 4.1.2. Belief Update:

Minding the conditional independence rules defined in (4), the prior categorical distribution at time  $k$ , (7),

is updated to reflect the observation at time  $k$  via the following equation:

$$\begin{aligned} p(X_k^V | Z_{1:k}) &= \frac{p(X_k^V, Z_k | Z_{1:k-1})}{p(Z_k | Z_{1:k-1})} \\ &= \frac{p(Z_k | X_k^V) \cdot p(X_k^V | Z_{1:k-1})}{\sum_{X_k^V=1}^8 p(Z_k | X_k^V) \cdot p(X_k^V | Z_{1:k-1})} \end{aligned} \quad (12)$$

where  $p(X_k^V | Z_{1:k-1})$  is the prior computed in (7). The conditional measurement likelihood,  $p(Z_k | X_k^V)$ , is found by counting the sensor returns from the discrete region of space corresponding to  $X_k^V$ , parameterized by the bearing and range intervals defined in (6):

$$p(Z_k | X_k^V) = \sum_{\ell=1}^{n_k^z} (\beta_{z_\ell} \in \mathbb{B}_{X_k^V}) \cap (\rho_{z_\ell} \in \mathbb{R}_{X_k^V}) \quad (13)$$

where  $\beta_{z_\ell}$  and  $\rho_{z_\ell}$  denote the bearing and range to sensor return  $z_\ell \in Z_k \forall \ell \in \{1, \dots, n_k^z\}$ . Note that the argument to the sum in (13) evaluates to 1, for points that lie within the discrete region of space corresponding to  $X_k^V$ , and 0 for those that do not; in this way, (13) counts the observations supporting qualitative state  $X_k^V$ .

## 4.2. Action Space Model

Action mode is reserved for objects of increasing significance to the consumer. For the autonomous driving example, these objects have a significant impact on the planning routine (i.e. the consumer), but are not at an immediate risk of collision [16], [17]. Therefore, a reasonable estimate of the object's position, velocity, and approximate size are desired to effectively anticipate their future behavior, and effectively plan around them. To this end, the extended object tracking approach chosen for the action space in the autonomous driving example is the random matrix method introduced in [23] and studied further in [24] and [5]. For the random matrix approach, the object state in action space at time  $k$ ,  $X_k^A$ , is defined as a random vector representing the objects position and velocity in the motion plane:

$$X_k^A = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix}_k \quad (14)$$

and the object extent in the motion plane is modeled as an ellipse by way of a symmetric positive definite random matrix,  $\mathbf{E}_k$ :

$$\mathbf{E}_k = \begin{bmatrix} e_x & e_{x,y} \\ e_{y,x} & e_y \end{bmatrix}_k \quad (15)$$

The tracking problem in (5) is then to estimate the joint distribution over the object state and elliptical extent given the history of measurements,  $Z_{1:k}$ :

$$p(X_k^A, \mathbf{E}_k | Z_{1:k}) \quad (16)$$

The joint distribution in (16) can be factored exactly into the product of vector and matrix valued distributions:

$$p(X_k^A, \mathbf{E}_k | Z_{1:k}) = p(X_k^A | \mathbf{E}_k, Z_{1:k})p(\mathbf{E}_k | Z_{1:k}) \quad (17)$$

where  $p(X_k^A | \mathbf{E}_k, Z_{1:k})$  is the vector valued distribution over the object state, modeled as a multivariate Gaussian, and  $p(\mathbf{E}_k | Z_{1:k})$  is the matrix valued distribution over the elliptical object extent, modeled to be inverse Wishart:

$$\begin{aligned} p(X_k^A | \mathbf{E}_k, Z_{1:k}) &= \mathcal{N}(\bar{X}_{k|k}^A, \mathbf{P}_{k|k}^A) \\ p(\mathbf{E}_k | Z_{1:k}) &= \mathcal{W}^{-1}(\Psi_{k|k}, \alpha_{k|k}) \end{aligned} \quad (18)$$

Thus, the posterior distribution in (16) is fully specified by the Gaussian mean,  $\bar{X}_k^A$ , and covariance,  $\mathbf{P}_k^A$ , coupled with the inverse Wishart scale matrix,  $\Psi_k$ , and degrees of freedom,  $\alpha_k$ .

The inverse Wishart distribution serves as the conjugate prior for the covariance matrix of a multivariate Gaussian. Further, the mean,  $\bar{\mathbf{E}}_k$ , variance of the  $(i, j)$ th element,  $(\sigma_{k|k}^{i,j})^2$ , and covariance between the  $(i, j)$ th and  $(\ell, m)$ th elements,  $\sigma_{k|k}^{(i,j),(\ell,m)}$ , of the extent matrix belief,  $p(\mathbf{E}_k | Z_{1:k})$ , are computed from the inverse Wishart parameters as:

$$\bar{\mathbf{E}}_{k|k} = \frac{\Psi_{k|k}}{\alpha_{k|k} + d - 1} \quad (19)$$

$$\begin{aligned} (\sigma_{k|k}^{i,j})^2 &= \frac{(\alpha_{k|k} - d + 1)(\psi_{k|k}^{i,j})^2 + (\alpha_{k|k} - d - 1)\psi_{k|k}^{i,i}\psi_{k|k}^{j,j}}{(\alpha_{k|k} - d)(\alpha_{k|k} - d - 1)^2(\alpha_{k|k} - d - 3)} \\ \sigma_{k|k}^{(i,j),(\ell,m)} &= \frac{2\psi_{k|k}^{i,j}\psi_{k|k}^{\ell,m} + (\alpha_{k|k} - d - 1)(\psi_{k|k}^{i,\ell}\psi_{k|k}^{j,m} + \psi_{k|k}^{i,m}\psi_{k|k}^{j,\ell})}{(\alpha_{k|k} - d)(\alpha_{k|k} - d - 1)^2(\alpha_{k|k} - d - 3)} \end{aligned} \quad (20)$$

where  $d$  is the dimension of  $\mathbf{E}_k$  ( $d = 2$  in this case).

#### 4.2.1 Belief Prediction:

The kinematic states evolve according to a stochastic continuous time differential equation model of the form in (66) provided in Appendix A. Therefore, the Kalman filter prediction equations in (69) are used to compute the mean and covariance of the prior state belief at time  $k$ :

$$p(X_k^A | Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k|k-1}^A, \mathbf{P}_{k|k-1}^A) \quad (21)$$

The dynamics model matrices,  $\mathbf{F}_k^A$  and  $\mathbf{G}_k^A$ , are defined as a function of the ego vehicle rotation matrix,  $\mathbf{M}_{g,k}^{\text{ego}}$ , and its time-derivatives:

$$\begin{aligned} \mathbf{F}_k^A &= \begin{bmatrix} \mathbf{0}_{2 \times 2} & \mathbf{I}_{2 \times 2} \\ (\mathbf{M}_{g,k}^{\text{ego}})^T \ddot{\mathbf{M}}_{g,k}^{\text{ego}} & -2(\mathbf{M}_{g,k}^{\text{ego}})^T \dot{\mathbf{M}}_{g,k}^{\text{ego}} \end{bmatrix} \\ \mathbf{G}_k^A &= \begin{bmatrix} \mathbf{0}_{2 \times 2} \\ -(\mathbf{M}_{g,k}^{\text{ego}})^T \end{bmatrix} \end{aligned} \quad (22)$$

where subscript ‘g’ indicates that the variable is described in a global coordinate frame, and superscript

‘ego’ denotes that the variable pertains to the ego vehicle dynamics, and is provided by an independent localization routine on-board the ego vehicle. The ego vehicle rotation matrix and its time-derivatives are defined as:

$$\begin{aligned} \mathbf{M}_{g,k}^{\text{ego}} &= \begin{bmatrix} \cos \phi_{g,k}^{\text{ego}} & -\sin \phi_{g,k}^{\text{ego}} \\ \sin \phi_{g,k}^{\text{ego}} & \cos \phi_{g,k}^{\text{ego}} \end{bmatrix} \\ \dot{\mathbf{M}}_{g,k}^{\text{ego}} &= \frac{\partial \mathbf{M}_{g,k}^{\text{ego}}}{\partial \phi_{g,k}^{\text{ego}}} \dot{\phi}_{g,k}^{\text{ego}} \\ \ddot{\mathbf{M}}_{g,k}^{\text{ego}} &= \frac{\partial \mathbf{M}_{g,k}^{\text{ego}}}{\partial \phi_{g,k}^{\text{ego}}} \ddot{\phi}_{g,k}^{\text{ego}} - \mathbf{M}_{g,k}^{\text{ego}} (\dot{\phi}_{g,k}^{\text{ego}})^2 \end{aligned} \quad (23)$$

The process noise represents the object acceleration in the ego reference frame, which is assumed to be driven by the following Gaussian white noise process:

$$W_k^A \sim \mathcal{N} \left( \begin{bmatrix} \ddot{x}_{g,k}^{\text{ego}} \\ \ddot{y}_{g,k}^{\text{ego}} \end{bmatrix}, \mathbf{Q}_k^A \right) \quad (24)$$

The elliptical object extent evolves according to the following rotation, accounting for the changing perspective of the ego-vehicle:

$$\mathbf{E}_k = \mathbf{M}_{g,\Delta k}^{\text{ego}} \mathbf{E}_{k-1} (\mathbf{M}_{g,\Delta k}^{\text{ego}})^T \quad (25)$$

where  $\mathbf{M}_{g,\Delta k}^{\text{ego}}$  is a rotation matrix accounting for the change in orientation of the ego-vehicle from  $k-1$  to  $k$ :

$$\mathbf{M}_{g,\Delta k}^{\text{ego}} = \begin{bmatrix} \cos(\phi_{g,k}^{\text{ego}} - \phi_{g,k-1}^{\text{ego}}) & -\sin(\phi_{g,k}^{\text{ego}} - \phi_{g,k-1}^{\text{ego}}) \\ \sin(\phi_{g,k}^{\text{ego}} - \phi_{g,k-1}^{\text{ego}}) & \cos(\phi_{g,k}^{\text{ego}} - \phi_{g,k-1}^{\text{ego}}) \end{bmatrix} \quad (26)$$

Therefore, the parameters of the inverse Wishart distribution are predicted over the time interval,  $\delta t$ , using the following equations:

$$\begin{aligned} \Psi_{k|k-1} &= \mathbf{M}_{g,\Delta k}^{\text{ego}} \Psi_{k-1|k-1} (\mathbf{M}_{g,\Delta k}^{\text{ego}})^T \\ \alpha_{k|k-1} &= \exp\left(\frac{-\delta t}{\tau}\right) (\alpha_{k-1|k-1} - 2) + 2 \end{aligned} \quad (27)$$

where  $\tau$  is the user-defined time constant governing the rate of change of the object extent.

#### 4.2.2. Belief Update:

The measurement of objects in action space is defined as an observation of the object centroid:

$$Z_k^A = \frac{1}{n_k^z} \sum_{\ell=1}^{n_k^z} z_k^\ell \quad (28)$$

where  $z_k^\ell \forall \ell \in \{1, \dots, n_k^z\}$  are the individual raw sensor returns at time  $k$ . The measurement in (28) is modeled as:

$$\begin{aligned} Z_k^A &= \mathbf{H} X_k^A + \nu_k \\ \mathbf{H} &= [\mathbf{I}_2, \mathbf{0}_2] \end{aligned} \quad (29)$$

where the measurement noise is defined as:

$$\begin{aligned} \nu_k &\sim \mathcal{N}(0_2, \bar{\mathbf{R}}_k) \\ \bar{\mathbf{R}}_{k|k-1} &= \Psi_{k|k-1} + \mathbf{R}_k^A \end{aligned} \quad (30)$$

where  $\mathbf{R}_k^A$  is the measurement noise covariance of each individual sensor return, which is typically provided by the sensor specification. Notice that the sensor uncertainty reflected in (30) is bloated by the object extent scale matrix,  $\Psi_{k|k-1}$ , thus tracking precision degrades for large objects.

Given the sensor model in (29), the measurement related parameters of the joint Gaussian distribution over the object state and measurement in (71) are given as:

$$\begin{aligned} \bar{Z}_k^A &= \mathbf{H}\bar{X}_{k|k-1}^A \\ \mathbf{P}_{Z_k}^A &= \mathbf{H}\mathbf{P}_{k|k-1}^A\mathbf{H}^T + \frac{1}{n_k^z}\bar{\mathbf{R}}_{k|k-1} \\ \mathbf{P}_{X_{k|k-1}Z_k}^A &= \mathbf{P}_{k|k-1}^A\mathbf{H}^T \end{aligned} \quad (31)$$

and the Kalman filter equations in (73), provided in Appendix A, are used to compute the posterior distribution over the object state in (17) and (18):

$$p(X_k^A | \mathbf{E}_k, Z_{1:k}) = \mathcal{N}(\bar{X}_{k|k}^A, \mathbf{P}_{k|k}^A) \quad (32)$$

The parameters of the inverse Wishart distribution are updated to get the posterior distribution over the object extent in (17) and (18),

$$p(\mathbf{E}_k | Z_{1:k}) = \mathcal{W}^{-1}(\Psi_{k|k}, \alpha_{k|k}) \quad (33)$$

using the following equations:

$$\begin{aligned} \Psi_{k|k} &= \frac{1}{\alpha_{k|k}}(\alpha_{k|k-1}\Psi_{k|k-1} + \hat{\mathbf{N}}_{k|k-1} + \hat{\Sigma}_{k|k-1}) \\ \alpha_{k|k} &= \alpha_{k|k-1} + n_k^z \end{aligned} \quad (34)$$

where:

$$\begin{aligned} \hat{\mathbf{N}}_{k|k-1} &= \Upsilon_{k|k-1}\mathbf{N}_{k|k-1}\Upsilon_{k|k-1}^T \\ \hat{\Sigma}_{k|k-1} &= \Xi_{k|k-1}\Sigma_k\Xi_{k|k-1}^T \end{aligned} \quad (35)$$

and:

$$\begin{aligned} \mathbf{N}_{k|k-1} &= (Z_k^A - \bar{Z}_{k|k-1}^A)(Z_k^A - \bar{Z}_{k|k-1}^A)^T \\ \Sigma_k &= \sum_{\ell=1}^{n_k^z} (z_k^\ell - Z_k^A)(z_k^\ell - Z_k^A)^T \\ \Upsilon_{k|k-1} &= \Psi_{k|k-1}^{1/2}(\mathbf{P}_{Z_k}^A)^{-1/2} \\ \Xi_{k|k-1} &= \Psi_{k|k-1}^{1/2}\bar{\mathbf{R}}_{k|k-1}^{-1/2} \end{aligned} \quad (36)$$

Matrix square roots when computing  $\Upsilon_{k|k-1}$  and  $\Xi_{k|k-1}$  in (36) are computed via the Cholesky factorization.

### 4.3. Personal Space Model

Personal space is reserved for objects with paramount relevance to the consumer; for the autonomous driving example, these are objects deemed to be at

immediate risk of collision. For this reason, precise estimates of object kinematics and occupied space are critical for safely interacting with objects in personal space, deeming the tracking approaches chosen for Vista and Actions spaces insufficient.

For the autonomous driving example, the personal space object state is defined as the position, velocity, and orientation of the object relative to some arbitrary initial orientation, described in a coordinate frame fixed to the ego-vehicle centroid:

$$X_k^P = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \\ \phi \end{bmatrix}_k \quad (37)$$

Note that the orientation,  $\phi$ , is decoupled from the object heading,  $\tan^{-1}(\dot{x}/\dot{y})$ , to accommodate arbitrary objects with a variety of latent motion constraints.

The object extent,  $\chi_k$ , is modeled as the most recent lidar scan returned from the object at each time step,

$$p(\chi_k | Z_{1:k}) = \mathcal{N}(Z_k, \mathbf{R}_k) \quad (38)$$

effectively maintaining a detailed, non-parametric, representation of the *immediately visible* object surface over a single time step.

#### 4.3.1. Belief Prediction:

Given that the first four object states in (37) are identical to  $X_k^A$  given in (14), they evolve according to the same model. The additional state, the orientation of the object, evolves according to the following scalar differential equation:

$$\dot{\phi}_k = \dot{\phi}_{g,k} - \dot{\phi}_{g,k}^{\text{ego}} \quad (39)$$

Therefore, the parameters of the prior distribution at  $k$ ,

$$p(X_k^P | Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k|k-1}^P, \mathbf{P}_{k|k-1}^P) \quad (40)$$

can be computed using the Kalman filter equations in (69), provided in Appendix A, using the following model:

$$\begin{aligned} \mathbf{F}_k^P &= \begin{bmatrix} \mathbf{F}_k^A & \mathbf{0}_{4 \times 1} \\ \mathbf{0}_{1 \times 4} & \mathbf{0} \end{bmatrix} \\ \mathbf{G}_k^P &= \begin{bmatrix} \mathbf{G}_k^A & \mathbf{0}_{2 \times 1} \\ \mathbf{0}_{1 \times 2} & -1 \end{bmatrix} \\ W_k^P &\sim \mathcal{N}(\bar{W}_k^P, \mathbf{Q}_k^P) \end{aligned} \quad (41)$$

where:

$$\begin{aligned} \bar{W}_k^P &= \begin{bmatrix} \bar{W}_k^A \\ \dot{\phi}_{g,k}^{\text{ego}} \end{bmatrix} \\ \mathbf{Q}_k^P &= \begin{bmatrix} \mathbf{Q}_k^A & \mathbf{0}_{2 \times 1} \\ \mathbf{0}_{1 \times 2} & q_{\dot{\phi}_{g,k}} \end{bmatrix} \end{aligned} \quad (42)$$

The object is assumed to be a rigid body, thus the object extent,  $\chi_k$ , is propagated forward in time via the following rigid body transform:

$$\chi_k = \mathbf{M}_{\Delta k}^{\text{Block}}(\chi_{k-1} - T_{k-1}^{\text{Block}}) + T_k^{\text{Block}} \quad (43)$$

where:

$$\begin{aligned} \mathbf{M}_{\Delta k}^{\text{Block}} &= \mathbf{I}_{n_{k-1}^z} \otimes \begin{bmatrix} \cos(\phi_k - \phi_{k-1}) & -\sin(\phi_k - \phi_{k-1}) \\ \sin(\phi_k - \phi_{k-1}) & \cos(\phi_k - \phi_{k-1}) \end{bmatrix} \\ T_k^{\text{Block}} &= \mathbf{1}_{n_{k-1}^z \times 1} \otimes \begin{bmatrix} x_k \\ y_k \end{bmatrix} \end{aligned} \quad (44)$$

$\otimes$  denotes the kronecker product, and  $(x_k, y_k, \phi_k)$  refer to the position and orientation object states in (37). The parameters of the prior distribution over the object extent at  $k$ ,

$$p(\chi_k | Z_{1:k-1}) = \mathcal{N}(\bar{\chi}_{k|k-1}, \mathbf{P}_{\chi_{k|k-1}}) \quad (45)$$

can be computed with the Sigma Point Transform [22].

