Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra kybernetiky

DIPLOMOVÁ PRÁCE

PLZEŇ 2020 **JAN KREJČÍ**

University of West Bohemia

Faculty of Applied Sciences

Department of Cybernetics

MASTER'S THESIS

Tracking of Moving Object

PILSEN 2020 **JAN KREJČÍ**

ZÁPADOČESKÁ UNIVERZITA V PLZNI

Fakulta aplikovaných věd

Akademický rok: 2019/2020

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Bc. Jan KREJČÍ**
Osobní číslo:	**A18N0056P**
Studijní program:	**N3918 Aplikované vědy a informatika**
Studijní obor:	**Kybernetika a řídicí technika**
Téma práce:	**Sledování pohybujícího se objektu**
Zadávající katedra:	**Katedra kybernetiky**

Zásady pro vypracování

1) Seznamte se s fomulací úlohy sledování pohybujícího se objektu.
2) Seznamte se se základními přístupy řešícími tuto úlohu.
2) Vybrané přístupy implementujte.
3) Porovnejte výsledky dosažené implementovanými přístupy.

Rozsah diplomové práce:	**40-50**
Rozsah grafických prací:	**dle potřeby**
Forma zpracování diplomové práce:	**tištěná**
Jazyk zpracování:	**Angličtina**

Seznam doporučené literatury:

Dodá vedoucí diplomové práce.

Vedoucí diplomové práce:	**Doc. Ing. Ondřej Straka, Ph.D.** Katedra kybernetiky

Datum zadání diplomové práce:	**1. října 2019**
Termín odevzdání diplomové práce:	**25. května 2020**

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
L.S.

Radová

Doc. Dr. Ing. Vlasta Radová
děkanka

Prof. Ing. Josef Psutka, CSc.
vedoucí katedry

V Plzni dne 1. října 2019

PROHLÁŠENÍ

Předkládám tímto k posouzení a obhajobě diplomovou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

V Plzni dne . . . .

vlastnoruční podpis

PODĚKOVÁNÍ

Tímto bych rád poděkoval panu doc. Ing. Ondřeji Strakovi, Ph.D za odborné vedení, cenné rady a připomínky, poskytnuté materiály, trpělivost, čas a ochotu při zpracovávání této diplomové práce.

Anotace

Tato diplomová práce se zabývá některými zobecněnými úlohami nelineární filtrace, které se vyskytují v reálných situacích při sledování pohybujícího se objektu. Konkrétně se jedná o připuštění nedetekování objektu, přítomnost falešných měření, či připuštění neexistence objektu. Tato zobecnění jsou klíčová pro úlohy sledování více objektů, kterým je zde kladen zvláštní důraz. Tato diplomová práce si bere za cíl seznámit čtenáře s přístupy k modelování, s potřebným matematickým aparátem a se základními filtračními algoritmy. Prezentované koncepce jsou srovnány jak z teoretického, tak z praktického pohledu. Konkrétně je diskutována aplikace na sledování polohy objektů na videu. Implementace je řešena pomocí gaussovských směsí.

klíčová slova: nelineární filtrace, sledování více cílů, asociace dat, náhodné množiny, bodové procesy

Annotation

This thesis addresses some generalizations of the nonlinear filtering problem that appear in real world scenarios when tracking moving objects. In particular, the generalizations include admittance of the object being undetected, presence of extraneous measurements, or assumption that the object is not necessarily present. The generalizations are the key in the area of multiple target tracking, to which a stress is laid. The goal is to present the modeling approaches, the necessary mathematical tools, and the resulting basic filtering algorithms. The concepts are compared in both theoretical and practical points of view. Specifically, a visual tracking application is discussed. Gaussian sum implementations are used.

keywords: nonlinear filtering, multiple target tracking, data associations, random sets, point processes

Contents

Nomenclature

$\emptyset$	=	empty set
$\mathbf{1}_{\mathcal{S}}(\mathbf{x})$	=	Indicator function of the set $\mathcal{S}$
E	=	Multi-target state-space
$\mathrm{E}[\cdot]$	=	Expectation operator
$G_{\Xi}[h]$	=	Probability generating functional of the random set Ξ
h_k	=	Injective function, that indexes hypothesis θ^k onto the set of integers
$\mathbf{I}$	=	Identity matrix of appropriate dimensions.
k	=	Time instant, $k \in \{0, \ldots, k_F\} \subseteq \mathbb{N}_0$
$l_k(\mathbf{z}\|\mathbf{x})$	=	Measurement likelihood function evaluated at the measurement $\mathbf{z}$, given $\mathbf{x}$
$\mathcal{N}(\mathbf{x}, \mathbf{m}, \mathbf{P})$	=	Gaussian PDF with the mean $\mathbf{m}$ and covariance matrix $\mathbf{P}$ evaluated at $\mathbf{x}$
$\Pr(E)$	=	Probability of the event E
$p_{\boldsymbol{\xi}}(\mathbf{x})$	=	Probability density function of the random vector $\boldsymbol{\xi}$ evaluated at $\mathbf{x}$
$p_{\Xi}(X), f_{\Xi}(X)$	=	Probability density function of the RFS Ξ evaluated at the set X
$\mathrm{Sym}(m)$	=	Set of all permutations of integers up to m
$v(\mathcal{S})$	=	Lebesgue measure of a set $\mathcal{S}$
$\mathbf{x},\ \mathbf{x}_k$	=	Single-target state-vector at time k
$\mathcal{X}$	=	Single-target state space
$X,\ X_k$	=	Multi-target state-set
$\|X\|$	=	Cardinality of the set X
$\mathbf{z},\ \mathbf{z}_k$	=	Measurement-vector at time k
$\mathbf{z}^k$	=	Time-sequence of previous measurements, $\mathbf{z}_1, \ldots, \mathbf{z}_k$
$Z,\ Z_k$	=	Measurement-set collected at time k
Z^k	=	Time-sequence of previous measurement-sets, $Z_1, \ldots, Z_k$
$\mathcal{Z}$	=	Single-measurement space
$\beta_{\Xi}(\mathcal{S})$	=	Belief mass measure of the random set Ξ evaluated at the set $\mathcal{S}$
$\delta_{\mathbf{x}}(\cdot)$	=	Dirac delta centered at $\mathbf{x} \in \mathcal{X}$
$\delta_n^{\mathrm{Kron}}(\cdot)$	=	Kronecker delta centered at $n \in \mathbb{N}$
λ	=	Parameter of the Poisson distribution (expected number of targets)
θ_k	=	Data association hypothesis at the time step k
θ^k	=	Sequence of data association hypothesis $(\theta_1, \ldots, \theta_k)$
$\psi_{k\|k-1}(\mathbf{y}\|\mathbf{x})$	=	Markov transition density evaluated at $\mathbf{y}$, given $\mathbf{x}$
Ξ	=	Random finite set or point process (as a random variable)
$\int_{\mathcal{S}} f(X)\delta X$	=	Set integral of the function f on region $\mathcal{S} \subseteq \mathcal{X}$
$X_1 \uplus X_2$	=	Disjoint union of the sets X_1 and X_2

Acronyms and Abbreviations

BRRs	...	Bayesian recursive relations
δ-GLMB	...	Delta-generalized labeled multi Bernoulli
EAP	...	Expected aposteriori
EM	...	Expectation-maximization
FISST	...	Finite-set statistics
GM	...	Gaussian mixture
GNN	...	Global nearest neighbor
IPDA	...	Integrated probabilistic data association
MAP	...	Maximum aposteriori
MeMBr	...	Multi-target multi Bernoulli
MHT	...	Multiple hypothesis tracking
MMSE	...	Minimum mean square error
MPH	...	Most probable hypothesis
MTT	...	Multi-target tracking or multiple target tracking
NN	...	Nearest neighbor
i.i.d.	...	Independent and identically distributed
JIPDA	...	Joint integrated probabilistic data association
JoTT	...	Joint target detection and tracking
JPDA	...	Joint probabilistic data association
KF	...	Kalman filter
PDA	...	Probabilistic data association
PDF	...	Probability density function
PET	...	Positron emission tomography
PGFL	...	Probability generating functional
PHD	...	Probability hypothesis density
PMBM	...	Poisson multi Bernoulli mixture
PPP	...	Poisson point process
RFS	...	Random finite set
TBD	...	Track before detect
YOLOv3	...	You only look once version 3

CHAPTER 1

Introduction

Tracking of moving objects has been a topic in a large variety of practical interests such as surveillance, defense, positioning and navigation, autonomous vehicles, computer vision and air traffic control. Today, tracking problems often include phenomena such as multiple targets[1] of different nature and behavior, multiple sensors, sensor faults, and many others. See for example [1, 2, 3, 4, 5] and references therein. Algorithms used to deal with such problems are usually beyond the classic textbook-level of the optimal filtering theory [6]. To establish a connection between the theoretical basis and real tracking scenarios, we could look back into the history of target tracking.

In the beginnings, the aim was to solve a problem of tracking exactly one object present, given measurement data. Ever since the early works of Wiener, and Kolomogorov, the task has been reformulated and addressed with various assumptions and concepts. About twenty years after the Wiener's rather measure theoretic approach, Kalman [7] has pushed the theory forward, introducing a state space approach for a specific linear problem. After further generalizations, the optimal nonlinear estimation/filtering theory emerged, based on Bayesian inference [8]. This theory, which today became a textbook-level [6, 9], settled down a standardized approach for various problems. Tracking moving object is addressed to be such a problem. Research and textbooks often deal with a solution to its nonlinear basis with techniques such as the Gaussian sum filtering [10], extended and unscented filtering [11, 12], sequential Monte Carlo techniques, point mass filtering [13] and other approaches, see for example [14, 15, 16]. In this thesis, we take the solutions to the underlying nonlinear basis for granted. We will build upon the general case, while dealing with another research direction, inspired by requirements of qualitatively different applications.

Other research is dedicated to solving rather practical tracking scenarios, trying to deal with more relaxed assumptions, or generalizations to the classic problem of target tracking. In other words, in the real-world scenarios, the nature does not always fit into the single-target Bayesian filtering framework. Such a case has risen in the late seventies, when [17,

[1] The terms *targets* and *objects* are assumed to have the same meaning and will be used interchangeably.

18, 19] proposed methods to solve the problem of tracking multiple objects in cluttered environment, which has emerged into a new research direction, today referred to as multiple[2] target tracking (MTT). Originally, this problem was solved rather heuristically, forming the so-called data association hypotheses. The Multiple hypothesis tracking (MHT) approach has been established in [17, 18, 19], and reviewed later in [20]. Both the moving objects and measurements were modeled as random tuples of vectors. Later, in the mid-nineties, Mahler [21] introduced a new mathematically tractable approach to modeling random finite sets of vectors, that gave rise to a multi-target case of the Bayesian framework. Recently, the Mahler's approach has gained a considerable research interest and many appealing algorithms has been derived [22]. Also, another approach to modeling targets as sets, referred to as the point process approach, has been introduced by Streit [23, 24, 25], and also others [26], lately. A form of animosity has originated in the field, since the approaches of Streit's and Mahler's differ mostly in phenomenological thoughts, such as Streit's modeling of "clutter target". Mahler saw these concepts (and also other aspects) as false and improperly attributed, and he released his scientific criticism in an arXiv paper [27].

Various books and number of papers have been dedicated to the introduced research direction. In fact, considerably general real tracking scenarios concerning multiple target tracking have been widely considered, such as multi-sensor multi-target tracking, extended object tracking, tracking based on intensity measurements[3] or other type of measurements, and others[4]. For a novice in this field, it may be hard to grasp the main concepts of the multi-target tracking. A possible solution might be a free online course on edX, introduced in mid-2019 [28][5]. This thesis is partially inspired by this online course, but aims to deepen the mathematical insight, i.e. to offer a comprehensive introduction to the underlying mathematics, for the reader to be able to understand the derivation techniques of various state-of-the-art filters. Note that we aim to give an appropriate introduction namely to tracking of multiple moving objects, stressing the connection to tracking a single moving object[6].

In particular, this thesis aims to cover the basic ideas, such as modeling undetection[7] of a target, introducing an extraneous "false alarm" measurements called the clutter, and

[2]The terms *multiple* and *multi*, when referring to the target tracking, will be used interchangeably.

[3]Known as the track-before-detect (TBD).

[4]Note that the real practical applications of MTT is often classified.

[5]The course consists of lecture videos, which are also uploaded on YouTube to be freely available.

[6]This should justify why this thesis is called simply *tracking of moving object* and not *tracking of multiple moving objects*.

[7]Undetection (also called missed detection) is an event such that an existing object have generated no measurement at a particular time instant.

letting a target to be randomly switching *on/off* [8]. These constructions are key in the area of multi-target tracking, for which an idea of approximating the targets to be independent and identically distributed (i.i.d.) is discussed. The stress is laid on the foundation of the practical multiple target filters. That is, the approaches, the underlying mathematical models and their application. Various notes are given to establish basic connections to more general cases described above. Note that attempting to keep the mathematics simple and easy to understand, the rigor may be slightly reduced. In such case, it is noted and references to appropriate literature are provided.

The thesis is organized as follows. In Chapter 2, the theoretical foundations are given for both the single-target and multi-target modeling. In particular, the classical filtering theory is summed up in Section 2.1, the introduction to modeling and treating random finite sets (RFS) especially using the finite set statistics (FISST) is given in Section 2.2, and a general Gaussian sum filtering principles are established in Section 2.3. The following three chapters are dedicated to the derivation of basic filtering algorithms, gradually relaxing assumptions of the classical filtering theory. In Chapter 3, we deal with the assumptions of target undetections and the presence of clutter. The data association hypotheses are introduced and the general single-target special case of the multiple hypothesis tracking (MHT) algorithm is derived. In Chapter 4 an assumption that the target is randomly switching *on/off* is added. This case is modeled using the FISST, yielding the Bernoulli filter. In Chapter 5, generalizations to tracking multiple targets is discussed, leading to the so-called PHD, or intensity filter. This case is modeled rather heuristically and intuitively, leaving the FISST-based derivation for an appendix. Also, an outlook of MTT is given to deepen the insight to the literature and related work. Next, in Chapter 6, an application to tracking objects from videos is discussed, providing a practical comparison of the algorithms. The thesis concludes in Chapter 7. Useful information and insights are provided in the following appendices.

[8] Such object will be referred to as Bernoulli object.

CHAPTER 2

Theory Preliminaries

In this chapter, the mathematical foundations needed to understand the derivations in the following chapters are given. As was stated in the introduction, in this thesis, we limit ourselves to a special case of the tracking problem, frequently appearing in the MTT literature. That is, we deal only with the so-called point objects[1] that all obey the same transition process, and point measurements that arrived from a single source[2]. We model discrete time tracking case, and we assume centralized data processing[3].

In particular, we assume the moving objects to be sufficiently modeled as points in some vector space $\mathcal{X} \subseteq \mathbb{R}^{n_x}$, i.e. we say that $\mathbf{x}_k$ is a single-target state-vector at the discrete time step $k \in \mathbb{N}_0$, if $\mathbf{x}_k \in \mathcal{X}$, and if $\mathbf{x}_k$ completely characterizes the physical target at the particular time k. We will call $\mathcal{X}$ the single-target state space. Also, the measurements are assumed to be points in some vector space $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$. Recalling the term *point objects*, it is to say we assume that no more than one measurement may be generated by a particular object.