#### 4.3.2. Belief Update:

The measurement of objects in personal space is inspired by [29], [30], and defined as an observation of the extremities of the object:

$$Z_k^{\text{P}} = \begin{bmatrix} \beta_k^{\text{cw}} \\ \beta_k^{\text{ccw}} \\ \rho_k \end{bmatrix} \quad (46)$$

where  $\beta_k^{\text{cw}}$ ,  $\beta_k^{\text{ccw}}$ , and  $\rho_k$  denote the clockwise and counterclockwise most bearings, and the minimum range to the object. The measurement model corresponding to (46) is defined as:

$$Z_k^{\text{P}} = h(\chi_k) + \nu_k \quad (47)$$

where  $\chi_k$  is the extent model, and  $h(\cdot)$  is a function extracting the measurement metadata in (46) from  $\chi_k$ .

The measurement related parameters of the joint distribution over the measurement metadata,  $Z_k^{\text{P}}$ , and object state,  $X_k^{\text{P}}$ , in (71), specifically,  $\bar{Z}_k^{\text{P}}$ ,  $\mathbf{P}_{Z_k}^{\text{P}}$ , and  $\mathbf{P}_{X_{k|k-1}Z_k}^{\text{P}}$ , are computed from the prior distributions over the state, (40), and object extent, (45), using the Sigma Point Transform [22]. Finally, the parameters of the posterior state distribution,

$$p(X_k^{\text{P}} | Z_{1:k}) = \mathcal{N}(\bar{X}_{k|k}^{\text{P}}, \mathbf{P}_{k|k}^{\text{P}}) \quad (48)$$

are computed with the Kalman filter update equations in (73), and the distribution over the object extent is updated with (38); i.e. replacing the prior extent belief with the most recent lidar scan.

## 5. MODE TRANSITIONS

The mode transitions among, Vista, Action, and Personal spaces, depicted in Fig. 3, are fully defined by their guards,  $\mathbb{G}$ , and invariants,  $\mathbb{I}$ , informed by the object relevancy metrics, as well as the state transition functions of the form,  $X_k^i = g_{j \rightarrow i}(X_k^j)$ , which transform

TABLE II  
Example relevance-based metrics

Relevance definition	Associated Metric
• Proximity	• Distance to object
• Danger	• Probability of collision
• Anomalous/erratic behavior	• $\chi^2$ test on tracker innovations
• Object of interest	• Object recognition probability

the state belief from the source mode representation,  $j$ , to that of the destination mode,  $i$ .

In hybrid system theory, the *invariants* are a set of conditions that must be satisfied for the system to operate within each discrete mode. This is in contrast to the *guards*, which are a set of conditions that must be satisfied to invoke each discrete mode transition. For the autonomous vehicle example, the hybrid system of Fig. 3 is deterministic; i.e. the invariants are chosen to be perfectly aligned with the guards, such that, at any given instant, there is a single valid mode of operation, and all 3 modes are reachable. The focus of this work is to invoke high precision EOT methods for objects that are *relevant* to the consumer, and inexpensive EOT methods for objects of peripheral relevance. Therefore, metrics informing the guards and invariants should be chosen to reflect a measure of the consumer's definition of object *relevance*. A list of example relevance definitions coupled with suggestions for relevancy metrics is provided in Table II. The definitions and metrics chosen for the autonomous vehicle example are presented in the following sections along with their associated guards, invariants, and mode transition functions.

### 5.1. Vista $\leftrightarrow$ Action

#### 5.1.1. Probabilistic object relevancy metric:

For the autonomous driving example, the relevancy metric informing transitions between Vista and Action modes is chosen as the probability that the object is 'Far' from the ego-vehicle,  $p(\text{Far}_k | Z_{1:k})$ ; 'Far' is defined by the discrete abstraction in Fig. 4, as the object occupying any of the first four qualitative states. Therefore, the far probability metric is computed as:

$$\begin{aligned} p(\text{Far}_k | Z_{1:k}) &= p\left(\bigcup_{X_k^{\text{V}}=1}^4 X_k^{\text{V}} | Z_{1:k}\right) \\ &= \sum_{X_k^{\text{V}}=1}^4 p(X_k^{\text{V}} | Z_{1:k}) \end{aligned} \quad (49)$$

where the equality of the first and second lines of (49) is conditioned on the fact that the qualitative states are disjoint. For the transition from Vista to Action,  $\text{V} \rightarrow \text{A}$ ,  $p(X_k^{\text{V}} | Z_{1:k})$  is the current state belief posterior computed during the measurement update step. For the reverse transition,  $\text{V} \leftarrow \text{A}$ ,  $p(X_k^{\text{V}} | Z_{1:k})$  is computed from

(12) with a uniform prior over the qualitative states,  $X_k^V$ . Specifically,

$$p(X_k^V | Z_{1:k}) = \frac{p(Z_k | X_k^V)}{\sum_{X_k^V} p(Z_k | X_k^V)} \quad (50)$$

where  $p(Z_k | X_k^V)$  is the vista mode measurement likelihood defined in (13).

### 5.1.2. Guards:

The guards are defined by thresholding the probabilistic object relevancy metric defined above. Specifically,

$$\begin{aligned} \mathbb{G}_{V \leftarrow A} &\triangleq p(\text{Far}_k | Z_{1:k}) \leq p_{V,A} \\ \mathbb{G}_{V \leftarrow A} &\triangleq p(\text{Far}_k | Z_{1:k}) > p_{A,V} \end{aligned} \quad (51)$$

where  $p_{A,V} \geq p_{V,A}$  are user-defined probability thresholds.

### 5.1.3. Transition functions:

Given that the Vista model, by design, reflects only low fidelity qualitative information about the object position, it does not have much to offer the continuous metrical Action model in terms of a transition function; therefore, at the  $V \rightarrow A$  transition, the Action model is initialized directly from the lidar scan. Specifically, the inverse Wishart parameters of the extent model distribution are computed as:

$$\begin{aligned} \Psi_k &= \frac{1}{n_k^z} \Sigma_k \\ \alpha_k &= n_k^z \end{aligned} \quad (52)$$

where  $\Sigma_k$  is defined in (36), and  $n_k^z$  is the number of measurement returns at time  $k$ . The Gaussian parameters of the state distribution are computed as:

$$\begin{aligned} \bar{X}_k^A &= \begin{bmatrix} Z_k^A \\ -\dot{x}_{g,k}^{\text{ego}} \\ -\dot{y}_{g,k}^{\text{ego}} \end{bmatrix} \\ \mathbf{P}_k^A &= \begin{bmatrix} \frac{1}{n_k^z} \bar{\mathbf{R}}_k & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{P}_{V_0}^A \end{bmatrix} \end{aligned} \quad (53)$$

where  $Z_k^A$  and  $\bar{\mathbf{R}}_k$  are the mean and covariance of the Action space centroid measurement, defined in (28) and (30), respectively, and  $\mathbf{P}_{V_0}^A$  is set to a large diagonal matrix reflecting the large amount of uncertainty in the velocity initialization. Note that the velocity state initialization is naive in assuming the object is static in the global reference frame. However, the linear properties of the dynamics and measurement model in (22) and (28) allow for a quick estimate convergence from a potentially poor initialization, as demonstrated in the results section. More elaborate initialization schemes can be implemented without loss of generality.

At the  $V \leftarrow A$  transition, the Vista model is initialized from (50), which was already computed to evaluate the guard,  $\mathbb{G}_{V \leftarrow A}$ , in (51).

## 5.2. Action $\leftrightarrow$ Personal

### 5.2.1. Probabilistic object relevancy metric:

For the autonomous driving example, the relevancy metric governing the transitions among Action and Personal modes is chosen as the anticipated probability of collision with the object over a defined time horizon,  $h$ ,  $p(C_{k:k+h} | Z_{1:k})$ . The anticipated probability of collision,  $p(C_{k:k+h} | Z_{1:k})$ , is taken as the maximum *instantaneous* collision probability over each time step in the horizon,  $h$ :

$$p(C_{k:k+h} | Z_{1:k}) = \max_{\ell \in \{1, \dots, h\}} [p(C_{k+\ell} | Z_{1:k})] \quad (54)$$

Conceptually, the instantaneous collision probabilities,  $p(C_{k+\ell} | Z_{1:k}) \forall \ell \in \{1, \dots, h\}$ , are computed as the probability that the space occupied by the object intersects that of the ego vehicle at each future instant,  $k + \ell$ . As demonstrated in Fig. 5, mathematically this is equivalent to the probability that the ego-vehicle centroid (i.e. the origin of the tracking coordinate frame) lies within the anticipated *collision region*,  $\mathcal{O}_{k+\ell}$ , defined as the dilation of the uncertain object extent at time  $k + \ell$  by the known ego vehicle extent.

Thus the anticipated instantaneous collision probability is calculated as:

$$p(C_{k+\ell} | Z_{1:k}) = \exp\left(-\frac{1}{2}(\bar{D}_{\mathcal{O}_{k+\ell|k}}^{\min})^T \mathbf{P}_{D_{\mathcal{O}_{k+\ell|k}}^{\min}}^{-1} \bar{D}_{\mathcal{O}_{k+\ell|k}}^{\min}\right) \quad (55)$$

where:

$$D_{\mathcal{O}_{k+\ell|k}}^{\min} \sim \mathcal{N}(\bar{D}_{\mathcal{O}_{k+\ell|k}}^{\min}, \mathbf{P}_{D_{\mathcal{O}_{k+\ell|k}}^{\min}}) \quad (56)$$

is the vector from the ego vehicle to the closest point in the collision region,  $\mathcal{O}_{k+\ell}$ , the mean and covariance of which can be computed using the sigma point transform [22].

In the simplest case, object state anticipation over the time horizon is accomplished by iterating over the usual filter prediction step. However, for highly dynamic scenes or large time horizons, the state uncertainty can quickly explode to produce an uninformative belief. In these cases, it is recommended to leverage more intelligent, specialized anticipation methods that integrate advanced features such as traffic lane following controllers, traffic laws, etc. [16], [17].

### 5.2.2. Guards:

The guards are defined by thresholding the probabilistic object relevancy metric defined above. Specifically,

$$\begin{aligned} \mathbb{G}_{A \rightarrow P} &\triangleq p(C_{k:k+h} | Z_{1:k}) \geq p_{A,P} \\ \mathbb{G}_{A \rightarrow P} &\triangleq p(C_{k:k+h} | Z_{1:k}) < p_{P,A} \end{aligned} \quad (57)$$

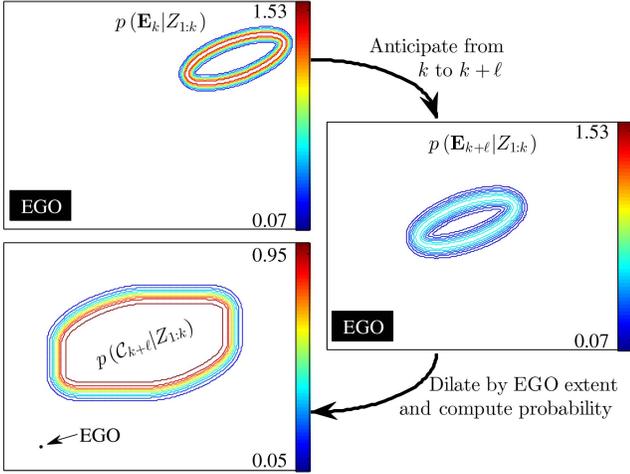


Fig. 5. Demonstration of the instantaneous collision probability calculation for an object in Action space. **Top:** Belief of the elliptical object extent at time  $k$ . **Right:** Anticipated belief of the elliptical object extent at time  $k + \ell$ . **Bottom:** Probability of collision. Note that the ego-vehicle centroid was swept over the space to generate the probability contours for demonstration purposes, however, in practice, the collision probability only needs to be evaluated at the ego vehicle centroid labeled EGO in the figure.

where  $p_{A,P} \geq p_{P,A}$  are user-defined probability thresholds.

### 5.2.3. Transition functions:

Given the similarity of the state representations in (14) and (37), the state transition functions for  $A \rightarrow P$  is:

$$\begin{aligned} \bar{X}_k^P &= \begin{bmatrix} \bar{X}_k^A \\ 0 \end{bmatrix} \\ \mathbf{P}_k^P &= \begin{bmatrix} \mathbf{P}_k^A & 0_{4 \times 1} \\ 0_{1 \times 4} & \epsilon \end{bmatrix} \end{aligned} \quad (58)$$

where  $\epsilon$  is a small positive number indicating perfect knowledge of the initial relative orientation,  $\phi_k$ , while maintaining the positive definite requirement of  $\mathbf{P}_k^P$ ; since the extent models do not identify the *front* of the object,  $\phi_k$  is defined as the orientation relative to some arbitrary initialization, and thus can be initialized with absolute certainty to any numerical value. The distribution over the object extent in Personal space,  $p(X_k | Z_{1:k})$ , is initialized from the current lidar scan returned from the object via (38).

The state transition function from  $A \leftarrow P$  is defined as:

$$\begin{aligned} \bar{X}_k^A &= \bar{X}_k^P(1:4) \\ \mathbf{P}_k^A &= \mathbf{P}_k^P(1:4, 1:4) \end{aligned} \quad (59)$$

and the inverse Wishart parameters of the distribution over the object extent are initialized from the lidar scan as in (52).

### 5.3. Invariants

For the autonomous driving example, the invariants are chosen such that as a guard enables a transition

between two modes, the invariant for the source mode is violated, and the invariant for the destination mode is satisfied. Specifically:

$$\begin{aligned} \mathbb{I}_V &\triangleq p(\text{Far}_k | Z_{1:k}) > p_{A,V} \\ \mathbb{I}_A &\triangleq (p(\text{Far}_k | Z_{1:k}) \leq p_{V,A}) \cap \dots \\ &\quad (p(C_{k:k+h} | Z_{1:k}) < p_{A,P}) \\ \mathbb{I}_P &\triangleq p_{P,A} \leq p(C_{k:k+h} | Z_{1:k}) \end{aligned} \quad (60)$$

In this way, deterministic transitions are triggered as soon the guard is satisfied.

## 6. SIMULATION RESULTS

To demonstrate the ability of the proposed priority-based framework in Fig. 3 to automatically trade computation for tracking precision as a function of object relevance, the framework, as parameterized in Sections 4 and 5 for the autonomous driving example, is evaluated over the two simulated scenarios depicted in Figs. 6 and 7. The scenario depicted in Fig. 6 involves a star shaped object maneuvering with continuously variable orientation and velocity along a spiral trajectory centered on the stationary ego vehicle; this scenario is intended to represent a somewhat arbitrary, unstructured, and challenging tracking application. The scenario depicted in Fig. 7 represents a common autonomous driving scenario in which the ego-vehicle and object (modeled as rectangles) pass each other with less than 0.5 m clearance in a four-way controlled intersection. The vehicles initially approach the intersection at a constant cruising speed of 12 m/s ( $\approx 26.8$  mph), decelerate to a full stop at the edge of the intersection, pause for 3 s, then accelerate straight through the intersection until they reach their initial cruising speed.

Data is simulated for a 360° field-of-view planar lidar firing at 12.5 Hz with 0.5° bearing resolution, fixed to the centroid of the ego vehicle. Random sensor noise is sampled independently for each beam in each scan from  $\mathcal{N}(0, 1 \text{ cm}^2)$  and added to the lidar range returns to simulate the accuracy of realistic lidar sensors. The filter parameters used for both simulations are defined in Table III. All simulations were coded in Matlab with all feature accelerators and code optimizers turned off, and run on a single thread of an Intel® Core™ i7-4770 CPU @ 3.40 GHz. Given that the code is written in an interpreted language and has not been optimized, only discussion about *relative* computational effort among the tracker modes is meaningful.

For this evaluation, *precision*,  $\gamma_k^{n-\sigma}$ , is defined as:

$$\gamma_k^{n-\sigma} = \bar{A}^{-1}(p(X_k | Z_{1:k}), n) \quad (61)$$

where  $\bar{A}(p(X_k | Z_{1:k}), n)$  denotes the expected value of the area enclosed by the  $n - \sigma$  confidence bound. To obtain a fair comparison with objects in Vista space,

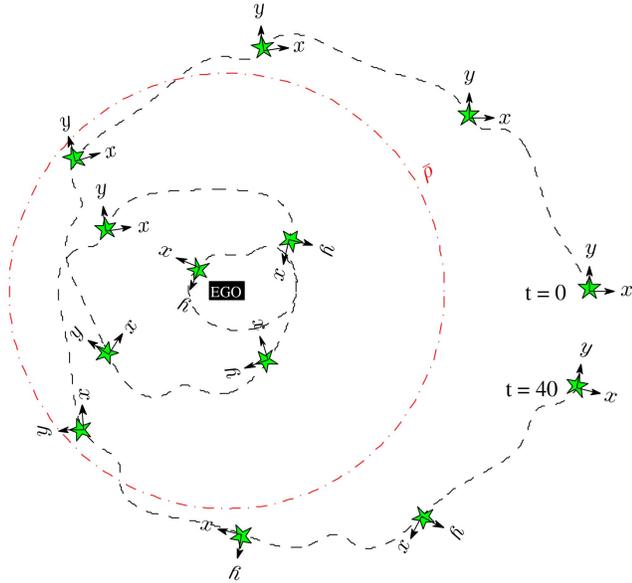


Fig. 6. Simulated scenario of a star shaped object spiraling in toward the ego-vehicle and then back out over a period of 40 s; 13 time steps are shown.

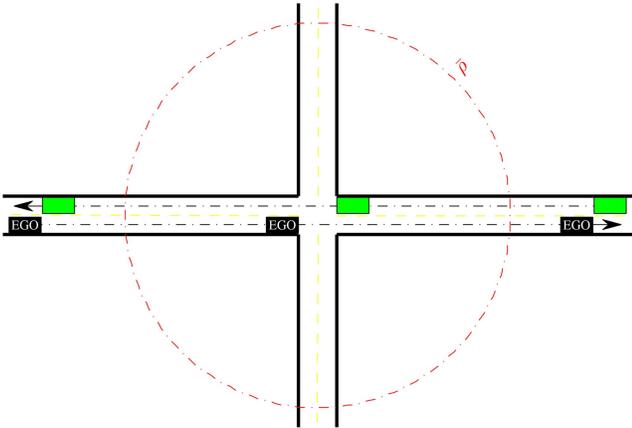


Fig. 7. Simulated scenario of a rectangular object passing the ego vehicle head-on in close proximity at an intersection; three time steps are shown.

TABLE III  
Hybrid tracking parameters

Mode Transitions	Vista	Action	Personal
$p_{V,A}$	0.45	$\bar{\rho}$ 30 m	$\tau$ 60 s
$p_{A,V}$	0.55	$p_{\mathcal{I}}$ 0.2	
$p_{A,P}$	0.5	$p_{\mathcal{A}}$ 0.1	
$p_{P,A}$	0.1	$p_{\mathcal{A}^c}$ 0.01	
$h$	1 s		

the object position belief is taken to be uniformly distributed over the qualitative region represented by the state, and the  $n - \sigma$  confidence bound is interpreted as the area required to enclose the same probability,  $p(n)$ , as the  $n - \sigma$  confidence bound of a Gaussian, where  $n - \sigma$  refers to the Mahalanobis distance from the mean

of the distribution:

$$p(n) = 1 - \exp\left(-\frac{n^2}{2}\right) \quad (62)$$

Specifically, for the discrete abstraction in Fig. 4:

$$\bar{A}(p(X_k | Z_{1:k}), n) = \begin{cases} p(n) \cdot \pi \cdot \frac{\mathbb{E}[\rho_o^2 - \rho_i^2]}{4} & \text{if Vista} \\ n^2 \cdot \pi \cdot \sqrt{|\mathbf{P}_{k|k}^{xy}|} & \text{otherwise} \end{cases} \quad (63)$$

where  $\rho_i$  and  $\rho_o$  denote the inner and outer radii of the qualitative regions associated with the Vista states. The expected value in the numerator of the Vista case of (63) becomes:

$$\mathbb{E}[\rho_o^2 - \rho_i^2] = p(X_k \in \text{Far} | Z_{1:k}) \cdot \rho_{\max}^2 \dots - [2 \cdot p(X_k \in \text{Far} | Z_{1:k}) - 1] \cdot \bar{\rho}^2 \quad (64)$$

where  $\rho_{\max}$  denotes the maximum range of the sensor,  $p(\text{Far} | Z_{1:k})$  is defined in (49). Note that the expression in (64) reflects that ‘Far’ states in Vista space are bounded at the sensor range,  $\rho_{\max} = 80$  m; while this is not technically an attribute of the abstraction in Fig. 4, it is a sensible bound to avoid infinite area (and infinitesimal precision) given that, inherent in the event that the object returns a sensor measurement, is the fact that the object must be within the range of the sensor.

For the purposes of this evaluation, computational effort,  $\epsilon_k$ , is defined as the clock time required to compute each filter recursion,  $\delta t_{\text{computation}}$ , normalized by the filter time step dictated by the sensor frequency,  $f_{\text{sensor}}$ :

$$\epsilon_k = \delta t_{\text{computation}} \cdot f_{\text{sensor}} \quad (65)$$

Figs. 8 and 9 demonstrate the tracking performance for the star and intersection scenarios, respectively, by comparing the *maximum-a-posteriori* (MAP) velocity estimates to the simulated truth values. In both scenarios, the filter appears inconsistent (under-confident) when in Action space, i.e. it is overestimating the filter uncertainty. This is an artifact of some over-simplifying assumptions in the object extent model limiting the amount of information that can be extracted from the lidar scan. Specifically, despite the centimeter-level precision of the lidar sensor, the measurement model in (29) reflects the naive and highly uncertain expectation that the origin of each lidar return is the centroid of the elliptical extent; a direct consequence of the extent model lacking a concept of object *surface*. Note that this naiveté is also a subtle but critical feature enabling the random matrix approach (Action space) to be computationally simple and efficient while simultaneously remaining flexible and robust in tracking objects from a variety of classes and applications. This measurement origin uncertainty is reflected in the measurement noise model of (30), which scales directly with the size of the object extent, and is ultimately the source of the degraded tracking precision in Action space.

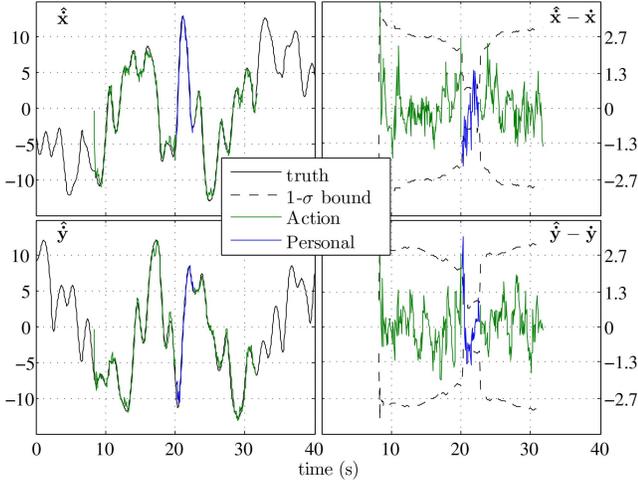


Fig. 8. Performance in tracking the velocity states for the star scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and  $1-\sigma$  bounds. Note that there is not a concept of velocity in Vista space.

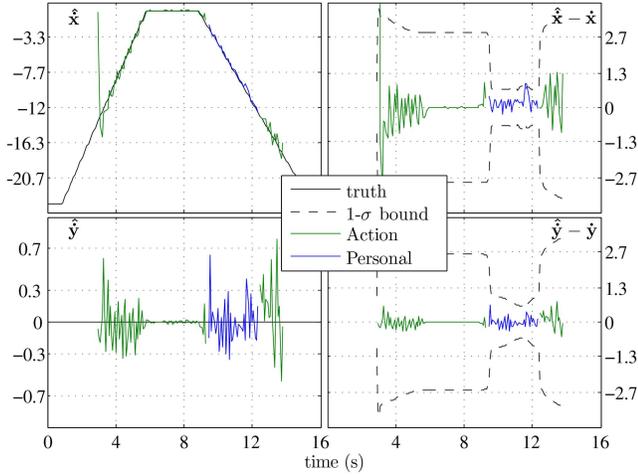


Fig. 9. Performance in tracking the velocity states for the intersection scenario: **Left:** MAP estimates overlaid on ground truth, **Right:** tracking error and  $1-\sigma$  bounds. Note that there is not a concept of velocity in Vista space.

Also apparent in Figs. 8 and 9 is that, as objects approach the ego vehicle, the likelihood that the ego vehicle may interact with the object increases and the tracker transitions to Personal mode. This transition triggers a dramatic improvement in the estimate uncertainty, which is a critical feature enabling the ego vehicle to safely maneuver in close proximity with uncooperative dynamic objects. In both scenarios, the filter quickly recovers from the naive velocity initialization defined in (53) within two time steps (0.16 s) of the filter transition from Vista to Action mode. Further, the filter seamlessly transitions between Action and Personal modes in both directions, mitigating the need for elaborate initialization schemes; a direct consequence of the synergy between models.

Figs. 10 and 11 plot the precision,  $\epsilon$  defined in (61), and computational effort,  $\gamma$  defined in (65), over time,

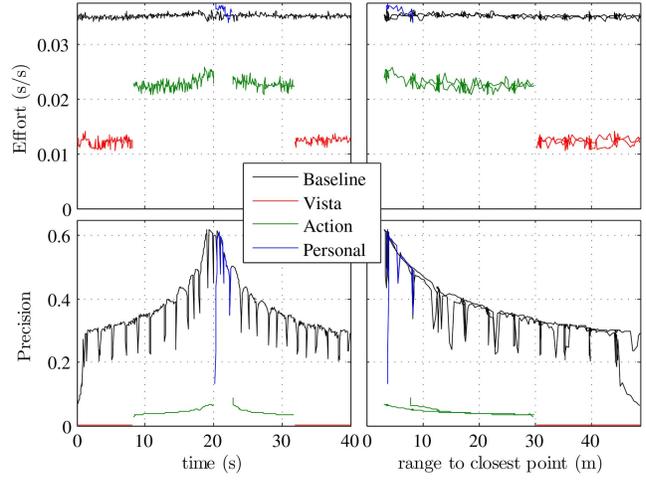


Fig. 10. Computational effort,  $\epsilon$ , and precision,  $\gamma$ , as a function of time (left) and range to the closest point on the object (right) for the star scenario.

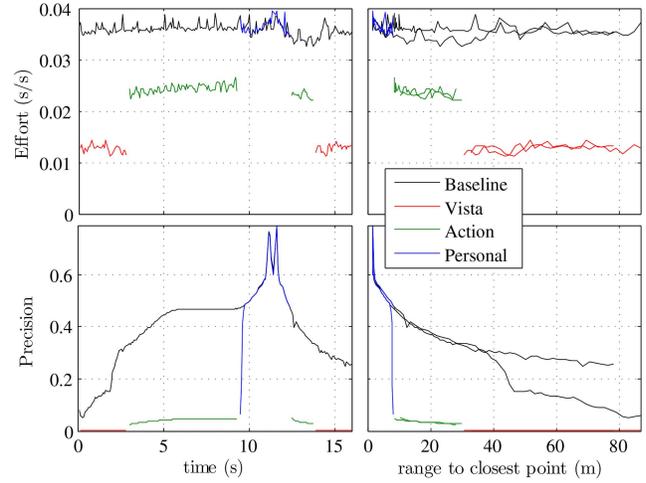


Fig. 11. Computational effort,  $\epsilon$ , and precision,  $\gamma$ , as a function of time (left) and range to the closest point on the object (right) for the intersection scenario.

TABLE IV  
Efficiency vs. Precision

	Vista	Action	Personal
Mean Effort, $\bar{\epsilon}$	0.01203	0.02180	0.03653
% of max Effort	32.9%	59.7%	100%
Mean Precision, $\bar{\gamma}$	0.00126	0.04257	0.52372
% of max precision	0.24%	8.1%	100%

for both the star and intersection scenarios, respectively; the combined summary of these metrics is provided in Table IV. *Baseline* refers to a tracker that operates solely in Personal mode to emphasize the contribution of the hybrid framework depicted in Fig. 3. The Personal model is used for comparison, as it is the only model of the three that achieves the tracking precision required for interacting with objects in close proximity—a requirement of many robotics applications, including autonomous driving. As designed, both the computational

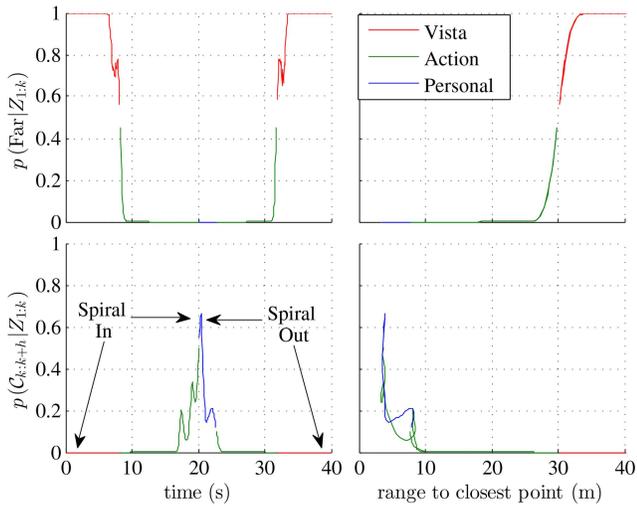


Fig. 12. Probabilistic object relevancy metric trajectories for the star scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is ‘Far’ away, governing the transitions between Vista and Action modes. **Bottom Row:** Anticipated collision probability for a  $h = 1$  s time horizon, governing the transitions between Action and Personal modes. **Left Column:** Variables plotted against time. **Right Column:** Variables plotted against distance to the closest point on the object.

effort and the tracking precision increase as the filter transitions from Vista through Action to Personal mode, and the reciprocal trend exists for transitions in the opposite direction. Specifically, in terms of computational effort, roughly 3 objects can be tracked in Vista mode for every 2 in Action mode, and every 1 in Personal mode, at the cost of decreased tracking precision. The periodic spike in the Personal mode precision for the star scenario in Fig. 10 is a direct consequence of the measurement model in (47) reflecting latent characteristics of the object shape, which invokes a relatively strong viewpoint-dependence for the state observability compared to the other tracking modes. This characteristic is not as apparent in Fig. 11 for the intersection scenario, due to the relatively simple object shape, and slow, acyclic, viewpoint changes compared to the star scenario; however, it is briefly apparent as the vehicles pass each other in close proximity at  $t \approx 13$  s, when the viewpoint is changing most rapidly.

Lastly, Figs. 12 and 13 demonstrate the trajectories of the probabilistic object relevancy metrics for both scenarios. The flat region in the bottom left of Fig. 13 in the approximate range,  $5.5 \text{ s} < t < 8.5 \text{ s}$ , corresponds to the 3 second pause of both vehicles before proceeding through the intersection. Notice that, while range to the closest point on the object inherently factors into the collision probability, it is not an accurate predictor in itself. This is most apparent in the bottom right of Fig. 13, in that the collision probability is strictly higher as the vehicles approach each other at the center of the intersection (portion of the curve labeled ‘On coming’ in Fig. 13) than it is after they depart the intersection in

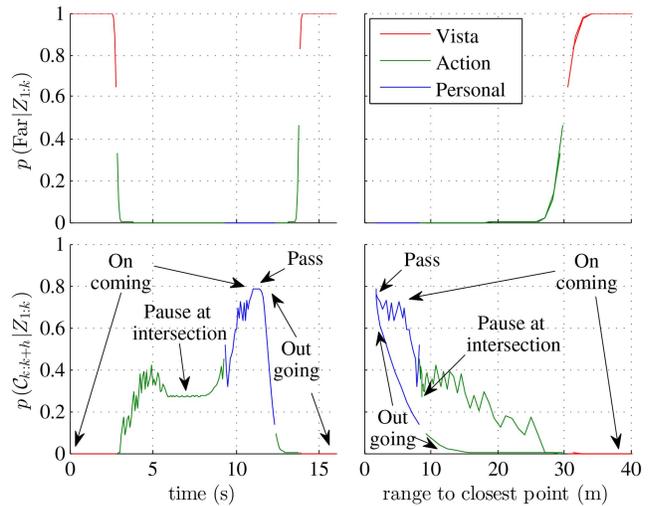


Fig. 13. Probabilistic object relevancy metric trajectories for the intersection scenario. Annotations referring to events in the ground truth scenario are provided for perspective. **Top Row:** Probability that the object is ‘Far’ away, governing the transitions between Vista and Action modes. **Bottom Row:** Anticipated collision probability for a  $h = 1$  s time horizon, governing the transitions between Action and Personal modes. **Left Column:** Variables plotted against time. **Right Column:** Variables plotted against distance to the closest point on the object.

opposing directions (portion of the curve labeled ‘Out going’ in Fig. 13). This is a direct result of the anticipatory nature of the collision probability. Specifically, as the vehicles approach, the algorithm anticipates that the distance between them continues to narrow, increasing the likelihood of an impending collision; conversely, as the vehicles depart, the algorithm anticipates that the distance between the objects continues to grow, decreasing the likelihood of an impending collision. This characteristic is not as apparent in Fig. 12 due to the spiral object trajectory. Specifically, given that the object approaches the ego vehicle without ever driving directly at it, the anticipation routine predicts that this behavior continues, and the probability of impending collision is relatively small until the object is within approximately 3 m of the ego vehicle. Given this attribute, and the exact symmetry of the spiral trajectory about  $t = 20$  s, the minor asymmetries in the collision probability in Fig. 12 (bottom) can be predominately attributed to the increase in the precision of the state belief in the latter half of the scenario (labeled ‘Spiral Out’ in Fig. 12)—a direct consequence of the hybrid mode transition to Personal space.

## 7. CONCLUSION

Inspired by human perception, this paper introduces a novel method to dynamically allocate algorithmic and computational resources to achieve variable precision tracking of extended objects. Many sensible extended object tracking (EOT) methods exist, with the main distinction being the model chosen to represent the object extent. In general, simple extent models result in com-

putationally efficient EOT, but engender low precision tracking by way of imprecise sensor models (i.e. large measurement source uncertainty); conversely, detailed and complex extent models tend to be computationally expensive, but engender high precision tracking by enabling complementary detailed and precise sensor models.