Following the classical filtering theory, we model the targets to be non-deterministic, i.e. we describe them with probability measures, or densities. The goal is to find a probabilistic representation of the targets, given all measurements available, i.e. up to the present time instant k.

Single object modeling and the classic filtering theory are reviewed in Section 2.1, followed by tools to model the generalized scenarios. To grasp a graphical intuition, a possible single target tracking scenario with position measurements and assumptions of the classic filtering theory is contrasted with a multi-target tracking scenario in Fig. 2.1.

[1] For tracking of extended objects see e.g. the survey [29].

[2] For sensor-focused discussion, such as track before detect (TBD) and multi-sensor fusion, see e.g. [1, 2, 4, 30].

[3] For distributed estimation see e.g. the survey [31]

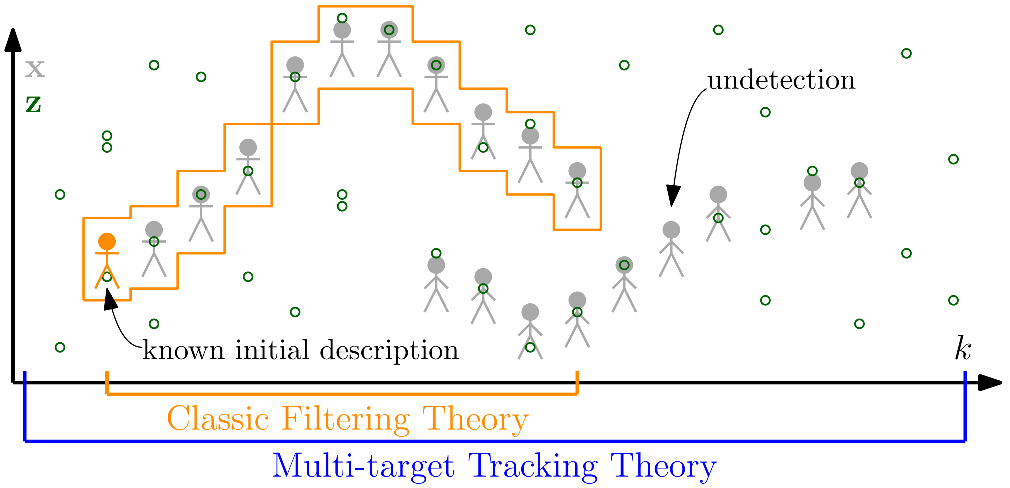


Fig. 2.1: Illustration of the scenarios.

2.1 Classic Single-target Modeling

In this chapter, we review the classical filtering theory as depicted in Fig. 2.1, i.e. it is assumed that the object exists for the whole tracking time horizon, its initial description is known, and a measurement at every time step is a single vector, known to be generated from the object.

2.1.1 Uncertainty Representation

Let the (moving) single-target be a stochastic process $\{\boldsymbol{\xi}_k\}_{k=0}^{k_F}$ continuous in values, and discrete in time, such that $\mathbf{x}_k \in \mathcal{X}\ \forall k$ is a realization of the random vector $\boldsymbol{\xi}_k$. Denote the probability measure of $\boldsymbol{\xi}_k$ by $P_{\boldsymbol{\xi}_k}(\mathcal{S})$, where $\mathcal{S}$ is an arbitrary Borel subset of $\mathcal{X}$. If the measure is absolutely continuous, then by the Radon-Nikodým theorem, its probability density function (PDF) $p_{\boldsymbol{\xi}_k}(\mathbf{x}_k)$ exists, such that the probability $\Pr(\boldsymbol{\xi}_k \in \mathcal{S})$ of $\boldsymbol{\xi}_k$ being in $\mathcal{S} \subseteq \mathcal{X}$ can be computed as

$$\Pr(\boldsymbol{\xi}_k \in \mathcal{S}) = P_{\boldsymbol{\xi}_k}(\mathcal{S}) = \int_{\mathcal{S}} dP_{\boldsymbol{\xi}_k}(\mathbf{x}_k) = \int_{\mathcal{S}} p_{\boldsymbol{\xi}_k}(\mathbf{x}_k) d\mathbf{x}_k \tag{2.1}$$

In the following, we assume a PDF always exists. From now on, the notation for the random vectors at their measures and densities will be omitted for convenience. Moreover, we will abuse the notation $\boldsymbol{\xi}_k = \mathbf{x}_k$ for the sake of simplicity.

2.1.2 Statistical Moments

Omitting the time index, the first and second statistical moments of a random vector $\mathbf{x}$ are the mean $\mathbf{m}$ and the covariance matrix $\mathbf{P}$, defined as

$$\mathbf{m} = \mathrm{E}[\mathbf{x}] = \int_{\mathcal{X}} \mathbf{x}\, p(\mathbf{x}) d\mathbf{x} \tag{2.2}$$

$$\mathbf{P} = \mathrm{E}[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T] = \int_{\mathcal{X}} (\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T p(\mathbf{x}) d\mathbf{x} \tag{2.3}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator and T stands for matrix transpose. They represent summary statistics of the random vector $\mathbf{x}$. The mean is the expected realization of the random vector and the covariance matrix represents a spread of realizations from the mean.

2.1.3 Gaussian and Gaussian Mixture PDF

A random vector $\mathbf{x}$ is said to be Gaussian, if its PDF is given by

$$\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}) = \frac{1}{\sqrt{\det(2\pi\mathbf{P})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1} (\mathbf{x}-\mathbf{m})\right) \tag{2.4}$$

where the parameters $\mathbf{m}$ and $\mathbf{P}$ are the first two moments of the random vector $\mathbf{x}$, respectively. In this thesis, for simplicity, we are concerned with the Gaussian sum implementation technique only. All the densities referring single-objects are assumed to be Gaussian, or Gaussian mixtures.

A Gaussian mixture (GM) PDF is a sum of Gaussian PDFs $\mathcal{N}(\mathbf{x}; \mathbf{m}^i, \mathbf{P}^i),\ i = 1, \ldots, M$, called the components or terms, weighted by $w^i \geq 0\ \forall i$ called the weights,

$$p_{\mathrm{GM}}(\mathbf{x}) = \sum_{i=1}^{M} w^i \mathcal{N}(\mathbf{x}; \mathbf{m}^i, \mathbf{P}^i), \qquad \text{such that} \quad \sum_{i=1}^{M} w^i = 1 \tag{2.5}$$

2.1.4 State-space Model

Let the single-target $\mathbf{x}_k$ be a hidden Markov process, which generates the measurement $\mathbf{z}_k$ at every time step k. Then, it can be described in terms of the autonomous stochastic state-space model,

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{w}_{k-1}), \qquad \text{with the PDF} \qquad \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) \tag{2.6}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k), \qquad \text{with the PDF} \qquad l_k(\mathbf{z}_k|\mathbf{x}_k) \tag{2.7}$$

where $\mathbf{f}_{k-1}$ describes the system dynamics, $\mathbf{h}_k$ is the measurement function, $\mathbf{w}_k$ and $\mathbf{v}_k$ are the state and measurement noises with PDFs $p_{\mathbf{W}_k}(\mathbf{w}_k)$ and $p_{\mathbf{V}_k}(\mathbf{v}_k)$, respectively. We will use the equivalent probabilistic description with the Markov transition density function $\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$, and the measurement likelihood function $l_k(\mathbf{z}_k|\mathbf{x}_k)$. Note that the subscripts may be omitted, if not causing confusion within a given context.

2.1.5 Single-target Bayesian Recursive Relations

Assume the PDF of $\mathbf{x}_k$ at time $k = 0$ is known and denoted by

$$p_{0|0}(\mathbf{x}_0|\mathbf{z}^0) \triangleq p_{\mathbf{X}_0}(\mathbf{x}_0) \tag{2.8}$$

We assume a single measurement arrives at a time. The goal is, at the given time k (i.e. *online*), to find the conditional PDF of $\mathbf{x}_k$ given all the measurements up to the time k. Denote the desired PDF $p_{k|k}(\mathbf{x}_k|\mathbf{z}^k)$, where $\mathbf{z}^k \triangleq (\mathbf{z}_1, \dots, \mathbf{z}_k)$ is the sequence of measurements up to the time step k, which are assumed known.

The core result in the nonlinear filtering theory are the Bayesian recursive relations (BRRs), which give a solution to the problem providing two steps as follows,

$$p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1}) = \int_{\mathcal{X}} \psi(\mathbf{x}_k|\mathbf{x}_{k-1}) p_{k-1|k-1}(\mathbf{x}_{k-1}|\mathbf{z}^{k-1}) d\mathbf{x}_k \tag{2.9}$$

$$p_{k|k}(\mathbf{x}_k|\mathbf{z}^k) = \frac{l(\mathbf{z}_k|\mathbf{x}_k) p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1})}{p(\mathbf{z}_k|\mathbf{z}^{k-1})} \propto l(\mathbf{z}_k|\mathbf{x}_k) p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1}) \tag{2.10}$$

where the eq. (2.9) is the Chapman-Kolomogorov equation, which represents the prediction, also called time-update; and eq. (2.10) is the Bayes rule, which represents the filtering step, also called Bayes-update, or measurement-update. We call $p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1})$ the prior PDF and $p_{k|k}(\mathbf{x}_k|\mathbf{z}^k)$ the posterior PDF. The BRRs are solved recursively as the time progresses using various techniques briefly stated in the introduction [10, 11, 12, 13, 14, 15, 16].

2.1.6 Linear-gaussian Case

As was said before, we are interested only in the Gaussian sum implementation technique [10]. The core of this technique is the "PDF-based" Kalman filter [6, 9, 32]. This refers to a specific form of the underlying problem, i.e. of the Markov transition density (2.6), measurement likelihood (2.7) and the initial PDF (2.8), that lead to a closed-form solution to the BRRs (2.9-2.10) [4].

The underlying estimation problem is said to be linear-gaussian, if

$$\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}_{k-1}\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}) \tag{2.11}$$

$$l_k(\mathbf{z}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k) \tag{2.12}$$

$$p_{0|0}(\mathbf{x}_0|\mathbf{z}^0) = \mathcal{N}(\mathbf{x}_0; \mathbf{m}_{0|0}, \mathbf{P}_{0|0}) \tag{2.13}$$

where $\mathbf{F}_{k-1} \in \mathbb{R}^{n_x \times n_x}$ is the state transition matrix, $\mathbf{Q}_{k-1} \in \mathbb{R}^{n_x \times n_x}$ is a covariance matrix of the additive state noise, $\mathbf{H}_k \in \mathbb{R}^{n_z \times n_x}$ is the measurement matrix, $\mathbf{R}_k \in \mathbb{R}^{n_z \times n_z}$ is a covariance matrix of the additive measurement noise, $\mathbf{m}_{0|0} \in \mathbb{R}^{n_x}$ is the initial mean vector and $\mathbf{P}_{0|0} \in \mathbb{R}^{n_x \times n_x}$ is the initial covariance matrix. The time indices of $\mathbf{F}_{k-1}, \mathbf{Q}_{k-1}, \mathbf{H}_k$ and $\mathbf{R}_k$ will be omitted for the sake of simplicity.

Suppose that $\mathbf{Q}, \mathbf{R}$ and $\mathbf{P}_{0|0}$ are positive definite. The solution to the BRRs (2.9-2.10) is then given by the prior and posterior being

$$p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}, \mathbf{P}_{k|k-1}) \tag{2.14}$$

$$p_{k|k}(\mathbf{x}_k|\mathbf{z}^k) = \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}, \mathbf{P}_{k|k}) \tag{2.15}$$

[4]For a future discussion, note that such result is closely connected to the term *conjugate prior*, which refers to case when the prior and the posterior distributions, if connected by the Bayes rule, are both from the same family of distributions. In another words a specific choice of the prior distribution on the Bayes rule entry results to the posterior having the same distribution that differ only in value of its parameters.

with the means and covariance matrices given by the Kalman filter (KF) equations

$$\mathbf{m}_{k|k-1} = \mathbf{F}\mathbf{m}_{k-1|k-1} \tag{2.16a}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q} \tag{2.16b}$$

$$\mathbf{m}_{k|k} = \mathbf{m}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\mathbf{m}_{k|k-1}) \tag{2.16c}$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1} \tag{2.16d}$$

$$\text{with} \qquad \mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R})^{-1} \tag{2.16e}$$

where $\mathbf{I} \in \mathbb{R}^{n_y \times n_x}$ is the identity matrix. This result is a corollary of the well known identity for Gaussian densities [33, pp. 9-11]. Later in the text we will need a further insight into Eq. (2.10), noting the particular form of the measurement prediction $p(\mathbf{z}_k|\mathbf{z}^{k-1})$. Using the law of total probability and noting it results into (2.14 with 2.16a,2.16b) for the linear-Gaussian case, it can be shown that

$$p(\mathbf{z}_k|\mathbf{z}^{k-1}) = \int_{\mathcal{X}} l(\mathbf{z}_k|\mathbf{x}_k)p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1})d\mathbf{x} = \mathcal{N}(\mathbf{z}_k; \hat{\mathbf{z}}_{k|k-1}, \mathbf{P}^{\hat{\mathbf{z}}}_{k|k-1}) \tag{2.17}$$

where

$$\hat{\mathbf{z}}_{k|k-1} = \mathbf{H}\mathbf{m}_{k|k-1} \tag{2.18}$$

$$\mathbf{P}^{\hat{\mathbf{z}}}_{k|k-1} = \mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R} \tag{2.19}$$

Therefore, the numerator of the Bayes update (2.10) for the linear-Gaussian case, is

$$l(\mathbf{z}_k|\mathbf{x}_k)p_{k|k-1}(\mathbf{x}_k|\mathbf{z}^{k-1}) = \mathcal{N}(\mathbf{z}_k; \hat{\mathbf{z}}_{k|k-1}, \mathbf{P}^{\hat{\mathbf{z}}}_{k|k-1}) \cdot \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}, \mathbf{P}_{k|k}) \tag{2.20}$$

Having reviewed these results, we are ready to start with the multi-target and "multi-measurement" modeling with finite sets.

2.2 Finite Set Modeling

At any fixed time instant, assume there is an unknown number of targets in the tracking area, whose states are unknown as well. A possible approach to constructing MTT algorithms is to handle every single target as an isolated entity and to extend the single-target tracking methods to the multi-target case "straightforwardly". The measurement origin uncertainty problem then arises, which can be tackled by evaluating the co-called data association hypotheses. Such approach, referred here to as *hypothesis-based approach*, takes the MHT and similar algorithms such as the probabilistic data association (PDA)-based filters. One of the major drawbacks of this approach is that there is no explicit modeling to estimate the number of objects in the tracking area [3, pp. 223-225]. A possible solution is to assign a "unifying" probabilistic description to the targets and to model them together as a single entity. An appropriate, phenomenologically correct model of multiple targets (and measurements as well) seems to be a finite set of distinct vectors. To continue with non-deterministic modeling, such set should be modeled similarly to a random variable. Random entities, whose outcomes are such sets, can be called random finite sets (RFSs), or finite point processes. These can be viewed as sets containing a finite random number of targets, whose states are random as well.