With the assertion that objects in a given scene are often of variable importance to the consumer of the tracker output, a priority-based tracking framework is proposed, enabling objects of critical importance to the consumer to be tracked with relatively expensive, high precision methods, and objects of peripheral importance to be tracked with relatively efficient, low precision methods. The proposed priority-based framework is a direct analog to the human perception concepts of *attention* and *focus*.

The priority-based EOT framework is parameterized for an example autonomous vehicle application in which the consumer of the tracking output is an anticipatory planner. Probabilistic object relevancy metrics are derived to convey the priority of an object to the consumer, and inform mode transitions in the hybrid model implementation of the priority-based EOT framework. Simulation results for two different scenarios are presented and compared to a baseline high precision EOT algorithm. The results demonstrate that the priority-based framework enables a significant computational savings by relaxing its precision requirements for objects deemed to be of peripheral importance, while maintaining high precision tracks for objects regarded as essential to the consumer (i.e. the anticipatory planner).

## APPENDIX A KALMAN FILTER

This section provides the Kalman filter prediction and update equations [4], [11], [33].

### A.1. Prediction

Given a stochastic linear vector differential equation model the form:

$$\begin{aligned} \dot{X}_k &= \mathbf{F}_k X_k + \mathbf{G}_k W_k \\ W_k &\sim \mathcal{N}(\bar{W}_k, \mathbf{Q}_k) \end{aligned} \quad (66)$$

describing the object dynamics, and a Gaussian belief of the posterior object state at time  $k-1$ ,

$$p(X_{k-1} | Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \quad (67)$$

the Gaussian prior distribution at time  $k$  is obtained by predicting the posterior at  $k-1$  over the time interval  $\delta t$ :

$$p(X_k | Z_{1:k-1}) = \mathcal{N}(\bar{X}_{k|k-1}, \mathbf{P}_{k|k-1}) \quad (68)$$

where the mean and covariance are computed using the following equations:

$$\begin{aligned} \bar{X}_{k|k-1} &= \int_{k-1}^k \dot{X}_t dt \\ \mathbf{P}_{k|k-1} &= \Phi_{k-1} \mathbf{P}_{k-1|k-1} \Phi_{k-1}^T + \Gamma_{k-1} (\mathbf{Q}_k \cdot \delta t) \Gamma_{k-1}^T \end{aligned} \quad (69)$$

where:

$$\Phi_{k-1} = \int_{k-1}^k \mathbf{F}_t dt, \quad \text{and} \quad \Gamma_{k-1} = \int_{k-1}^k \mathbf{G}_t dt \quad (70)$$

$\bar{X}_{k|k-1}$ ,  $\Phi_{k-1}$ , and  $\Gamma_{k-1}$  are computed using numerical integration techniques, such as Runge-Kutta.

### A.2. Update

Given that the object state,  $X_k$ , and measurement,  $Z_k$ , are jointly Gaussian:

$$p(X_k, Z_k | Z_{1:k-1}) = \mathcal{N} \left( \begin{bmatrix} \bar{X}_{k|k-1} \\ \bar{Z}_k \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{k|k-1}, & \mathbf{P}_{X_{k|k-1} Z_k} \\ \mathbf{P}_{X_{k|k-1} Z_k}^T, & \mathbf{P}_{Z_k} \end{bmatrix} \right) \quad (71)$$

where  $\bar{X}_{k|k-1}$  and  $\mathbf{P}_{k|k-1}$  are the mean and covariance of the prior state distribution computed in (69), and  $\bar{Z}_k$ ,  $\mathbf{P}_{Z_k}$ , and  $\mathbf{P}_{X_{k|k-1} Z_k}$  are the measurement mean, covariance, and state-measurement covariance derived from the particular sensor model. Then, the posterior distribution over the object state conditioned on the measurement is also Gaussian:

$$p(X_k | Z_{1:k}) = \mathcal{N}(\bar{X}_{k|k}, \mathbf{P}_{k|k}) \quad (72)$$

with parameters computed as:

$$\begin{aligned} \bar{X}_{k|k} &= \bar{X}_{k|k-1} + \mathbf{K}_k (Z_k - \bar{Z}_k) \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_{Z_k} \mathbf{K}_k^T \end{aligned} \quad (73)$$

where  $\mathbf{R}_k$  is the measurement noise covariance provided by the sensor specification, and the Kalman gain,  $\mathbf{K}_k$ , is defined as:

$$\mathbf{K}_k = \mathbf{P}_{X_{k|k-1} Z_k} \mathbf{P}_{Z_k}^{-1} \quad (74)$$

## ACKNOWLEDGMENT

This work was supported by the NSF Robust Intelligence grant IIS-1320490 and ARO grant W911NF-09-1-0466.

## REFERENCES

- [1] A. Andriynko, S. Roth, and K. Schindler  
An analytical formulation of global occlusion reasoning for multi-object tracking,  
in *Proceedings of the 13th IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011.
- [2] D. Arbuckle, A. Howard, and M. Mataric  
Temporal occupancy grids: a method for classifying the spatio-temporal properties of the environment,  
in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, October 2004, pp. 409-414.

- [3] J. Aue, M. R. Schmid, T. Graf, J. Effertz, and P. Muehlfellner  
Object tracking from medium level stereo camera data providing detailed shape estimation using local grid maps, in *Proceedings of the IEEE Intelligent Vehicles Symposium*, June 2013.
- [4] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan  
*Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.
- [5] M. Baum, M. Feldmann, D. Franken, U. D. Hanebeck, and W. Koch  
Extended object and group tracking: A comparison of random matrices and random hypersurface models, in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2010)*, October 2010.
- [6] M. Baum and U. D. Hanebeck  
Extended object tracking based on combined set-theoretic and stochastic fusion, in *Proceedings of the 12th International Conference on Information Fusion (FUSION)*, July 2009, pp. 1288–1295.
- [7] M. Baum and U. D. Hanebeck  
Random hypersurface models for extended object tracking, in *Proceedings of the 9th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2009)*, December 2009.
- [8] M. Baum and U. D. Hanebeck  
Tracking an extended object modeled as an axis-aligned rectangle, in *Proceedings of Informatik 2009*, 2009.
- [9] M. Baum and U. D. Hanebeck  
Shape tracking of extended objects and group targets with star-convex rhms, in *Proceedings of the International Conference on Information Fusion (FUSION 2011)*, July 2011.
- [10] M. Baum, B. Noack, and U. D. Hanebeck  
Extended object and group tracking with elliptic random hypersurface models, in *Proceedings of the International Conference on Information Fusion (FUSION 2010)*, July 2010.
- [11] J. L. Crassidis and J. L. Junkins  
*Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC Press, 2004.
- [12] J. E. Cutting  
Reconceiving perceptual space, in *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space*. MIT Press, June 2003, ch. 11.
- [13] J. E. Cutting and P. M. Vishton  
Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth, in *Handbook of perception and cognition, Vol. 5; Perception of space and motion*. Academic Press, 1995, pp. 69–117.
- [14] K. Gilholm and D. Salmond  
Spatial distribution model for tracking extended objects, in *IEEE Proceedings—Radar, Sonar, and Navigation*, vol. 152, no. 5, October 2005, pp. 364–371.
- [15] K. Granstrom and O. Orguner  
A phd filter for tracking multiple extended targets using random matrices, *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5657–5671, 2012.
- [16] J. Hardy and M. Campbell  
Contingency planning over probabilistic obstacle predictions for autonomous road vehicles, *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 913–929, 2013.
- [17] F. Havlak and M. Campbell  
Discrete and continuous, probabilistic anticipation for autonomous robots in urban environments, *Transactions on Robotics*, vol. 30, no. 2, pp. 461–474, December 2013.
- [18] D. Held, J. Levinson, and S. Thrun  
Precision tracking with sparse 3d and dense color 2d data, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2013)*, May 2013.
- [19] D. Held, J. Levinson, S. Thrun, and S. Savarese  
Combining 3d shape, color, and motion for robust anytime tracking, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2014)*, June 2014.
- [20] E. Ilg, R. Kummerle, W. Burgard, and T. Brox  
Reconstruction of rigid body models from motion distorted laser range data using optical flow, in *Proceedings of the Robotics Science and Systems Conference (RSS 2014)*, July 2014.
- [21] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon  
Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera, in *ACM Symposium on User Interface Software and Technology*, October 2011. [Online]. Available: <http://research.microsoft.com/apps/pubs/default.aspx?id=155416>.
- [22] S. Julier and J. Uhlmann  
A new extension of the kalman filter to nonlinear systems, in *Proceedings of the 11th International Symposium on Aerospace/Defence*, 1997, pp. 401–422.
- [23] W. Koch  
Bayesian approach to extended object and cluster tracking using random matrices, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, 2008.
- [24] W. Koch and M. Feldmann  
Cluster tracking under kinematical constraints using random matrices, *Robotics and Autonomous Systems*, vol. 57, no. 3, pp. 296–309, 2009.
- [25] M. S. Landy, L. T. Maloney, E. Johnston, and M. Young  
Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Research*, vol. 35, no. 3, pp. 389–412, February 1995.
- [26] C. Lundquist, K. Granstrom, and O. Orguner  
Estimating object shape of targets with a phd filter, in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, July 2011, pp. 1–8.
- [27] R. Luo and M. Kay  
Data fusion and sensor integration: state-of-the-art, in *1990s, Data Fusion in Robotics and Machine Intelligence*, 1992, pp. 7–136.
- [28] M. McClelland, T. Estlin, and M. Campbell  
Qualitative relational mapping for planetary rover exploration, in *Proceedings of the AIAA Guidance, Navigation and Control Conference*, 2013.
- [29] I. Miller, M. Campbell, et al.  
Team cornell’s skynet: Robust perception and planning in an urban environment, *Journal of Field Robotics*, vol. 25, no. 8, pp. 493–527, 2008.
- [30] I. Miller, M. Campbell, and D. Huttenlocher  
Efficient, unbiased tracking of multiple dynamic obstacles under large viewpoint changes, *IEEE Transactions on Robotics*, vol. 27, no. 1, pp. 29–46, 2011.

- [31] A. Petrovskaya and S. Thrun  
Model based vehicle detection and tracking for autonomous urban driving,  
*Autonomous Robots Journal*, vol. 26, no. 2–3, pp. 123–139, 2009.
- [32] J. Saarinen, H. Andreasson, and A. Lilienthal  
Independent markov chain occupancy grid maps for representation of dynamic environment,  
in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, October 2012, pp. 3489–3495.
- [33] S. Thrun, W. Burgard, and D. Fox  
*Probabilistic Robotics*.  
MIT Press, 2006.
- [34] K. Wyffels and M. Campbell  
Modeling and fusing negative information for dynamic extended multi-object tracking,  
in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2013)*, May 2013.
- [35] K. Wyffels and M. Campbell  
Negative observations for multiple hypothesis tracking of dynamic extended objects,  
in *Proceedings of the American Controls Conference (ACC 2014)*, June 2014.
- [36] K. Wyffels and M. Campbell  
Joint tracking and non-parametric shape estimation of arbitrary extended objects,  
in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1–8.
- [37] K. Wyffels and M. Campbell  
Negative information for occlusion reasoning in dynamic extended multiobject tracking,  
*IEEE Transactions on Robotics*, vol. 31, no. 2, pp. 425–442, April 2015.



**Kevin Wyffels** received his B.S. degree from the University at Buffalo in 2005, M.S. degree from Rochester Institute of Technology in 2007, and M.S./Ph.D. degrees from Cornell University in 2014/2016. He is currently a research scientist at Ford Motor Company specializing in estimation, probabilistic inference, and robotic perception for fully autonomous vehicles. He is a member of IEEE.



**Mark Campbell** received his B.S. degree from CMU in Mechanical Engineering, and his M.S./Ph.D. degrees in Control and Estimation from MIT in 1993/1996. He is currently a Professor and the S. C. Thomas Sze Director of the Sibley School of Mechanical and Aerospace Engineering at Cornell University. In 2005–06, he was Visiting Scientist at the Insitu group, and an ARC International Fellow at the Australian Centre of Field Robotics. He received best paper awards from AIAA Propulsion and GNC conferences, and Frontiers in Education conference, and teaching awards from Cornell, University of Washington, and ASEE. He has been an invited speaker for the NAE Frontiers in Engineering Symposium and NAS Kavli Frontiers of Science Symposium. His research interests are in the areas of autonomous systems.

# Extended Object Tracking with Exploitation of Range Rate Measurements

STEVEN BORDONARO  
PETER WILLETT  
YAAKOV BAR-SHALOM  
MARCUS BAUM  
TOD LUGINBUHL

In active sonar and radar target tracking, measurements consist of position and often also include range rate. Tracking algorithms use these measurements over time to estimate target state comprising position, velocity and, where applicable, turn rate. In most cases there is an underlying assumption in the tracking algorithm that the target is a “point target” (i.e. the target has no physical extent). Another common assumption is that at most one measurement per scan originates from the target. For certain combinations of transmitted waveform and target type, the resolution of the waveform is such that the target is “over-resolved” (i.e. the sensor resolution is high enough that closely spaced scatter centers can be resolved). For such cases the point target assumption must be replaced with an extended target assumption. This work provides a methodology to exploit the extended nature of the target for the case of a rigid target whose spatial characteristics are fixed with respect to the line of motion. By employing a combination of the expectation maximization (EM) algorithm and allowing more than one measurement per scan to originate from the target, a technique is developed that uses a single scan of raw measurements that include range, bearing and range rate to provide an estimate of target position, velocity, heading and turn rate. This single scan estimate is then used in a nearly constant turn rate extended Kalman filter to provide a multi-scan estimate of the target state.

Manuscript received December 30, 2015; revised May 11, 2016; released for publication January 16, 2017.

Refereeing of this contribution was handled by Paolo Braca.

Authors’ addresses: S. Bordonaro and T. Luginbuhl, Naval Undersea Warfare Center, Newport, RI 02841 (E-mail: steven.bordonaro@navy.mil, tod.luginbuhl@navy.mil). P. Willett, Y. Bar-Shalom, and M. Baum, Dept. of Electrical and Computer Engr., University of Connecticut, Storrs, CT 06269-2157, USA (E-mail: willett@engr.uconn.edu, ybs@engr.uconn.edu, baum@engineer.uconn.edu).

1557-6418/17/\$17.00 © 2017 JAIF

## I. INTRODUCTION

In active sonar and radar target tracking systems, the goal is often to provide an estimate of the target’s state using measurements of range, bearing and range rate. Target dynamics are best modeled in Cartesian coordinates and consist of position, velocity and often include acceleration or turn rate. Common models for target dynamics are the nearly constant velocity, nearly constant acceleration and coordinated turn models [1].

In the formulation of the tracking algorithm it is common to assume that the target has no physical extent. This assumption is reasonable if the resolution of the transmitted waveform is greater than or equal to the size of the target. If, however, the resolution of the measurements is small enough that the spatial characteristics of the target can be measured, this “point target” assumption must be relaxed.

If the sensor is capable of resolving individual measurement sources within an extended target and detailed knowledge is available to model these sources, the target can be modeled as a set of discrete measurements sources within an extended object [2]. An alternative is to estimate the overall shape of the target as opposed to individual components. Within this shape estimation approach, numerous models exist. Two approaches that represent the extended target as an ellipse are [14], which uses symmetric, positively definite (SPD) random matrices; and the approach of [3] which employs a random hypersurface model (RHM). The RHM approach has been extended to more complex shapes in [2] by using star-convex RHMs. Irregular shapes are handled in [15] by using multiple (possibly overlapping) ellipses. Another approach to modeling spatial extent uses the assumption that the number of target measurements is Poisson distributed, with the measurement(s) drawn from a spatial distribution [9], [10], [11].

While these approaches are excellent and fairly liberal with regards to shape, a different approach is chosen here that aims to fully exploit the range rate measurements at the expense of using a somewhat more restrictive target model. The target model chosen in this research is that of a target “template” that characterizes the locations of target highlights (i.e. the active reflectors of the target). While the size and orientation of the target is unknown, the relationships of the highlight locations are assumed to be known a priori. It is also assumed that the target is rigid and has spatial characteristics that are fixed with respect to the line of motion. (The model can be viewed as a parameterized version of a “discrete” spatial distribution, as discussed in [5], [10]. The idea of using a set of reflectors can also be found in [13]; however, in [13] the relative positions of the reflectors are based on a known target size, while in the model proposed here, the size is unknown.) With this parametrized model, a single scan estimate of position, velocity, heading and turn rate can be made. This single scan estimate can then be utilized in a multi-scan tracker

(e.g. an extended Kalman filter) with a coordinated turn motion model (nearly constant turn rate) [1].

To provide the target estimate, the measurements from the extended target must be assigned to the individual target highlights. This is achieved by employing a combination of the EM algorithm and a version of the probabilistic multi-hypothesis tracker (PMHT) association model [18]. Unlike many tracking approaches, the PMHT (even for a single point target) does not assume there is at most a single measurement per target. There is therefore a natural compatibility between the PMHT and extended objects, which have multiple measurements per target. Also advantageous is that the algorithm is very flexible and easy to extend [22]. A pertinent example of this is that the PMHT has been successfully employed in extended object tracking using random matrices [20], [21]. The relationship of the PMHT association model with spatial distributions is also discussed in [10] and [9].