The RFSs can be attributed as representations of the finite point processes that are treated in an unconventional manner. It is quite hard to make things completely clear. For the theory presented in this chapter, we consider an RFS as a subclass of the point processes that model the underlying physical reality with sets, no matter how they are treated (thus to be phenomenologically interchangeable). However, to attribute the so-called *point process approach* to the MTT, the term point processes in this thesis refers to the case of modeling RFSs with the conventional probabilistic manner (which might be misleading [27, 34]).

For the RFSs, analogous probabilistic descriptions and laws to the single-target case can be introduced. In particular they can be described by a density and obey appropriately defined BRRs. The finite set statistics (FISST) is a collection of tractable mathematical tools to work with RFSs, describing them with non-additive measures. This theory was used to derive tracking algorithms, some of which are presented in this thesis. To verify that these algorithms are correct in terms of the standard probability treatment, it is argued that FISST densities can be viewed as the standard probability densities in Appendix A. In the case of tracking discussed in this thesis, we therefore argue that the theory of point processes and

RFSs with FISST can be used to model the physical reality interchangeably. The following discussion covers some basics of both FISST and point process theory. Since the following text includes only some basics of the theory, the discussion is extended in Appendix A.

2.2.1 Uncertainty Representation and the Set Integral

The moving objects are modeled as a sequence for $k = k_0, \ldots, k_F$ of finite sets of single-object states $\{\Xi_k\}_{k=k_0}^{k_F}$, with X_k being a realization of the RFS $\Xi_k = \{\mathbf{x}_k^1, \ldots, \mathbf{x}_k^{n_k}\}$. Both the cardinality n_k and the states are a assumed random[5]. The vectors $\mathbf{x}_k^i \in \mathcal{X}\ \forall k\ \forall i$ represent the states of the targets. If $|X_k| = 0$, then $X_k = \emptyset$ is an empty configuration. The notion of RFSs can be straightforwardly generalized to model both the moving targets[6] and measurements. Hence, in the following, the time indices will be omitted, and Ξ will denote a general RFS.

The belief mass measure $\beta_\Xi(\mathcal{S})$ of the RFS Ξ is a set function used to describe the RFS with FISST analogically to the probability measure when treated conventionally,

$$\beta_\Xi(\mathcal{S}) \triangleq \int_{\mathcal{S}} p_\Xi(X)\delta X \tag{2.21}$$

where $\int_{\mathcal{S}} f(X)\delta X$ denotes the FISST set integral of the set-valued function $f(X)$ over $\mathcal{S} \subseteq \mathcal{X}$ defined as [21, 4]

$$\int_{\mathcal{S}} f(X)\delta X \triangleq \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{S}^n} f(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n \tag{2.22}$$

It is usual to call the function $p_\Xi(X)$ in 2.21 the FISST density, multi-target PDF, or simply a PDF. Analogically to the single-target case, $\beta_\Xi(\mathcal{X}) = 1$, but note that both the set integral and the belief mass measure are non-additive in general. For a deeper understanding and connection to the conventional probabilistic approach, see Appendix A.

Note that

$$p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) = p_N(n) \cdot p_{\Xi|n}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}|n) \tag{2.23}$$

[5] We assume that Ξ_k is a *simple* process, which means we do not allow multiplicities in X_k, i.e. $\mathbf{x}_k^i \neq \mathbf{x}_k^j$, $i \neq j$, i.e. all realizations (vectors in X_k) are assumed distinct.

[6] Referred to as *multi-target*, and *multi-object*.

is the joint PDF of both the states $\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$ and their number n, with $p_N(n),\ n \in \mathbb{N}_0$ being the cardinality distribution, i.e. the distribution of $N \triangleq |\Xi|$.

2.2.2 Ordered Realization

An RFS can be arbitrarily ordered and modeled as a random tuple, or a matrix $\mathbf{X} \triangleq (\mathbf{x}^1, \ldots, \mathbf{x}^n)$. This structure is obviously no longer a set, thus precisely speaking it is not an RFS but still a point process. Its PDF is given by

$$p_{\mathbf{X}|N}(\mathbf{x}^1, \ldots, \mathbf{x}^n|n) = \frac{1}{n!} \cdot p_{\Xi|N}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}|n) \tag{2.24}$$

This introduces generally no phenomenological bias, however, it is not usual to define such orderings in terms of FISST, as it might potentially lead to some unknown or hidden statistical bias [4, 35]. Such ordering will be used e.g. in Chapter 3, when modeling data association hypotheses arising from arbitrarily ordering a given set of measurements.

2.2.3 FISST Multi-target Bayesian Recursive Relations

The common probabilistic tools, such as the law of total probability and conditional probability[7], are based on the notion of additivity of the underlying probability measures [36]. If the measures are absolutely continuous, which we assume here, those tools apply to the PDFs also. Since the FISST densities are "in essence" the probability densities[8], we can define the multi-target BRRs with the FISST set integral as [21, 4]

$$p_{k|k-1}(X_k|Z^{k-1}) = \int_{\mathcal{X}} \psi_{k|k-1}(X_k|X_{k-1}) p_{k-1|k-1}(X_{k-1}|Z^{k-1}) \delta X_{k-1} \tag{2.25}$$

$$p_{k|k}(X_k|Z^k) = \frac{l(Z_k|X_k) p_{k|k-1}(X_k|Z^{k-1})}{\int_{\mathcal{X}} l(Z_k|X_k) p_{k|k-1}(X_k|Z^{k-1}) \delta X_k} \propto l(Z_k|X_k) p_{k|k-1}(X_k|Z^{k-1}) \tag{2.26}$$

where $Z^k = Z_1, \ldots, Z_k$ is a sequence of previous measurement-sets, and the set functions $\psi_{k|k-1}(X_k|X_{k-1})$ and $l(Z_k|X_k)$ are of course far from those introduced in Section 2.1 when dealing with the single-target case, but it should be clear from the context or from the fact that the arguments are sets. To grasp a notion, the meaning of the particular densities appearing in the multi-target BRRs may be described as follows.

[7] And therefore the use of Chapman-Kolomogorov equation, and Bayes rule.

[8] See Appendix A, and [37].

- The posterior density $p_{k|k}(X_k|Z^k)$ describes the positions and the nature of the targets, especially the relationships among each other. The prior density $p_{k|k-1}(X_k|Z^{k-1})$ has the same essential meaning for the predicted set of targets.
- The transition density $\psi_{k|k-1}(X_k|X_{k-1})$ describes the transition of the whole set of states. For each target, it allows to model its survival, birth (appearance), death (disappearance), and spawning, from one time instant to another.
- The measurement likelihood function $l(Z_k|X_k)$ can be very general. It describes how the measurements are likely to be generated by the underlying state-set. For each target, it allows to model its undetections, the way the measurements are generated and related to each other.

The main difference from the single-target models is, at least in this thesis, that we are essentially interested in describing the differences among the vectors in the random sets, rather than how they behave separately. In the following, the aim is to present some important properties of the RFSs, in particular the multi-target densities that can be used to model the multi-target state-set, and also the tools that can be used to form the transition and measurement processes. We need to be able to model unions of random sets. For this, the probability generating functionals (PGFLs) and their derivatives are useful. Note that these are not essentially needed to understand the "special-case" derivation of the filters in the following chapters, but they are readily used to derive the general cases, and also many other popular multi-target filters rely on them. Their practical use can be seen in Appendix B. Most of the following concepts can be found e.g. in [4, 21, 38].

2.2.4 Probability Hypothesis Density

The probability hypothesis density (PHD) corresponding to the RFS Ξ, is defined as a function $D_\Xi(\mathbf{x}) : \mathcal{X} \to \mathbb{R}$, such that integrating it over a certain area of the single-target state space results into the expected number of targets in that area,

$$\mathrm{E}[|\Xi \cap \mathcal{S}|] = \int_{\mathcal{S}} D_\Xi(\mathbf{x}) d\mathbf{x} \tag{2.27}$$

It turns out that the PHD can be computed as the first moment density[9] $D(\{\mathbf{x}\})$, defined with the set integral as

$$D_\Xi(\mathbf{x}) = D_\Xi(\{\mathbf{x}\}) \triangleq \int_{\mathcal{X}} p_\Xi(\{\mathbf{x}\} \cup W)\delta W = \int_{\mathcal{X}} \delta_{\mathbf{x}}(W) p_\Xi(W)\delta W \tag{2.28}$$

where

$$\delta_{\mathbf{x}}(W) \triangleq \begin{cases} 0, & \text{if } W = \emptyset \\ \sum_{\mathbf{w}\in W} \delta_{\mathbf{x}}(\mathbf{w}), & \text{otherwise} \end{cases} \tag{2.29}$$

with $\delta_{\mathbf{x}}$ being the Dirac delta centered at $\mathbf{x}$. Therefore, PHD can be viewed as summary statistics of the RFS Ξ.

Since the integral (2.27) is additive (it is not a set integral), we can assume the PHD can be written as a product of a constant and a function that integrates to one. To derive a formula for the constant and the function, ordering of a state-set (2.24) to deal with the PDF should be introduced (remind that this is not typical for FISST). Then, it turns out that the constant is the expected number $\overline{N} = \mathrm{E}_N[N] = \int_{\mathcal{X}} D_\Xi(\mathbf{x})d\mathbf{x}$ of targets in the whole single-target state space $\mathcal{X}$, and the function that integrates to one is a normalized sum of marginal PDFs of ordered $\Xi|N$ being a singleton set, denoted by $p(\mathbf{x})$,

$$D_\Xi(\mathbf{x}) = \overline{N} \cdot p(\mathbf{x}) \tag{2.30}$$

To the author's best knowledge, the result of specifying $p(\mathbf{x})$ does not appear in the literature. A proof of (2.30) and a further discussion about the function $p(\mathbf{x})$ is in Appendix A.5.1.

2.2.5 Probability-generating Functional

The probability generating functional (PGFL) $G_\Xi[h]$ of the RFS Ξ is a tractable representation of the RFS, which allows working with yet engineering-friendly functional calculus, rather than with calculus defined for belief mass measures. It is defined as an integral

[9] The PHD and the first moment density are equal almost everywhere, for more details see Appendix A and Eq. (A.19a).

transform of its PDF, as

$$G_\Xi[h] \triangleq \int_\mathcal{X} h^X \; p_\Xi(X)\delta X \tag{2.31}$$

where

$$h^X = \begin{cases} 1, & \text{if } X = \emptyset \\ \prod_{\mathbf{x}\in X} h(\mathbf{x}), & \text{otherwise} \end{cases} \tag{2.32}$$

with $h : \mathcal{X} \to \mathbb{R}$ being a real-valued function (also called the *test function*), which is the input of the PGFL. It can be viewed as a generalization of the belief mass measure, since $\beta_\Xi(\mathcal{S}) = G_\Xi[\mathbf{1}_\mathcal{S}]$, where $\mathbf{1}_\mathcal{S}$ is the indicator function of the set $\mathcal{S} \subseteq \mathcal{X}$.

2.2.6 Union of Random Sets

Let $\Xi_1, \ldots, \Xi_m$ be statistically independent RFSs with the PDFs $p_{\Xi_1}(X_1), \ldots, p_{\Xi_m}(X_m)$, and PFGLs $G_{\Xi_1}[h], \ldots, G_{\Xi_m}[h]$. Then, their union $\Xi = \Xi_1 \cup \cdots \cup \Xi_m$ is an RFS with PDF given by the convolution formula

$$p_\Xi(X) = \sum_{X_1 \uplus \cdots \uplus X_m = X} p_{\Xi_1}(X_1) \cdots p_{\Xi_m}(X_m) \tag{2.33}$$

where $\uplus$ stands for disjoint union. The term convolution is due to the analogy with a sum of random vectors. In case of having union of exactly two RFSs, the resulting formula for PDFs resembles the usual convolution for random vectors. An example is in Chapter 3, Eq. (3.10c), when deriving a measurement process for a single-target scenario.

The PGFL of Ξ is given by the product formula

$$G_\Xi[h] = G_{\Xi_1}[h] \cdots G_{\Xi_m}[h] \tag{2.34}$$

An analogous product formula holds for the probability measure and belief-mass measure also. We can view those as integral transforms of the PDFs such as Laplace or Fourier transforms, which give rise to the well known relation between convolution and product within the corresponding transformation. Since the convolution is intractable in general and we will need to form e.g. the measurement process as a union of various RFSs, we will rather work with the PGFLs, when constructing more general tracking algorithms.

2.2.7 Functional Derivatives

If we use PGFLs, we have to be able to recover the PHD and PDF back from a PGFL. To do so, the functional derivatives are used. We will define them in a simplified way. For further discussion see [4, 21, 27, 38, 39].

The directional (or Gâteaux) derivative of the functional $G[h]$ in the direction of a function g is given by

$$\frac{\partial G_\Xi[h]}{\partial g} \triangleq \lim_{\epsilon \to 0} \frac{G_\Xi[h + \epsilon g] - G_\Xi[h]}{\epsilon} \tag{2.35}$$

If heuristically $g = \delta_{\mathbf{x}}$ is a Dirac delta centered at $\mathbf{x}$ (or more precisely by Radon-Nikodým derivative [40]), the directional derivative can be called the functional derivative. Abbreviate

$$\frac{\delta G_\Xi[h]}{\delta \mathbf{x}} \triangleq \frac{\partial G_\Xi[h]}{\partial \delta_{\mathbf{x}}} \tag{2.36}$$

For iterated derivatives with respect to $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$ with $\mathbf{x}^1, \ldots, \mathbf{x}^n$ distinct, denote

$$\frac{\delta G_\Xi[h]}{\delta X} \triangleq \frac{\delta^n G_\Xi[h]}{\delta \mathbf{x}^1 \cdots \delta \mathbf{x}^n} = \frac{\partial^n G_\Xi[h]}{\partial \delta_{\mathbf{x}^1} \cdots \partial \delta_{\mathbf{x}^n}} \tag{2.37}$$

which in the case of $X = \emptyset$ is simply $G_\Xi[h]$. Functional derivatives obey standard differentiation rules, well known in undergraduate calculus, such as product rule, exponential rule, etc., see [4, pp. 387-395].

2.2.8 Recovery of the Representations

The key result of FISST calculus is that both the PDF $p_\Xi(X)$ and the PHD $D_\Xi(X)$ of the RFS Ξ can be easily recovered from its PFGL $G_\Xi[h]$ using the functional derivatives,

$$p_\Xi(X) = \left[\frac{\delta G_\Xi[h]}{\delta X}\right]_{h=0} \tag{2.38}$$

$$D_\Xi(\mathbf{x}) = D_\Xi(\{\mathbf{x}\}) = \left[\frac{\delta G_\Xi[h]}{\delta \mathbf{x}}\right]_{h=1} \tag{2.39}$$

Recalling that the PGFL is a generalization of the belief-mass measure, it can be shown that the functional derivative is a co-called set derivative [38, Proposition 1]. Eq. (2.38) can be seen as a set derivative [21, pp. 159-161], constructed using the Radon-Nikodým derivative of a corresponding belief-mass measure. A proof of Eq. (2.39) is given in Appendix A.5.2. Note, that also higher-order moment densities can be recovered from a PGFL similarly to Eq. (2.39).