The combination of this target extent model and EM based estimation results in an algorithm with similar characteristics to one from a different field (image processing). This concept of aligning measured points to a template can be viewed as a version of surface registration. The iterative closest point algorithm (ICP) [23] is a common approach for surface registration. Its extension, the multi-scale EM-ICP [12], uses a similar formulation to the one proposed here; however, the approach of the present paper allows for the more general measurement error model needed for radar/sonar processing and utilizes range rate measurements. The novel aspect of the approach proposed here is the employment of a template based target model and utilization of existing techniques (EM, observed information matrix and the EKF) in an innovative way to exploit the extended nature of the target to improve state estimates.

The remainder of this paper is organized as follows: Section II introduces the model for the extended target and the measurements; Section III describes the approach for single and multi-scan estimation; Section IV provides a simulation of the algorithm and examines the resulting performance and Section V provides some concluding remarks. This paper is a continuation of the work presented in [7], with portions of [7] repeated here for continuity. This paper extends [7] by (i) modifying the measurement model to improve performance, (ii) providing an estimate of the converted measurement error covariance using the observed information matrix and (iii) utilizing the converted measurement in an extended Kalman filter.

## II. THE MODEL

### A. Extended Target Model with Discrete Reflectors

In active radar and sonar processing, the transmitted signal is reflected off the target and returns to the receiver, resulting in measurements of range, bearing and range rate. The reflections are due to a finite number

of strong reflectors, such as the nose and engines of an aircraft or the bow and sail of a submarine. For waveforms with high spatial resolution, it is possible to resolve the individual reflectors from the target as opposed to the integration of all the reflectors. In many cases there is general knowledge of the relative locations of the primary reflectors for a given target class (e.g. a military aircraft), that can reasonably represent a number of targets in that class. Using this premise, an extended target model approach can be developed as in [7].

The target is therefore represented as a set of  $M$  highlights (i.e. reflectors) forming a template for a general target. Each reflector,  $j = 1 \dots M$ , is specified with a probability of detection,  $\zeta_j$ , and a position in 2D Cartesian coordinates,

$$\mathbf{t}_j = \begin{bmatrix} x_r(j) \\ y_t(j) \end{bmatrix} \quad (1)$$

relative to the center of the target. While the shape of the target is known, the orientation,  $\psi$ , location (of the center),  $\mathbf{x} = [x \ y]^T$ , and size,  $s$ , are unknown.

An assumption is made that the direction of travel of the object is along the orientation,  $\psi$ , of the target (i.e. the plane flies forward, not sideways).<sup>1</sup> Furthermore, we assume the target is following a coordinated turn (nearly constant speed and turn rate) motion model. Using these assumptions the turn rate,  $\dot{\psi}$ , and speed,  $v$ , can be estimated using a single scan of data.

### B. Measurement Model

The measurement vector for a single scan of  $N$  measurements for time step  $k$  is

$$\mathbf{z}_{\text{RAW}_i}(k) = \begin{bmatrix} r_m(i,k) \\ \alpha_m(i,k) \\ \dot{r}_m(i,k) \end{bmatrix} \quad i = 1, \dots, N \quad (2)$$

where the measurement vector includes range,  $r$ , bearing,  $\alpha$ , and range rate,  $\dot{r}$ .

The measurement error for the raw measurements is assumed to be Gaussian with covariance matrix

$$\mathbf{R}_{\text{RAW}} = \begin{bmatrix} \sigma_r^2 & 0 & \rho\sigma_r\sigma_{\dot{r}} \\ 0 & \sigma_\alpha^2 & 0 \\ \rho\sigma_r\sigma_{\dot{r}} & 0 & \sigma_{\dot{r}}^2 \end{bmatrix} \quad (3)$$

where  $\sigma_r$ ,  $\sigma_\alpha$ , and  $\sigma_{\dot{r}}$  are the standard deviations of the range, bearing and range rate measurement noise. The correlation coefficient between the range and range rate measurement noise is  $\rho$ .

## III. ESTIMATION APPROACH

An overview of the new approach is shown in Fig.

1. First the raw measurements from a single scan are

<sup>1</sup>Note that, due to wind forces, the direction of movement of the aircraft, or track, is not necessarily same as the heading. Although this difference is neglected here, the difference may not be negligible in certain scenarios.

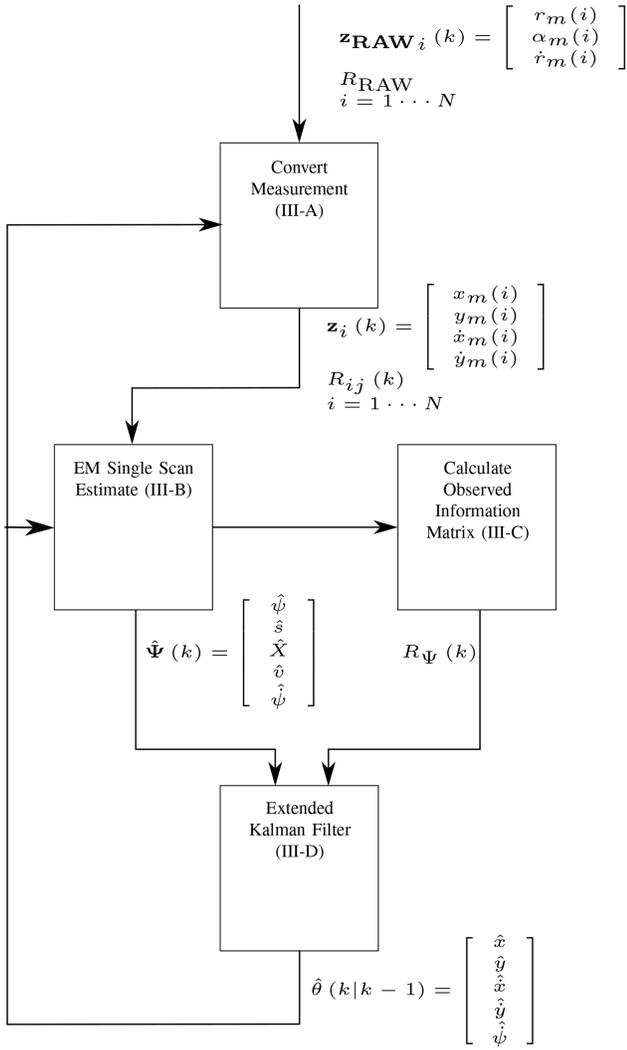


Fig. 1. Overview of the extended target tracking approach.

converted to Cartesian coordinates using the approach of [6]. These converted measurements are used in an EM algorithm for a single scan estimate of target position, size, heading, velocity and turn rate. The observed information matrix is calculated and used as a surrogate for the error covariance of this estimate. Finally, an extended Kalman filter with a coordinated turn motion model is used to combine the single scan estimates into a multi-scan estimate of the target state.

#### A. Measurement Conversion for an Individual Measurement

It is advantageous to first convert the raw measurements into Cartesian before processing. The raw measurements are converted into measurements of Cartesian position,  $x$  and  $y$ , and velocity,  $\dot{x}$  and  $\dot{y}$  using a simplified version of the method described in [6].

$$\mathbf{z}_i(k) = \begin{bmatrix} x_m(i,k) \\ y_m(i,k) \\ \dot{x}_m(i,k) \\ \dot{y}_m(i,k) \end{bmatrix} \quad (4) \quad \text{where}$$

$$\mathbf{z}_{\text{RAW}_i}(k) = \begin{bmatrix} r_m(i) \\ \alpha_m(i) \\ \dot{r}_m(i) \end{bmatrix} = e^{\sigma_a^2/2} \begin{bmatrix} r_m(i,k) \cos \alpha_m(i,k) \\ r_m(i,k) \sin \alpha_m(i,k) \\ \dot{r}_m(i,k) \cos \alpha_m(i,k) \\ \dot{r}_m(i,k) \sin \alpha_m(i,k) \end{bmatrix} \quad (5)$$

The conversion from range rate into Cartesian velocity assumes that the cross range rate is zero and accounts for any error in this assumption by setting the variance in the cross range rate dimension to infinity (or equivalently, setting the inverse to zero). This is implemented using the inverse converted measurement covariance,  $R_{ij}(k)^{-1}$ , which has a dimension of four by four, but is rank 3.

The converted measurement error covariance,  $R_{ij}(k)$ , is calculated according to Appendix A.

#### B. EM Single Scan Estimate from Multiple Measurements

1) Likelihood Model: Using the set of  $N$  measurements in combination with the target model, a single scan estimate of target position, size, heading, speed and turn rate can be calculated. The unknown parameters to be estimated form the vector  $\Psi$

$$\Psi = [\mathbf{x}^T \quad s \quad \psi \quad v \quad \dot{\psi}]^T \quad (6)$$

The following probabilistic model is used for the likelihood function of  $\Psi$ :

$$p_{\mathbf{z}}(\mathbf{z}_i | \Psi) = \sum_{j=1}^M \pi_j p_{ij}(\mathbf{z}_i | \Psi) \quad (7)$$

where,  $\pi_j$  is treated as the prior probability of a measurement originating from reflector  $j$  and  $p_{\mathbf{z}}$  is the conditional probability density for a single measurement given  $\Psi$ . This value is approximated using the probabilities of detection ( $\zeta_j$ ,  $j = 1, \dots, M$ ) by assuming each measurement comes from one of the reflectors, namely,

$$\pi_j = \frac{\zeta_j}{\sum_{l=1}^M \zeta_l} \quad (8)$$

The probability density function (pdf) for a given measurement-to-reflector combination,  $p_{ij}$  is given by

$$p_{ij}(\mathbf{z}_i | \Psi) = |2\pi R_{ij}|^{-1/2} \cdot \left\{ -\frac{1}{2} \nu_{ij}(\Psi, \mathbf{z}_i)^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \right\} \quad (9)$$

where  $\nu_{ij}(\Psi, \mathbf{z}_i)$ , the difference between measurement  $i$  and reflector  $j$ , is

$$\nu_{ij}(\Psi, \mathbf{z}_i) = \mathbf{z}_i - \begin{bmatrix} sD(\psi)\mathbf{t}_j + \mathbf{x} \\ v \cos \psi - s\dot{\psi}|\mathbf{t}_j| \sin(\psi + \theta_j) \\ v \sin \psi + s\dot{\psi}|\mathbf{t}_j| \cos(\psi + \theta_j) \end{bmatrix} \quad (10)$$

$$D(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \quad (11)$$

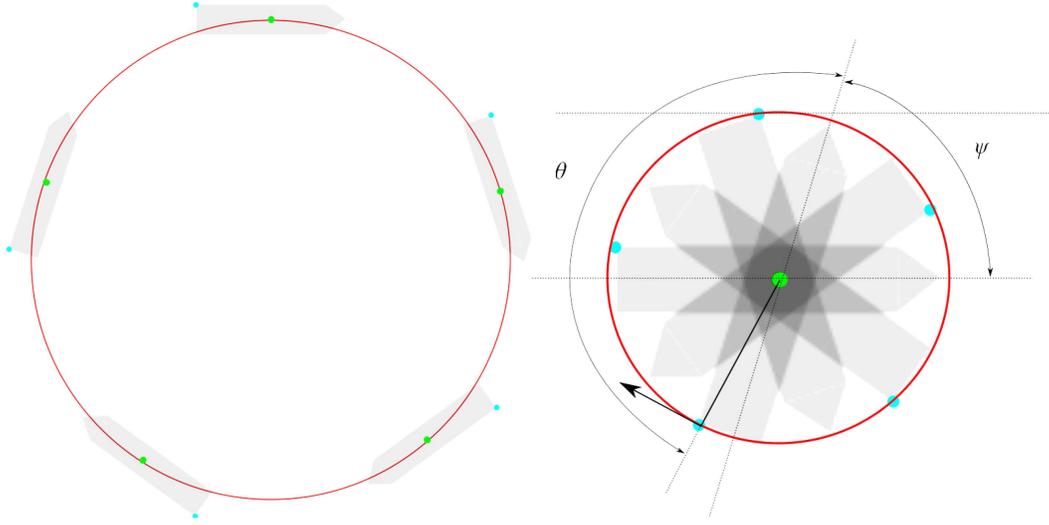


Fig. 2. Turn rate contribution to range rate.

is the rotation matrix,

$$|\mathbf{t}_j| = \sqrt{x_t(j)^2 + y_t(j)^2} \quad (12)$$

is the distance from reflector  $j$  to the target center,

$$\theta_j = \text{atan2}(y_t(j), x_t(j)) \quad (13)$$

is the angle of the line from the center to reflector  $j$ ; relative to the reference direction, and  $R_{i,j}$  is the converted measurement error covariance matrix (see appendix A).

The term  $sD(\psi)\mathbf{t}_j + \mathbf{x}$  provides the position of target highlight  $j$ , scaled by the size,  $s$ , rotated by the heading  $\psi$ , and translated by the position of the target center  $\mathbf{x}$ .

In order to simultaneously estimate target speed (along its heading) and turn rate, the contribution of these terms to the measured instantaneous velocity must be separated. The terms  $v \cos \psi$  and  $v \sin \psi$  are the contributions of the target center's velocity to measured velocity. The terms  $s\psi|\mathbf{t}_j| \sin(\psi + \theta_j)$  and  $s\psi|\mathbf{t}_j| \cos(\psi + \theta_j)$  are the contributions due to turn rate. Fig. 2 shows the path of the target on the left [7]. When the motion of the target center is removed (as shown on the right), the motion of the individual highlights due to turn rate is evident.

The incomplete-data log-likelihood of  $\Psi$  based on all the measurements  $\mathcal{Z}$  is given by [4]:

$$\begin{aligned} \ln \mathcal{L}(\Psi; \mathcal{Z}) &= \ln p_{\mathcal{Z}}(\mathcal{Z} | \Psi) \\ &= \ln \prod_{i=1}^N p_{\mathcal{Z}}(\mathbf{z}_i | \Psi) \\ &= \sum_{i=1}^N \ln \left( \sum_{j=1}^M \pi_j p_{ij}(\mathbf{z}_i | \Psi, \mathbf{t}_j) \right) \end{aligned} \quad (14)$$

where  $p_{\mathcal{Z}}$  is the conditional probability density of the set of measurements  $\mathcal{Z}$ , given  $\Psi$ . For each measurement,  $\mathbf{z}_i$ ,

one has here the summation of its pdf if originated from reflector  $j$  and weighted by  $\pi_j$ .

2) Solving for  $\Psi$ : To estimate  $\Psi$ , one can find the vector that maximizes (14). The difficulty with (14) is the log of a sum. However, by recognizing (7) as a mixture model, the problem can be approached with the EM algorithm. The inside summation can be rewritten according to the EM approach using binary multipliers as missing data. The ‘‘missing’’ data are association variables that declare which reflector produced each measurement [16]. These association variables are expressed as binary vectors where each element in the binary vector corresponds to a reflector.

The binary vectors are defined as

$$\mathcal{Y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_N^T]^T \quad (15)$$

where  $\mathbf{y}_i = [y_{i1}, \dots, y_{iM}]^T$  is a  $M$ -dimensional binary vector (0 or 1), such that  $y_{ij}$  is one if measurement  $i$  is a reflection from reflector  $j$ , and zero otherwise. Each vector  $\mathbf{y}_i$  contains only one nonzero element. The complete log-likelihood, based also on  $\mathcal{Y}$  is

$$\begin{aligned} \ln \mathcal{L}_c(\Psi; \mathcal{Z}, \mathcal{Y}) &= \ln p_c(\mathcal{Z}, \mathcal{Y} | \Psi) \\ &= \sum_{i=1}^N \ln \left( \sum_{j=1}^M y_{ij} \pi_j p_{ij}(\mathbf{z}_i | \Psi) \right) \\ &= \sum_{i=1}^N \sum_{j=1}^M y_{ij} \ln(\pi_j p_{ij}(\mathbf{z}_i | \Psi)) \end{aligned} \quad (16)$$

where  $p_c$  is the conditional probability density of the complete data,  $\mathcal{Z}$  and  $\mathcal{Y}$ , given  $\Psi$ . If we view the missing data,  $\mathcal{Y}$ , as random variables, the EM  $\mathcal{Q}$  function can now be found. In the EM algorithm, the  $\mathcal{Q}$  function is iteratively maximized. This function is the expectation of the complete log-likelihood, with the expectation operation conducted with respect to the unknown data

$\mathcal{Y}$ , given the observed data,  $\mathcal{Z}$ , and the estimate of  $\Psi$  from the previous iteration,  $\Psi^{(l)}$ , namely,

$$\mathcal{Q}(\Psi; \Psi^{(l)}, \mathcal{Z}) = E\{\ln \mathcal{L}_c(\Psi; \mathcal{Z}, \mathcal{Y}) \mid \mathcal{Z}, \Psi^{(l)}\} \quad (17)$$

$$\begin{aligned} \mathcal{Q}(\Psi; \Psi^{(l)}, \mathcal{Z}) &= \sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \ln(\pi_j p_{ij}(\mathbf{z}_i \mid \Psi)) \\ &= \sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \\ &\quad \left[ \ln(\pi_j) - \frac{1}{2} \ln(|2\pi R_{ij}|) \right. \\ &\quad \left. - \frac{1}{2} \nu_{ij}(\Psi, \mathbf{z}_i)^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \right] \quad (18) \end{aligned}$$

where  $w_{ij}$  is the estimate of the posterior association probabilities  $y_{ij}$  given the measurements and the previous estimate  $\Psi^{(l)}$ , allowing for more than one measurement to be a reflection from a single reflector. Since this association model allows for more than one measurement to be a reflection from a single reflector, the model is an application of the PMHT association model [18]. The association probabilities are

$$\begin{aligned} w_{ij}(\Psi^{(l)}, \mathbf{z}_i) &= p_y(y_{ij} \mid \mathbf{z}_i, \Psi^{(l)}) \\ &= \frac{\pi_j p_{ij}(\mathbf{z}_i \mid \Psi^{(l)})}{\sum_{m=1}^M \pi_m p_{im}(\mathbf{z}_i \mid \Psi^{(l)})} \quad (19) \end{aligned}$$

where  $p_y$  is the conditional probability of an association pair, given  $\Psi^{(l)}$  and measurement  $\mathbf{z}_i$ . The  $w_{ij}$  calculation given above assumes a clutter free environment. The extension to a cluttered environment is quite straightforward and simply requires an additional clutter distribution term in the denominator of the expression for  $w_{ij}$  [18] and the appropriate modification to  $\mathcal{Y}$ .

For the M step of EM, the  $\mathcal{Q}$  function is maximized with respect to  $\Psi$ . The  $\Psi$  that maximizes (18) can be found by solving

$$\nabla_{\Psi} \mathcal{Q}(\Psi; \Psi^{(l)}, \mathcal{Z}) = 0 \quad (20)$$

to yield  $\Psi^{(l+1)}$ , where

$$\begin{aligned} \nabla_{\Psi} \mathcal{Q}(\Psi; \Psi^{(l)}, \mathcal{Z}) &= -\frac{1}{2} \left[ \nabla_{\Psi} \sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \right. \\ &\quad \left. \cdot \nu_{ij}(\Psi, \mathbf{z}_i)^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \right] \quad (21) \end{aligned}$$

Since  $R_{ij}^{-1}$  is symmetric and using

$$\nabla_{\mathbf{x}} \{f(\mathbf{x}^T) A f(\mathbf{x})\} = 2(\nabla_{\mathbf{x}} f(\mathbf{x}))^T A f(\mathbf{x}) \quad (22)$$

one can simplify (21)

$$\begin{aligned} \nabla_{\Psi} \mathcal{Q}(\Psi; \Psi^{(l)}, \mathcal{Z}) &= -\sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \\ &\quad \cdot (\nu'_{ij}(\Psi))^{T} R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (23) \end{aligned}$$

where

$$\nu'_{ij}(\Psi) = \nabla_{\Psi} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (24)$$

The components of  $\nu'_{ij}(\Psi)$  are

$$\nu'_{ij}(\Psi) = - \begin{bmatrix} 1 & 0 & a_{13} & a_{14} & 0 & 0 \\ 0 & 1 & a_{23} & a_{24} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & a_{35} & a_{36} \\ 0 & 0 & a_{34} & a_{44} & a_{45} & a_{46} \end{bmatrix} \quad (25)$$

where

$$\begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{34} \end{bmatrix} = \begin{bmatrix} D(\psi) \mathbf{t}_j \\ -\dot{\psi} |\mathbf{t}_j| \sin(\psi + \theta_j) \\ \dot{\psi} |\mathbf{t}_j| \cos(\psi + \theta_j) \end{bmatrix} \quad (26)$$

$$\begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \\ a_{44} \end{bmatrix} = \begin{bmatrix} sD'(\psi) \mathbf{t}_j \\ -v \sin \psi - s\dot{\psi} |\mathbf{t}_j| \cos(\psi + \theta_j) \\ v \cos \psi - s\dot{\psi} |\mathbf{t}_j| \sin(\psi + \theta_j) \end{bmatrix} \quad (27)$$

$$\begin{bmatrix} a_{35} \\ a_{45} \end{bmatrix} = \begin{bmatrix} \cos \psi \\ \sin \psi \end{bmatrix} \quad (28)$$

$$\begin{bmatrix} a_{36} \\ a_{46} \end{bmatrix} = \begin{bmatrix} -s|\mathbf{t}_j| \sin(\psi + \theta_j) \\ s|\mathbf{t}_j| \cos(\psi + \theta_j) \end{bmatrix} \quad (29)$$

and

$$D'(\psi) = \begin{bmatrix} -\sin \psi & -\cos \psi \\ \cos \psi & -\sin \psi \end{bmatrix} \quad (30)$$

Since (20) cannot be solved directly, a first order Taylor expansion is used to find  $\Psi^{(l+1)}$ , the maximizing  $\Psi$ , given  $\Psi^{(l)}$ , namely,

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \nu'_{ij}(\Psi)^T R_{ij}^{-1} \\ \cdot [\nu_{ij}(\Psi, \mathbf{z}_i) + \nu'_{ij}(\Psi)(\Psi^{(l+1)} - \Psi^{(l)})] = 0 \quad (31) \end{aligned}$$

which leads to

$$\begin{aligned} \Psi^{(l+1)} &= \Psi^{(l)} \\ &+ \left( \sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \nu'_{ij}(\Psi)^T R_{ij}^{-1} \nu'_{ij}(\Psi) \right)^{-1} \\ &\cdot \left( \sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(l)}, \mathbf{z}_i) \nu'_{ij}(\Psi)^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \right) \Big|_{\Psi^{(l)}} \quad (32) \end{aligned}$$

The resulting EM algorithm is defined as follows:

- 1) Initialize  $\Psi^{(l)}$
- 2) Calculate  $w$  using (19)
- 3) Solve for  $\Psi^{(l+1)}$  using (32).
- 4) Iteratively repeat steps 2 and 3 until a convergence criterion is met (e.g. when the increase in the complete log-likelihood is below a threshold).
- 5) Set  $\hat{\Psi}(k) = \Psi^{(L)}$  at the last iteration,  $l = L$ .

### C. Observed Information Matrix

In order to provide a measure of uncertainty for the estimate  $\hat{\Psi}$ , critical information for tracking, the observed information matrix is used as a surrogate for the inverse covariance matrix. Oakes' formula [17] for the observed information matrix is used (see Appendix B).

$$-\nabla_{\Psi} \nabla_{\Psi}^T \ln \mathcal{L}(\Psi; \mathcal{Z}) = -[\nabla_{\Psi} \nabla_{\Psi}^T Q(\Psi; \Psi^{(L)}, \mathcal{Z}) + \nabla_{\Psi} \nabla_{\Psi^{(L)}}^T Q(\Psi; \Psi^{(L)}, \mathcal{Z})] \quad (33)$$

Evaluating (33) using  $\Psi = \Psi^{(L)}$  results in the "observed information matrix,"  $I(\hat{\Psi}; \mathcal{Z})$  [16].

The first term on the right hand side of (33) is the observed information if the associations were known,  $I_c(\hat{\Psi}; \mathcal{Z})$ :

$$I_c(\hat{\Psi}; \mathcal{Z}) = -\nabla_{\Psi} \nabla_{\Psi}^T Q(\Psi; \Psi^{(L)}, \mathcal{Z}) \quad (34)$$

$$= -\sum_{i=1}^N \sum_{j=1}^M w_{ij}(\Psi^{(L)}, \mathbf{z}_i) \cdot [(\nu'_{ij}(\Psi))^T R_{ij}^{-1} \nu'_{ij}(\Psi) + B_{ij}(\Psi, \mathbf{z}_i)] \quad (35)$$

where the  $B$  matrix is based on the second derivative of  $\nu$ ,

$$B_{ij}(\Psi, \mathbf{z}_i) = - \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{34} & 0 & b_{36} \\ 0 & 0 & b_{34} & b_{44} & b_{45} & b_{46} \\ 0 & 0 & 0 & b_{45} & 0 & 0 \\ 0 & 0 & b_{36} & b_{46} & 0 & 0 \end{bmatrix} \quad (36)$$

with components:

$$b_{34} = \begin{bmatrix} D'(\psi) \mathbf{t}_j \\ -\dot{\psi} |\mathbf{t}_j| \cos(\psi + \theta_j) \\ -\dot{\psi} |\mathbf{t}_j| \sin(\psi + \theta_j) \end{bmatrix}^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (37)$$

$$b_{36} = \begin{bmatrix} 0 \\ 0 \\ -|\mathbf{t}_j| \sin(\psi + \theta_j) \\ |\mathbf{t}_j| \cos(\psi + \theta_j) \end{bmatrix}^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (38)$$

$$b_{44} = \begin{bmatrix} -sD(\psi) \mathbf{t}_j \\ -v \cos \psi + s\dot{\psi} |\mathbf{t}_j| \sin(\psi + \theta_j) \\ -v \sin \psi - s\dot{\psi} |\mathbf{t}_j| \cos(\psi + \theta_j) \end{bmatrix}^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (39)$$

$$b_{45} = \begin{bmatrix} 0 \\ 0 \\ -\sin \psi \\ \cos \psi \end{bmatrix}^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (40)$$

$$b_{46} = \begin{bmatrix} 0 \\ 0 \\ -s|\mathbf{t}_j| \cos(\psi + \theta_j) \\ -s|\mathbf{t}_j| \sin(\psi + \theta_j) \end{bmatrix}^T R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (41)$$

The second term on the right hand side of (33) accounts for the association uncertainty,  $I_m(\hat{\Psi}; \mathcal{Z})$ . This is found by taking the derivative of  $Q$  with respect to  $\Psi$ , and taking the derivative with respect to  $\Psi^{(L)}$ , i.e.

$$I_m(\hat{\Psi}; \mathcal{Z}) = -\nabla_{\Psi} \nabla_{\Psi^{(L)}}^T Q(\Psi; \Psi^{(L)}, \mathcal{Z})$$

$$= -\sum_{i=1}^N \sum_{j=1}^M w'_{ij}(\Psi^{(L)}, \mathbf{z}_i) (\nu'_{ij}(\Psi))^{(L)T} R_{ij}^{-1} \nu_{ij}(\Psi, \mathbf{z}_i) \quad (42)$$

where  $w'$  is the derivative of (19)

$$w'_{ij}(\Psi^{(L)}, \mathbf{z}_i) = \pi_j p_{ij}(\mathbf{z}_i | \Psi^{(L)}) \quad (43)$$

$$\left\{ \left( \sum_{m=1}^M \pi_m p_{im}(\mathbf{z}_i | \Psi^{(L)}) \right)^{-2} \cdot \sum_{m=1}^M [\pi_m p_{im}(\mathbf{z}_i | \Psi^{(L)}) \cdot (\nu'_{im}(\Psi^{(L)}))^{(L)T} R_{im}^{-1} \nu_{im}(\Psi^{(L)}, \mathbf{z}_i)] - \frac{(\nu'_{ij}(\Psi^{(L)}))^{(L)T} R_{ij}^{-1} \nu_{ij}(\Psi^{(L)}, \mathbf{z}_i)}{\sum_{m=1}^M \pi_m p_{im}(\mathbf{z}_i | \Psi^{(L)})} \right\}$$

### D. Extended Kalman Filter for Multi-Scan Estimation

The single scan estimate of the target state can be used in an EKF to provide multi-scan estimates. The EKF for a coordinated turn motion model is well known for the case of position only measurements (pp. 466–470 of [1]). The state vector for the CT-EKF is

$$\hat{\boldsymbol{\theta}} = [x_{\hat{\theta}} \quad \dot{x}_{\hat{\theta}} \quad y_{\hat{\theta}} \quad \dot{y}_{\hat{\theta}} \quad \psi_{\hat{\theta}}]^T \quad (44)$$

Note that the subscript  $\hat{\boldsymbol{\theta}}$  is used to avoid confusion between the elements of  $\hat{\boldsymbol{\theta}}$  and  $\Psi$ .

The dynamic equation is

$$\boldsymbol{\theta}(k+1) = f[k, \boldsymbol{\theta}(k)] + \Gamma(k) v(k) \quad (45)$$

and the state prediction is:

$$\hat{\boldsymbol{\theta}}(k+1 | k) = f[k, \hat{\boldsymbol{\theta}}(k | k)] \quad (46)$$

where  $f[k, \hat{\boldsymbol{\theta}}(k | k)]$  is:

$$\begin{bmatrix} 1 & \frac{\sin(\dot{\psi}_{\hat{\theta}}(k)T)}{\dot{\psi}_{\hat{\theta}}(k)} & 0 & -\frac{1 - \cos(\dot{\psi}_{\hat{\theta}}(k)T)}{\dot{\psi}_{\hat{\theta}}(k)} & 0 \\ 0 & \cos(\dot{\psi}_{\hat{\theta}}(k)T) & 0 & -\sin(\dot{\psi}_{\hat{\theta}}(k)T) & 0 \\ 0 & \frac{1 - \cos(\dot{\psi}_{\hat{\theta}}(k)T)}{\dot{\psi}_{\hat{\theta}}(k)} & 1 & \frac{\sin(\dot{\psi}_{\hat{\theta}}(k)T)}{\dot{\psi}_{\hat{\theta}}(k)} & 0 \\ 0 & \sin(\dot{\psi}_{\hat{\theta}}(k)T) & 0 & \cos(\dot{\psi}_{\hat{\theta}}(k)T) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \hat{\boldsymbol{\theta}}(k | k) \quad (47)$$

and

$$\Gamma(k) = \begin{bmatrix} \frac{1}{2}T^2 & 0 & 0 \\ T & 0 & 0 \\ 0 & \frac{1}{2}T^2 & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{bmatrix} \quad (48)$$

The state prediction covariance is:

$$P(k+1|k) = F(k)P(k|k)F(k)^T + \Gamma(k)Q(k)\Gamma(k)^T \quad (49)$$

where  $Q$  is the covariance of the process noise,  $v$ , and

$$F(k) = \left. \frac{\partial f(k)}{\partial \theta} \right|_{\theta=\hat{\theta}(k+1|k)} \quad (50)$$

Two modifications to the CT-EKF in [1] are necessary for this application. This first is in the observation function. Using the single scan estimate,  $\hat{\Psi}$ , as the observation, the full state vector can be observed, thus eliminating the need for an observation matrix (commonly referred to as the  $H$  matrix, i.e., here  $H$  is the identity matrix). The single scan observation,  $\mathbf{z}_C$  and the inverse observation error covariance,  $R_C^{-1}$ , are:

$$\mathbf{z}_C(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos \psi & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sin \psi & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \Psi(k) \quad (51)$$

$$R_C(k)^{-1} = (A^{-1})^T \mathbf{I}(\hat{\Psi}; \mathcal{Z}) A^{-1} \quad (52)$$

where  $\mathbf{I}(\hat{\Psi}; \mathcal{Z})$  was defined following (33).

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sin \psi & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos \psi & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (53)$$

A second modification is required since  $R_C^{-1}$  is not necessarily invertible. The inverse error covariance matrices for the individual measurements,  $R_{ij}^{-1}$ , are not invertible due to the fact that the information related to cross range velocity is zero (it has a zero eigenvalue in the cross range rate direction). Although,  $R_C^{-1}$  will be invertible for most scans, it is not invertible if the target aspect is  $90^\circ$ , or if there is only one measurement in the scan. To allow for this possibility, the information form

of the EKF is utilized. The EKF update is:

$$W(k+1) = [P(k+1|k)^{-1} + R_C(k+1)^{-1}]^{-1} \cdot R_C(k+1)^{-1} \quad (54)$$

$$P(k+1|k+1) = [P(k+1|k)^{-1} + R_C(k+1|k)^{-1}]^{-1} \quad (55)$$

$$\hat{\theta}(k+1|k+1) = \hat{\theta}(k+1|k) + W(k+1) \cdot [\mathbf{z}_C(k) - \hat{\theta}(k+1|k)] \quad (56)$$

## IV. IMPLEMENTATION AND RESULTS

### A. Implementation

1) EM Initialization: As with any optimization approach, care must be taken when employing the algorithm during initialization to avoid convergence to a local maximum. The initialization approach chosen here is as follows.

The initial value for  $\mathbf{x}^{(0)}$  is simply the mean of all the position measurements. The initial value for size,  $s^{(0)}$ , is set to the ratio of the average distance from the measurement to  $\mathbf{x}^{(0)}$  and the average distance of the target highlight,  $\mathbf{t}$ , to the target center. The initial value for heading,  $\psi^{(0)}$ , is calculated by finding the covariance of the position measurements and estimating the heading based on the largest eigenvector. The initial speed and turn rate are set to 0.

A particular concern for local maximums for many target models is one at a heading of 180 degrees from the true heading. To avoid maximizing at this incorrect heading, the algorithm is optimized using two initial headings, 180 degrees apart, and the result with the highest likelihood is used.

Even with proper initialization, converging to a local instead of global maximum is a concern. To help, the  $R$  matrix is artificially inflated for the first few iterations of the algorithm. This tends to smooth the likelihood surface. Optimization on the augmented surface first reduces the probability of converging to a local maximum. This approach is related to the deterministic annealing EM algorithm [19].