We now discuss some of the multi-target densities that are crucial for real applications.

2.2.9 The Bernoulli RFS

An RFS Ξ is said to be Bernoulli, if its PDF is given by

$$p_\Xi(X) = \begin{cases} 1-q, & \text{if } X = \emptyset \\ q \cdot p(\mathbf{x}), & \text{if } X = \{\mathbf{x}\} \\ 0, & \text{otherwise} \end{cases} \tag{2.40}$$

i.e. it is singleton with probability $q \in [0,1]$ and the single element (random vector) $\mathbf{x}$ is distributed with the spatial PDF $p(\mathbf{x})$, and it is empty with probability $1-q$. The parameters q and $p(\mathbf{x})$ will be referred to as Bernoulli parameters. One can easily see it is a PDF, i.e. that $\int_\mathcal{X} p_\Xi(X)\delta X = 1 - q + q\int_\mathcal{X} p(\mathbf{x})d\mathbf{x} = 1$. The Bernoulli distribution is often used to model the following

- a target that randomly switches *off* and *on* (a Bernoulli object),
- a measurement that were generated from (a possibly undetected) target.

In case of the Bernoulli distribution, we will also need the PGFL, it is

$$G_\Xi[h] = \int_\mathcal{X} h^X p_\Xi(X)\delta X = \sum_{n=0}^{+\infty} \frac{1}{n!}\int_{\mathcal{X}^n} \prod_{i=1}^{n} h(\mathbf{x}^i) p_\Xi(\{\mathbf{x}^1,\ldots,\mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n \tag{2.41a}$$

$$= p_\Xi(\emptyset) + \int_\mathcal{X} h(\mathbf{x}) p_\Xi(\{\mathbf{x}\}) d\mathbf{x} + \ldots \tag{2.41b}$$

$$= 1 - q + q\underbrace{\int_\mathcal{X} h(\mathbf{x})p(\mathbf{x})d\mathbf{x}}_{p[h]\triangleq} = 1 - q + q \cdot p[h] \tag{2.41c}$$

2.2.10 The Poisson RFS and PPP

An RFS Ξ is said to be Poisson, if

1. its cardinality is Poisson distributed with parameter $\lambda > 0$, i.e.

$$p_N(n) = \frac{e^{-\lambda}\lambda^n}{n!} \tag{2.42}$$

2. the elements of Ξ are i.i.d. with spatial PDF $p(\mathbf{x})$, i.e.

$$p_{\Xi|N}(\{\mathbf{x}^1, \dots, \mathbf{x}^n\}|n) = n! \cdot p_{\mathbf{X}|N}(\mathbf{x}^1, \dots, \mathbf{x}^n|n) = n! \prod_{\mathbf{x} \in X} p(\mathbf{x}) \tag{2.43}$$

When connected together, we can say that Ξ is said to be Poisson, if its PDF is given by

$$p_\Xi(X) = p_N(n) \cdot p_{\Xi|N}(\{\mathbf{x}^1, \dots, \mathbf{x}^n\}|n) = e^{-\lambda} \prod_{\mathbf{x} \in X} \lambda p(\mathbf{x}) \tag{2.44}$$

Note that this is an FISST-based model of the Poisson point process (PPP) known in the point process literature.

In case of the Poisson RFS, we will need its PHD and PGFL. Its PHD is

$$\begin{aligned}
D_\Xi(\mathbf{x}) = \int_{\mathcal{X}} p_\Xi(\{\mathbf{x}\} \cup W)\delta W &= \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} p_\Xi(\{\mathbf{x}, \mathbf{w}^1, \dots, \mathbf{w}^n\}) d\mathbf{w}^1 \cdots d\mathbf{w}^n && \text{(2.45a)} \\
&= \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} e^{-\lambda} \lambda p(\mathbf{x}) \prod_{i=1}^{n} \lambda p(\mathbf{w}^i) d\mathbf{w}^1 \cdots d\mathbf{w}^n && \text{(2.45b)} \\
&= \lambda p(\mathbf{x}) \cdot e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \int_{\mathcal{X}^n} \prod_{i=1}^{n} p(\mathbf{w}^i) d\mathbf{w}^1 \cdots d\mathbf{w}^n && \text{(2.45c)} \\
&= \lambda p(\mathbf{x}) \cdot e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \prod_{i=1}^{n} \underbrace{\int_{\mathcal{X}} p(\mathbf{w}^i) d\mathbf{w}^i}_{=1\ \forall i} && \text{(2.45d)} \\
&= \lambda p(\mathbf{x}) \cdot e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} = \lambda p(\mathbf{x}) \cdot e^{-\lambda} e^{\lambda} = \lambda p(\mathbf{x}) && \text{(2.45e)}
\end{aligned}$$

Note that its PHD is also called intensity in the PPP terminology.

Its PGFL is

$$
\begin{aligned}
G_\Xi[h] = \int_{\mathcal{X}} h^X p_\Xi(X)\delta X &= \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} \prod_{i=1}^{n} h(\mathbf{x}^i) p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(2.46a)} \\
&= \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} \prod_{i=1}^{n} h(\mathbf{x}^i) e^{-\lambda} \prod_{j=1}^{n} \lambda p(\mathbf{x}^j) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(2.46b)} \\
&= e^{-\lambda} \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} \prod_{i=1}^{n} h(\mathbf{x}^i) \lambda^n p(\mathbf{x}^i) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(2.46c)} \\
&= e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \prod_{i=1}^{n} \int_{\mathcal{X}} h(\mathbf{x}^i) p(\mathbf{x}^i) d\mathbf{x}^i && \text{(2.46d)} \\
&= e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} \Bigg(\underbrace{\int_{\mathcal{X}} h(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}}_{p[h] \triangleq} \Bigg)^n = e^{-\lambda} e^{\lambda p[h]} = e^{\lambda p[h] - \lambda} && \text{(2.46e)}
\end{aligned}
$$

In the PPP literature (i.e. non-FISST literature), a superposition and independent (or Bernoulli) thinning are defined as manipulations with PPPs [23, pp. 28-33]. Those will be needed in derivation of the intensity filter in Chapter 5, in order to split targets on those detected and undetected, and merge back again after their respective Bayes-update.

Bernoulli Thinning

Bernoulli, or independent thinning is a process of taking points from a PPP, independently one by one, and keeping them in the realization with probability $q \in \mathbb{R}$ and deleting them with probability $1 - q$. The result is viewed as a PPP with intensity given by the product of the former intensity and q. In another words, for a PPP Ξ with intensity $\lambda p(\mathbf{x})$, the thinned point process is a PPP with intensity $q\lambda p(\mathbf{x})$.

Superposition

Superposition of two PPPs is their union. The result is a PPP with intensity given by summing the intensities of the superposed PPPs. In terms of PPPs Ξ_1 and Ξ_2, with intensities $\lambda_1 p_1(\mathbf{x})$ and $\lambda_2 p_2(\mathbf{x})$, respectively, the result is a PPP with intensity $\lambda p(\mathbf{x}) = \lambda_1 p_1(\mathbf{x}) + \lambda_2 p_2(\mathbf{x})$. Using the FISST, this can be easily verified with the PFGL of a Poisson RFS (2.46e) and the product formula (2.34).