2) Limitations: It is important to note the limitations of the algorithm in its ability to estimate velocity and turn rate. Regardless of the target model, the ability to estimate velocity from range rate measurements will be limited when the target is traveling across the line of sight from the sensor. Turn rate estimation will also be limited for targets that do not have significant width when the target is traveling directly towards or away from the sensor. To analyze these effects the Cramer-Rao low bound (CRLB) is examined for two target types. Evident in (33) is that the CRLB is a function of the measurements. The looser bound of (35) is used here and is calculated using the expected measurement. The bound provides a lower bound on the average square

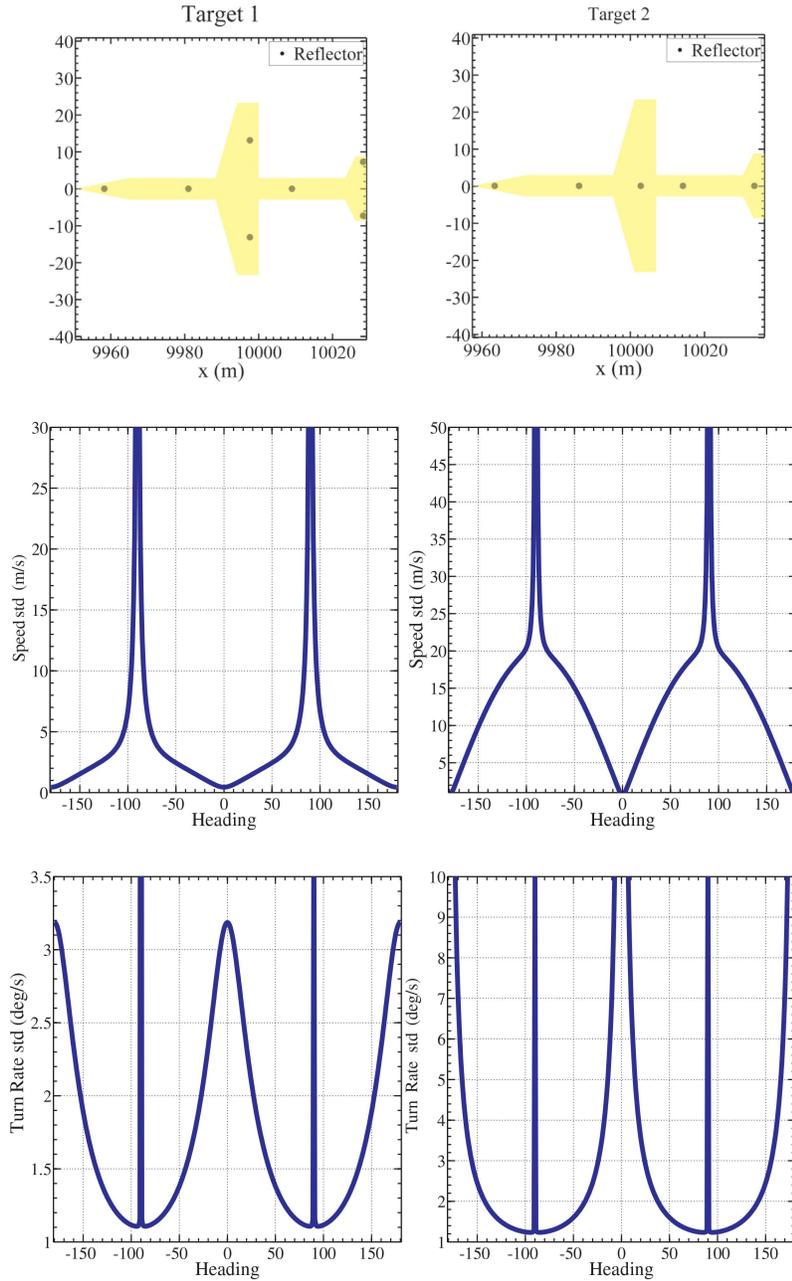


Fig. 3. CRLB analysis (using (35)) for a target that has width (target 1) and one without width (target 2).

error, but is looser than the CRLB due to the fact that it does not consider assignment uncertainty. Nevertheless, this bound is sufficient to demonstrate the limitations in the algorithm at various aspect angles. For this test the heading was varied from  $-180$  to  $180$  degrees, the probability of detection was set to 1, the size was set to  $70$  m, position set to  $[10 \ 0]^T$  km, the speed set  $120$  m/s and the turn-rate set to  $3$  deg/sec. The measurement error covariance was set as follows:

- 1)  $\sigma_r = 2$  m
- 2)  $\sigma_{\dot{r}} = 1$  m/s
- 3)  $\sigma_\alpha = 0.05$  deg
- 4)  $\rho\sigma_r\sigma_{\dot{r}} = 0$

As seen from Figure 3, if the target has width then turn-rate and speed can be estimated at all aspects with the exception of  $\pm 90$  degrees. In the case of a line-like target, such as target 2 in Figure 3, speed can be estimated at all aspects with the exception of aspects near  $\pm 90$  degrees, while turn-rate cannot be estimated at  $\pm 90$  degrees and near  $0$  or  $180$  degrees.

3) Implementation Details: There are three notable implementation details that are required for robust performance of the algorithm. The first is dealing with the inability to estimate velocity when the target aspect is  $90$  deg. Since the true error covariance of the single scan estimate is unknown, the observed information matrix serves as a surrogate. When the true target aspect is

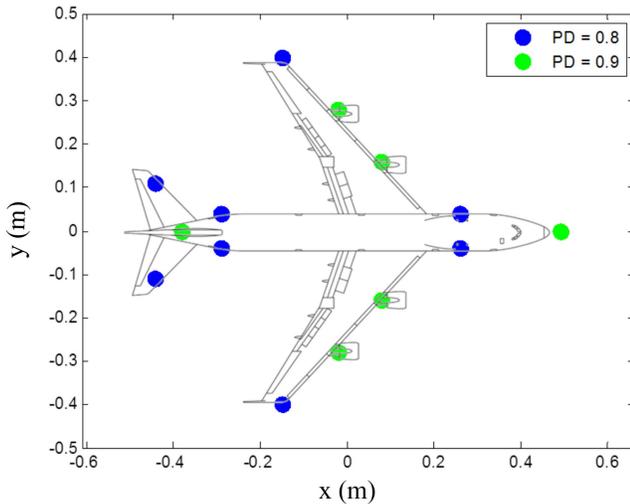


Fig. 4. Target template

90 deg, while the observed aspect is near, but not equal to 90, the observed information matrix will be overconfident in the velocity estimate. To avoid this, when the estimated aspect, based on  $\hat{\theta}(k+1|k)$ , is near 90 degrees, the velocity estimate should not be used. This is achieved by setting the appropriate rows and columns of  $R_C(k)^{-1}$  to zero. (For targets with little or no width, a similar test is required for turn-rate estimation at aspects near 0 or 180 deg.)

A related issue is that when the estimated target aspect is close to 90 deg, components of  $R_C(k)^{-1}$  may be close to zero, resulting in a badly conditioned matrix. In these cases, only the position portion of the single scan estimate is used. (Note that the first issue occurs when the true target aspect is 90 deg, while the second issue is when the estimated target aspect is close to 90 deg).

Finally, since the EM algorithm may converge on a local maximum, gating is used to validate the single scan estimate based on the innovation in the EKF update. If the innovation for either the velocity or turn-rate is too large, only the position portion of the single scan estimate is used. Again, this is achieved by setting the appropriate rows and columns of  $R_C(k)^{-1}$  to zero.

## B. Results

The new algorithm was tested in a aircraft tracking application. The target template is based on a commercial airliner (see Fig. 4), with probability of detections for the highlights at 0.8 and 0.9. The aircraft follows the path shown in Fig. 5. The measurement error covariance was set the same as Section IV-A.2. The EKF is implemented assuming the following process noise:

$$Q = \begin{bmatrix} (0.25)^2 & 0 & 0 \\ 0 & (0.25)^2 & 0 \\ 0 & 0 & \left(0.6 \frac{\pi}{180}\right)^2 \end{bmatrix} \quad (57)$$

Fig. 6 shows the average normalized estimation error squared (ANEES) [1] and mean square error

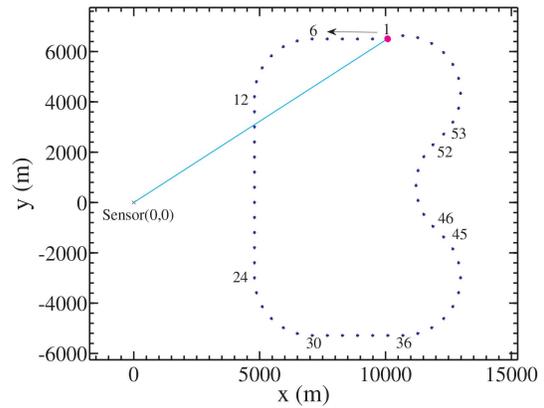


Fig. 5. Target path for the test case. The position of the target at each scan is shown. The first scan, as well as any scan that is starting a maneuver is labeled. The sensor position is at the origin. The line-of-sight for the first scan is also shown.

for position, velocity and turn rate. Errors are shown for the state prediction ( $\hat{\theta}(k|k-1)$ ) for the algorithm (EXTGTEKF). Since the primary advantage of the proposed algorithm is the exploitation of the target shape to extract speed and turn rate, a tracker that does not extract these quantities is used for comparison (POSEKF). The POSEKF is identical to the EXTGTEKF, with the important exception that only the position portion of the observation,  $z_C$ , is used. This is achieved by setting the appropriate rows and columns of  $R_C(k)^{-1}$  to zero. The proposed algorithm exhibits better consistency (ANEES closer to 1) and, in general, improved mean square error (MSE). Unlike the POSEKF algorithm, the EXTGTEKF does not lag in the turn-rate estimate since turn-rate is measured directly. The turn rate estimate for the EXTGTEKF is significantly better when the turn initiates, but it worse in steady state. This is due to the fact that the turn rate estimate for the POSEKF requires three position measurements, resulting in a smoother estimate. It is a trade-off between lag and smoothing. Performance of the EXTGTEKF is, as expected, degraded for target aspects near 90 deg. (e.g. near scan 19), as the EXTGTEKF reverts to position only measurements during those periods.

## V. CONCLUSION

A novel approach to extended object tracking has been presented. A target model has been developed for the target spatial characteristics that is appropriate for estimation, flexible enough to handle various target types, and loose enough such that exact knowledge of the target size is not required. By restricting the spatial characteristics to be fixed with respect to the line of motion, the resulting algorithm allows for single scan estimation of position, heading, size, velocity and turn rate by using measurements of position and range rate. These single scan measurements when used in a multi-scan tracking algorithm (i.e. extended Kalman filter) provide improved estimates of target position, velocity

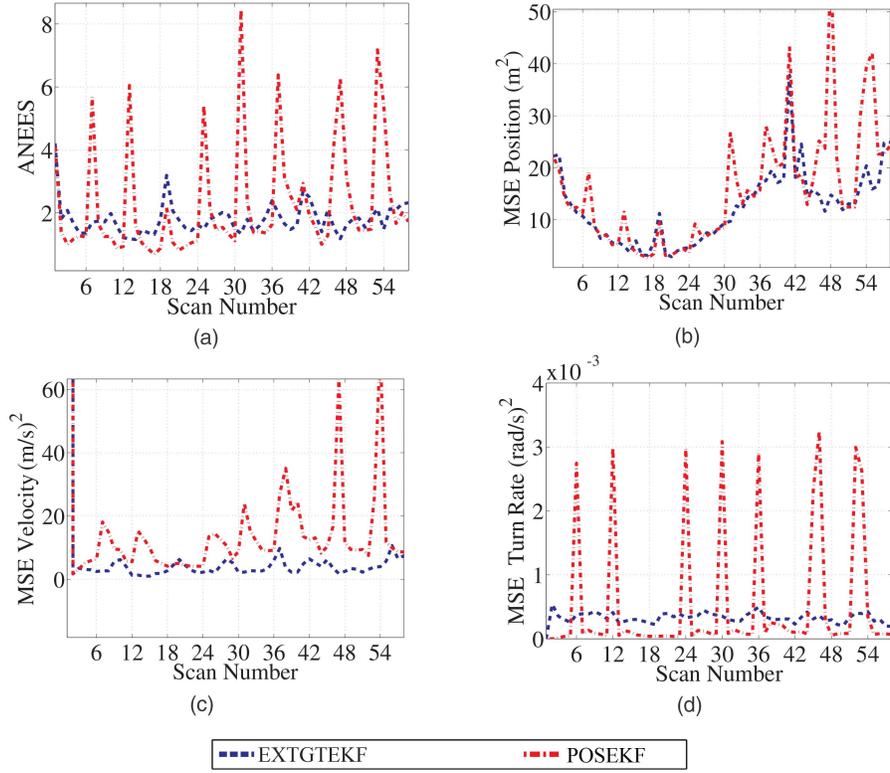


Fig. 6. Results of a 500 run Monte Carlo evaluation of the estimated state for the new EXTGTEKF and a position-only cluster tracker (CLUSTEREKF). a. Average Normalized Estimation Error Squared (ANEES). b. Mean Square Error (MSE) for Position Estimate. c. Mean Square Error (MSE) for Velocity Estimate. d. Mean Square Error (MSE) for Turn Rate Estimate.

and turn rate compared to a traditional cluster tracker using only position measurements. A primary advantage is that the new method, unlike methods using only position measurements, does not suffer from a lag in the estimation of turn rate and the resulting estimation errors.

#### APPENDIX A CONVERTED MEASUREMENT ERROR COVARIANCE

The converted measurement error covariance is approximated using a simplification of [6]. The calculation requires a prediction, which is based on the one step prediction,  $\hat{\theta}(k | k-1)$ . Using this prediction, in combination with the target template and the previous estimate of the size,  $s$ , the state of an individual highlight can be calculated (which will be referred to as  $\mathbf{x}_j$ ).

First the predicted highlight state is rotated into the estimate's line of sight (LOS) coordinate system. Noting that the inverse of the direction cosine matrix,  $D(\alpha_m)$ , is its transpose, the rotated state is calculated as:

$$\hat{\mathbf{x}}_R = D(\alpha_t)^T \hat{\mathbf{x}}_j \quad (58)$$

where the predicted bearing to the highlight is

$$\alpha_t = \tan^{-1} \left( \frac{\hat{\mathbf{x}}_{k+1|k}^2}{\hat{\mathbf{x}}_{k+1|k}^1} \right) \quad (59)$$

and  $\hat{\mathbf{x}}^n$  is the  $n$ th component of  $\hat{\mathbf{x}}$ .

$$R_R^{11} = \frac{1}{2} [(\hat{\mathbf{x}}_R^1)^2 + \sigma_r^2] (1 + e^{-2\sigma_\alpha^2}) e^{\sigma_\alpha^2} - (\hat{\mathbf{x}}_R^1)^2 \quad (60)$$

$$R_R^{12} = 0 \quad (61)$$

$$R_R^{13} = \frac{1}{2} (\hat{\mathbf{x}}_R^1 \hat{\mathbf{x}}_R^3 + \rho \sigma_r \sigma_f) (1 + e^{-2\sigma_\alpha^2}) e^{\sigma_\alpha^2} - \hat{\mathbf{x}}_R^1 \hat{\mathbf{x}}_R^3 \quad (62)$$

$$R_R^{22} = \frac{1}{2} [(\hat{\mathbf{x}}_R^1)^2 + \sigma_r^2] (1 - e^{-2\sigma_\alpha^2}) e^{\sigma_\alpha^2} \quad (63)$$

$$R_R^{23} = \frac{1}{2} (\hat{\mathbf{x}}_R^1 \hat{\mathbf{x}}_R^4) (1 - e^{-2\sigma_\alpha^2}) e^{\sigma_\alpha^2} \quad (64)$$

$$R_R^{33} = \frac{1}{2} [(\hat{\mathbf{x}}_R^3)^2 + \sigma_f^2] (1 + e^{-2\sigma_\alpha^2}) e^{\sigma_\alpha^2} - (\hat{\mathbf{x}}_R^3)^2 + \frac{1}{2} [(\hat{\mathbf{x}}_R^4)^2 + \sigma_c^2] (1 - e^{-2\sigma_\alpha^2}) e^{\sigma_\alpha^2} \quad (65)$$

Since the cross range rate measurement,  $\dot{c}_m$ , is non-informative, its standard deviation,  $\sigma_c$ , is infinite. One can, however, set the value of  $\sigma_c$  used in (65) based on an a priori estimate of the standard deviation of target cross range rate to capture the effect that the cross range rate has on the ability to measure the line of sight velocity. The remaining components of the measurement noise covariance in the coordinate system,  $R_R$  (e.g.  $R_R^{44}$ ,  $R_R^{34}$ ), are set to infinity to capture that  $\dot{c}_m$  is non-informative. It is therefore useful to deal with the inverse of  $R_R$  and note that for a positive definite covariance matrix,

$$\begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}^{-1} \Big|_{\sigma_2 \rightarrow \infty} = \begin{bmatrix} (\sigma_1^2)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \quad (66)$$

therefore

$$R_R^{-1} = \begin{bmatrix} & & & 0 \\ (R_R^{1:3,1:3})^{-1} & & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (67)$$

The measurement noise covariance for (5),  $R_{ij}$ , is

$$R_{ij}^{-1} = D(\alpha_i)R_R^{-1}D(\alpha_j)^T \quad (68)$$

Since  $R_C^{-1}$  is not invertible,  $R_{ij}$  is not available for use in the Kalman filter gain calculation; one has to use the information form of the Kalman filter. The determinant of  $R_{ij}$  (needed for (9) in the calculation of  $w_{ij}$  using (19)) is also not available, so the determinant of  $R_R$  is used as a surrogate.

This is a simplification of (35)–(38) in [6]. The simplification is warranted due to the more accurate measurement in the present manuscript when compared to the measurement accuracy of [6].

## APPENDIX B OAKES' FORMULA

In [17], a simple explicit formula is given for the observed information matrix. A summary of Oakes' work is provided below with the necessary background from [8].

$$\begin{aligned} \mathcal{L}(\Psi; \mathcal{Z}) &= p_{\mathcal{Z}}(\mathcal{Z} | \Psi) \\ &= \frac{p_c(\mathcal{Z}, \mathcal{Y} | \Psi)p_{\mathcal{Z}}(\mathcal{Z} | \Psi)}{p_c(\mathcal{Z}, \mathcal{Y} | \Psi)} \end{aligned} \quad (69)$$

where  $p_c(\mathcal{Z}, \mathcal{Y} | \Psi)$  is defined after (16). Let  $k(\mathcal{X} | \mathcal{Z}, \Psi)$  be the conditional probability of the complete data,  $\mathcal{X}$ , given the observed data,  $\mathcal{Z}$ , namely

$$k(\mathcal{X} | \mathcal{Z}, \Psi) = \frac{p_c(\mathcal{Z}, \mathcal{Y} | \Psi)}{p_{\mathcal{Z}}(\mathcal{Z} | \Psi)} \quad (70)$$

Therefore

$$\mathcal{L}(\Psi; \mathcal{Z}) = \frac{p_c(\mathcal{Z}, \mathcal{Y} | \Psi)}{k(\mathcal{X} | \mathcal{Z}, \Psi)} \quad (71)$$

and

$$\ln \mathcal{L}(\Psi; \mathcal{Z}) = \ln p_c(\mathcal{Z}, \mathcal{Y} | \Psi) - \ln k(\mathcal{X} | \mathcal{Z}, \Psi) \quad (72)$$

Taking the expectation of both sides with respect to the conditional distribution of  $\mathcal{X}$  given  $\mathcal{Z}$ , using the previous estimate  $\Psi^{(l)}$  for  $\Psi$  gives

$$\begin{aligned} \ln \mathcal{L}(\Psi; \mathcal{Z}) &= E\{\ln \mathcal{L}_c(\Psi; \mathcal{Z}, \mathcal{Y}) | \mathcal{Z}, \Psi^{(l)}\} \\ &\quad - E\{\ln k(\mathcal{X} | \mathcal{Z}, \Psi) | \mathcal{Z}, \Psi^{(l)}\} \\ &= Q(\Psi; \Psi^{(l)}, \mathcal{Z}) - H(\Psi; \Psi^{(l)}, \mathcal{Z}) \end{aligned} \quad (73)$$

using (17) and where

$$H(\Psi; \Psi^{(l)}, \mathcal{Z}) = E\{\ln k(\mathcal{X} | \mathcal{Z}, \Psi) | \mathcal{Z}, \Psi^{(l)}\} \quad (74)$$

In [8] the following is shown using Jensen's inequality,

$$H(\Psi; \Psi^{(l)}, \mathcal{Z}) \leq H(\Psi^{(l)}; \Psi^{(l)}, \mathcal{Z}) \quad (75)$$

for all  $\Psi$  in the parameter space. This is fundamental in the proof for EM convergence, and leads to

$$\nabla_{\Psi} H(\Psi; \Psi^{(l)}, \mathcal{Z})|_{\Psi=\Psi^{(l)}} = 0 \quad (76)$$

Assuming that the expectation with respect to  $\mathcal{X}$  and differentiation with respect to  $\Psi$  are interchangeable,

$$E\{\nabla_{\Psi} \ln k(\mathcal{X} | \mathcal{Z}, \Psi) | \mathcal{Z}, \Psi^{(l)}\} = 0 \quad (77)$$

Also, from equivalent statements of Fisher's information,

$$\begin{aligned} -E\{\nabla_{\Psi} \nabla_{\Psi}^T \ln k(\mathcal{X} | \mathcal{Z}, \Psi) | \mathcal{Z}, \Psi^{(l)}\} \\ = E\{\nabla_{\Psi} k(\mathcal{X} | \mathcal{Z}, \Psi) \nabla_{\Psi} k(\mathcal{X} | \mathcal{Z}, \Psi)^T | \mathcal{Z}, \Psi^{(l)}\} \end{aligned} \quad (78)$$

Differentiation of (73) with respect to  $\Psi$  gives

$$\begin{aligned} \nabla_{\Psi} \ln \mathcal{L}(\Psi; \mathcal{Z}) &= \nabla_{\Psi} Q(\Psi; \Psi^{(l)}, \mathcal{Z}) \\ &\quad - E\{\nabla_{\Psi} \ln k(\mathcal{X} | \mathcal{Z}, \Psi) | \mathcal{Z}, \Psi^{(l)}\} \end{aligned} \quad (79)$$

By evaluating (79) using  $\Psi^{(l)} = \Psi$  and noting (77), we obtain

$$\nabla_{\Psi} \ln \mathcal{L}(\Psi; \mathcal{Z}) = \nabla_{\Psi} Q(\Psi; \Psi^{(l)}, \mathcal{Z})|_{\Psi^{(l)}=\Psi} \quad (80)$$

Differentiation of (79) with respect to  $\Psi$  gives,

$$\begin{aligned} \nabla_{\Psi} \nabla_{\Psi}^T \ln \mathcal{L}(\Psi; \mathcal{Z}) \\ = \nabla_{\Psi} \nabla_{\Psi}^T Q(\Psi; \Psi^{(l)}, \mathcal{Z}) \\ - E\{\nabla_{\Psi} \nabla_{\Psi}^T \ln k(\mathcal{X} | \mathcal{Z}, \Psi) | \mathcal{Z}, \Psi^{(l)}\} \end{aligned} \quad (81)$$

Differentiation of (79) with respect to  $\Psi^{(l)}$  gives,

$$\begin{aligned} \mathbf{0} &= \nabla_{\Psi} \nabla_{\Psi^{(l)}}^T Q(\Psi; \Psi^{(l)}, \mathcal{Z}) \\ &\quad - E\{\nabla_{\Psi} \ln k(\mathcal{X} | \mathcal{Z}, \Psi) \nabla_{\Psi^{(l)}} \ln k(\mathcal{X} | \mathcal{Z}, \Psi)^T | \mathcal{Z}, \Psi^{(l)}\} \end{aligned} \quad (82)$$

where  $\mathbf{0}$  is the appropriately sized null matrix.

Substitution  $\Psi = \Psi^{(l)}$  and adding (81) and (82) results in

$$\begin{aligned} \nabla_{\Psi} \nabla_{\Psi}^T \ln \mathcal{L}(\Psi; \mathcal{Z}) &= \nabla_{\Psi} \nabla_{\Psi}^T Q(\Psi; \Psi^{(l)}, \mathcal{Z}) \\ &\quad + \nabla_{\Psi} \nabla_{\Psi^{(l)}}^T Q(\Psi; \Psi^{(l)}, \mathcal{Z}) \end{aligned} \quad (83)$$

This result is used in (33), using the  $\Psi^{(l)}$  from the last EM iteration for a scan (i.e.  $\Psi^{(L)}$ ).

## ACKNOWLEDGMENT

Steven Bordonaro was supported by ONR Independent Applied Research (IAR). Peter Willett was supported by ONR under contract N00014-13-1-0231. Yaakov Bar-Shalom was supported under ARO W991NF-10-1-0369.

## REFERENCES

- [1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan  
*Estimation with Applications to Tracking and Navigation*.  
New York & Boston: John Wiley and Sons, 2001.
- [2] M. Baum and U. Hanebeck  
“Shape tracking of extended objects and group targets with star-convex RHMs,”  
in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*. IEEE, 2011, pp. 1–8.
- [3] M. Baum, B. Noack, and U. Hanebeck  
“Extended object and group tracking with elliptic random hypersurface models,”  
in *Proceedings of the 13th Conference on Information Fusion (FUSION)*. IEEE, 2010, pp. 1–8.
- [4] J. A. Bilmes et al.  
“A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models,”  
*International Computer Science Institute*, vol. 4, no. 510, p. 126, 1998.
- [5] Y. Boers and J. N. Driessen  
“A track before detect approach for extended objects,”  
in *2006 9th International Conference on Information Fusion*. IEEE, 2006, pp. 1–7.
- [6] S. Bordonaro, P. Willett, and Y. Bar-Shalom  
“Consistent linear tracker with position and range rate measurements,”  
in *Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012, pp. 880–884.
- [7] S. Bordonaro, P. Willett, Y. Bar-Shalom, M. Baum, and T. Luginbuhl  
“Extracting speed, heading and turn-rate measurements from extended objects using the EM algorithm,”  
in *Aerospace Conference, 2015 IEEE*. IEEE, 2015, pp. 1–12.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin  
“Maximum likelihood from incomplete data via the EM algorithm,”  
*Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
- [9] K. Gilholm, S. Godsill, S. Maskell, and D. Salmond  
“Poisson models for extended target and group tracking,”  
*SPIE, Signal and Data Processing of Small Targets, San Diego, CA, USA*, vol. 5913, p. 59130R, 2005.
- [10] K. Gilholm and D. Salmond  
“Spatial distribution model for tracking extended objects,”  
in *Radar, Sonar and Navigation, IEE Proceedings*, vol. 152, no. 5. IET, 2005, pp. 364–371.
- [11] S. Godsill, J. Li, and W. Ng  
“Multiple and extended object tracking with Poisson spatial processes and variable rate filters,”  
in *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*. IEEE, 2005, pp. 93–96.
- [12] S. Granger and X. Pennec  
“Multi-scale EM-ICP: A fast and robust approach for surface registration,”  
in *Computer Vision—ECCV 2002*. Springer, 2002, pp. 418–432.
- [13] L. Hammarstrand, L. Svensson, F. Sandblom, and J. Sorstedt  
“Extended object tracking using a radar resolution model,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2371–2386, 2012.
- [14] W. Koch  
“Bayesian approach to extended object and cluster tracking using random matrices,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, 2008.
- [15] J. Lan and X. R. Li  
“Tracking of extended object or target group using random matrix—part II: Irregular object,”  
in *Proceedings of the 15th International Conference on Information Fusion (FUSION)*. IEEE, 2012, pp. 2185–2192.
- [16] G. McLachlan and T. Krishnan  
*The EM algorithm and extensions*.  
John Wiley & Sons, 2007.
- [17] D. Oakes  
“Direct calculation of the information matrix via the EM,”  
*Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 2, pp. 479–482, 1999.
- [18] R. Streit and T. Luginbuhl  
“Probabilistic multi-hypothesis tracking,”  
Technical Report 10428, Naval Undersea Warfare Center, Newport, RI February 1995 (available from DTIC), Tech. Rep., 1995.
- [19] N. Ueda and R. Nakano  
“Deterministic annealing EM algorithm,”  
*Neural Networks*, vol. 11, no. 2, pp. 271–282, 1998.
- [20] M. Wieneke and W. Koch  
“Probabilistic tracking of multiple extended targets using random matrices,”  
in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 769812–769812.
- [21] ———  
“A PMHT approach for extended objects and object groups,”  
*IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2349–2370, 2012.
- [22] P. Willett, Y. Ruan, and R. Streit  
“PMHT: Problems and some solutions,”  
*Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 3, pp. 738–754, 2002.
- [23] Z. Zhang  
“Iterative point matching for registration of free-form surfaces,”  
*Int. Journal of Computer Vision*, vol. 13, no. 2, pp. 119–152, 1994.



**Steven Bordonaro** received his B.S. degree in Electrical Engineering from Boston University in 1991, and his M.S. degree, also in Electrical Engineering, from the University of Massachusetts. He completed a Ph.D. at the University of Connecticut in 2015. He has worked at the Naval Undersea Warfare Center in Newport, RI since 1991. His primary areas of research for the Navy and at the University of Connecticut have been signal processing, classification (machine learning) and tracking.



**Peter Willett** received his B.A.Sc. (Engineering Science) from the University of Toronto in 1982, and his Ph.D. degree from Princeton University in 1986. He has been a faculty member at the University of Connecticut ever since, and since 1998 has been a Professor. He was awarded IEEE Fellow status effective 2003. His primary areas of research have been statistical signal processing, detection, machine learning, data fusion and tracking. He was editor-in-chief for *IEEE Transactions on Aerospace and Electronic Systems*, and continues as VP for Publications.

**Yaakov Bar-Shalom** received the B.S. and M.S. degrees from the Technion, Israel Institute of Technology, in 1963 and 1967 and the Ph.D. degree from Princeton University in 1970, all in electrical engineering. From 1970 to 1976 he was with Systems Control, Inc., Palo Alto, California. Currently he is Board of Trustees Distinguished Professor in the Dept. of Electrical and Computer Engineering and Marianne E. Klewin Professor in Engineering at the University of Connecticut. He is also Director of the ESP (Estimation and Signal Processing) Lab. His current research interests are in estimation theory, target tracking and data fusion. He has published over 450 papers and book chapters in these areas and in stochastic adaptive control. He coauthored the monograph *Tracking and Data Association* (Academic Press, 1988), the graduate texts *Estimation and Tracking: Principles, Techniques and Software* (Artech House, 1993), *Estimation with Applications to Tracking and Navigation: Algorithms and Software for Information Extraction* (Wiley, 2001), the advanced graduate texts *Multitarget-Multisensor Tracking: Principles and Techniques* (YBS Publishing, 1995), *Tracking and Data Fusion* (YBS Publishing, 2011), and edited the books *Multitarget-Multisensor Tracking: Applications and Advances* (Artech House, Vol. I, 1990; Vol. II, 1992; Vol. III, 2000). He has been elected Fellow of IEEE for “contributions to the theory of stochastic systems and of multi-target tracking.” He has been consulting to numerous companies and government agencies, and originated the series of Multitarget-Multisensor Tracking short courses offered via UCLA Extension, at Government Laboratories, private companies and overseas. During 1976 and 1977 he served as Associate Editor of the *IEEE Transactions on Automatic Control* and from 1978 to 1981 as Associate Editor of *Automatica*. He was Program Chairman of the 1982 American Control Conference, General Chairman of the 1985 ACC, and Co-Chairman of the 1989 IEEE International Conference on Control and Applications. During 1983–87 he served as Chairman of the Conference Activities Board of the IEEE Control Systems Society and during 1987–89 was a member of the Board of Governors of the IEEE CSS. He was a member of the Board of Directors of the International Society of Information Fusion (1999–2004) and served as General Chairman of FUSION 2000, President of ISIF in 2000 and 2002 and Vice President for Publications in 2004–13. In 1987 he received the IEEE CSS Distinguished Member Award. Since 1995 he is a Distinguished Lecturer of the IEEE AESS and has given numerous keynote addresses at major national and international conferences. He is co-recipient of the M. Barry Carlton Award for the best paper in the *IEEE Transactions on Aerospace and Electronic Systems* in 1995 and 2000 and recipient of the 1998 University of Connecticut AAUP Excellence Award for Research. In 2002 he received the J. Mignona Data Fusion Award from the DoD JDL Data Fusion Group. He is a member of the Connecticut Academy of Science and Engineering. In 2008 he was awarded the IEEE Dennis J. Picard Medal for Radar Technologies and Applications, and in 2012 the Connecticut Medal of Technology. He has been listed by academic.research.microsoft (top authors in engineering) as number 1 among the researchers in Aerospace Engineering based on the citations of his work.





**Marcus Baum** is an assistant research professor at the University of Connecticut. He received his Dipl.-Inform. in computer science from the the Universitaet Karlsruhe (TH), Germany, in 2007 and his a Dr.-Ing. (summa cum laude), at the Intelligent Sensor-Actuator-Systems Laboratory Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, which is a merger of Universitaet Karlsruhe and Forschungszentrum Karlsruhe. His research interests are in the field of extended object and group tracking, nonlinear estimation, and sensor data fusion. He is the associate administrative editor for the *Journal of Advances in Information Fusion (JAIF)*.



**Tod Luginbuhl** received a B.S. degree in electrical engineering from the University of Rochester, Rochester, NY, in 1985. He earned an M.S. and a Ph.D. in electrical engineering from the University of Connecticut, Storrs, in 1991 and 1999, respectively. He is a senior researcher in the Sensors and Sonar Systems Department at the Naval Undersea Warfare Center, where he has been employed since 1985. His research interests include data association and fusion, frequency tracking, multi-target tracking, probability density function estimation, search theory, sequential analysis, signal classification and detection, and spectrum estimation. Dr. Luginbuhl is an Associate Editor of *Sequential Analysis*, and a member of Eta Kappa Nu.

# INTERNATIONAL SOCIETY OF INFORMATION FUSION

ISIF Website: <http://www.isif.org>

## 2017 BOARD OF DIRECTORS\*

2013–2015	2014–2016	2015–2017
Jean Dezert	Sten F. Andler	Joachim Biermann
Gee-Wah Ng	Murat Efe	Frederik Gustafsson
Anne-Laure Jousset	Lyudmila Mihaylova	Jesús García

\*Board of Directors are elected by the members of ISIF for a three year term.

## PAST PRESIDENTS

Jean Dezert, 2016	Stefano Coraluppi, 2010	Xiao-Rong Li, 2003
Darin Dunham, 2015	Elisa Shahbazian, 2009	Yaakov Bar-Shalom, 2002
Darin Dunham, 2014	Darko Musicki, 2008	Pramod Varshney, 2001
Wolfgang Koch, 2013	Erik Blasch, 2007	Yaakov Bar-Shalom, 2000
Roy Streit, 2012	Pierre Valin, 2006	Jim Llinas, 1999
Joachim Biermann, 2011	W. Dale Blair, 2005	Jim Llinas, 1998
	Chee Chong, 2004	

## SOCIETY VISION

The International Society of Information Fusion (ISIF) is the premier professional society and global information resource for multidisciplinary approaches for theoretical and applied information fusion technologies.

## SOCIETY MISSION

### Advocate

To advance the profession of fusion technologies, propose approaches for solving real-world problems, recognize emerging technologies, and foster the transfer of information.

### Serve

To serve its members and engineering, business, and scientific communities by providing high-quality information, educational products, and services.

### Communicate

To create international communication forums and hold international conferences in countries that provide for interaction of members of fusion communities with each other, with those in other disciplines, and with those in industry and academia.

### Educate

To promote undergraduate and graduate education related to information fusion technologies at universities around the world. Sponsor educational courses and tutorials at conferences.

### Integrate

Integrate ideas from various approaches for information fusion, and look for common threads and themes—look for overall principles, rather than a multitude of point solutions. Serve as the central focus for coordinating the activities of world-wide information fusion related societies or organizations. Serve as a professional liaison to industry, academia, and government.

### Disseminate

To propagate the ideas for integrated approaches to information fusion so that others can build on them in both industry and academia.

## Call for Papers

The Journal of Advances in Information Fusion (JAIF) seeks original contributions in the technical areas of research related to information fusion. Authors are encouraged to submit their manuscripts for peer review <http://isif.org/journal>.

## Call for Reviewers

The success of JAIF and its value to the research community is strongly dependent on the quality of its peer review process. Researchers in the technical areas related to information fusion are encouraged to register as a reviewer for JAIF at <http://jaif.msubmit.net>. Potential reviewers should notify via email the appropriate editors of their offer to serve as a reviewer.