2.3 Gaussian Sum Filter Tools

Gaussian sum filtering (GSF) refers to the linear Gaussian case presented in Section 2.1. For multi-target scenarios, it can be viewed as framework providing tools to work with the Gaussian densities, representing the targets and their transition, measurements, measurement likelihood function, etc. The tools are gating, mixture reduction, and normalization.

2.3.1 Ellipsoidal Gating

When forming the so-called data-to-track association hypotheses, one tries to connect each measurement $\mathbf{z}_k^i$ that arrived at a particular time k with a trajectory of the object to be tracked. Gating refers to deciding which measurements are to be kept to continue working with, and which can be deleted right away. In another words, it means deciding, which measurements might have been generated from a particular object, and which "certainly" have not.

Suppose the measurement-set Z_k arrived at time k. Suppose the measurement prediction PDF $p(\mathbf{z}_k|\mathbf{z}^{k-1})$ is a Gaussian PDF of the form (2.17), i.e. $\mathcal{N}(\mathbf{z}_k; \hat{\mathbf{z}}_{k|k-1}, \mathbf{P}^{\hat{\mathbf{z}}}_{k|k-1})$. The contours of a Gaussian density are (hyper-)ellipsoids, whose shape is determined by its covariance matrix. The covariance matrix gives rise to the squared Mahalanobis distance

$$d^2(\mathbf{z}_k) = (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T (\mathbf{P}^{\hat{\mathbf{z}}}_{k|k-1})^{-1} (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \tag{2.47}$$

which can be evaluated for each measurement $\mathbf{z}_k^i \in Z_k$. Defining the value $G \geq 0$, called the gate, we can test the measurements,

- if $d^2(\mathbf{z}_k^i) > G \Rightarrow$ keep the measurement $\mathbf{z}_k^i$ in Z_k,
- if $d^2(\mathbf{z}_k^i) \leq G \Rightarrow$ delete the measurement $\mathbf{z}_k^i$ from Z_k.

Since $d^2(\mathbf{z}_k)$ obviously obeys the Chi-square distribution with n_z degrees of freedom, denoted with $\chi^2_{n_z}$, the gate G defines the size of the (hyper)-ellipsoid in the following sense. The value $\sqrt{G}$ represents the maximal "sigma-distance" from the mean $\hat{\mathbf{z}}_{k|k-1}$ in the direction of measurement $\mathbf{z}_k$. When passed into the cumulative distribution function of $\chi^2_{n_z}$, the value $P_{\chi^2_{n_z}}(G)$ represent the probability of valid observation falling into the ellipsoid. Graphical illustration of the ellipsoidal gating is in Fig. 2.2. More on Gating can be found e.g. in [1, pp. 107-111][2, pp. 334-338].

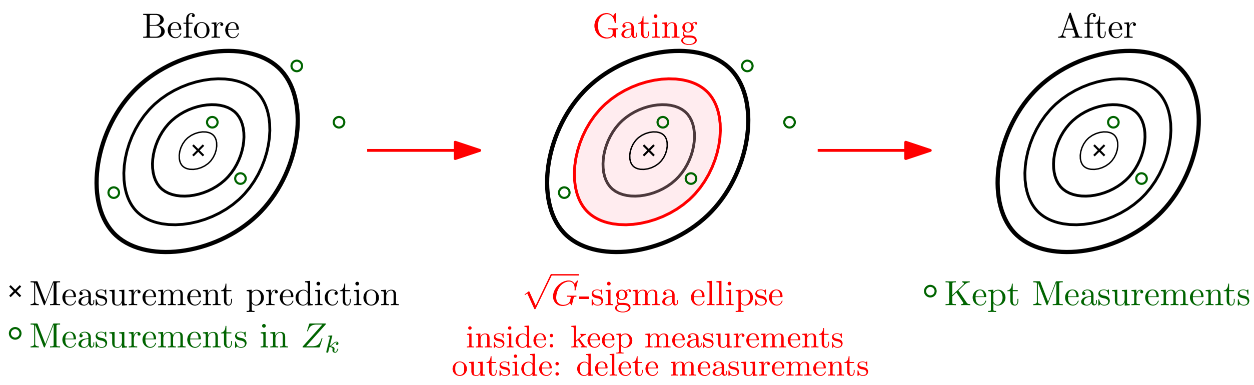


Fig. 2.2: Illustration of ellipsoidal gating for $n_z = 2$.

2.3.2 Reduction of Mixture densities

Suppose having a weighted sum of PDFs. The reduction then refers to decreasing the number of its summands (components).

Deleting components whose weights are lower than a threshold is called **pruning**. Keeping maximally some number of components with the largest weights is called **capping**. For pruning and capping, the weights w^i do not necessarily have to meet the condition $\sum_{i=1}^{M} w^i = 1$. Pruning and capping are considered trivial.

Fusing information from all the given components while yielding a single one is called **merging**. Assume that the given mixture is a GM of the form (2.5). Minimizing the Kullback-Liebler divergence between the former GM and the new, single PDF, leads to the new PDF being Gaussian $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P})$, with

$$\mathbf{m} = \sum_{i=1}^{M} w^i \mathbf{m}^i, \qquad \mathbf{P} = \sum_{i=1}^{M} w^i \Big(\mathbf{P}^i + (\mathbf{m} - \mathbf{m}^i)(\mathbf{m} - \mathbf{m}^i)^T \Big) \tag{2.48}$$

2.3.3 Normalization of Mixture Densities

Suppose having a weighted sum of PDFs whose weights w^i do not meet the condition $\sum_{i=1}^{M} w^i = 1$. This can happen e.g. after performing pruning or capping, but also when deriving a Bayes update using the proportional relation (2.10,2.26). Normalizing the sum means finding new weights, denoted as $\tilde{w}^i$, such that they meet the above condition. The new weights are easily given by $\tilde{w}^i = w^i / (\sum_{i=1}^{M} w^i)$.

CHAPTER 3

Single Object Tracking in Clutter

This chapter is devoted to tracking an object, which is known to be present for the whole tracking time horizon, in presence of undetections and false alarm measurements called the clutter. Graphical illustration of a possible scenario is given in Fig. 3.1.

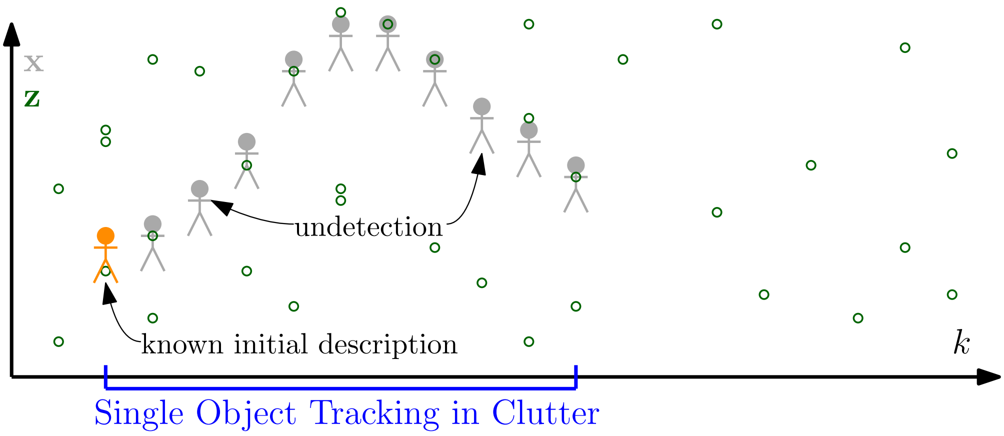
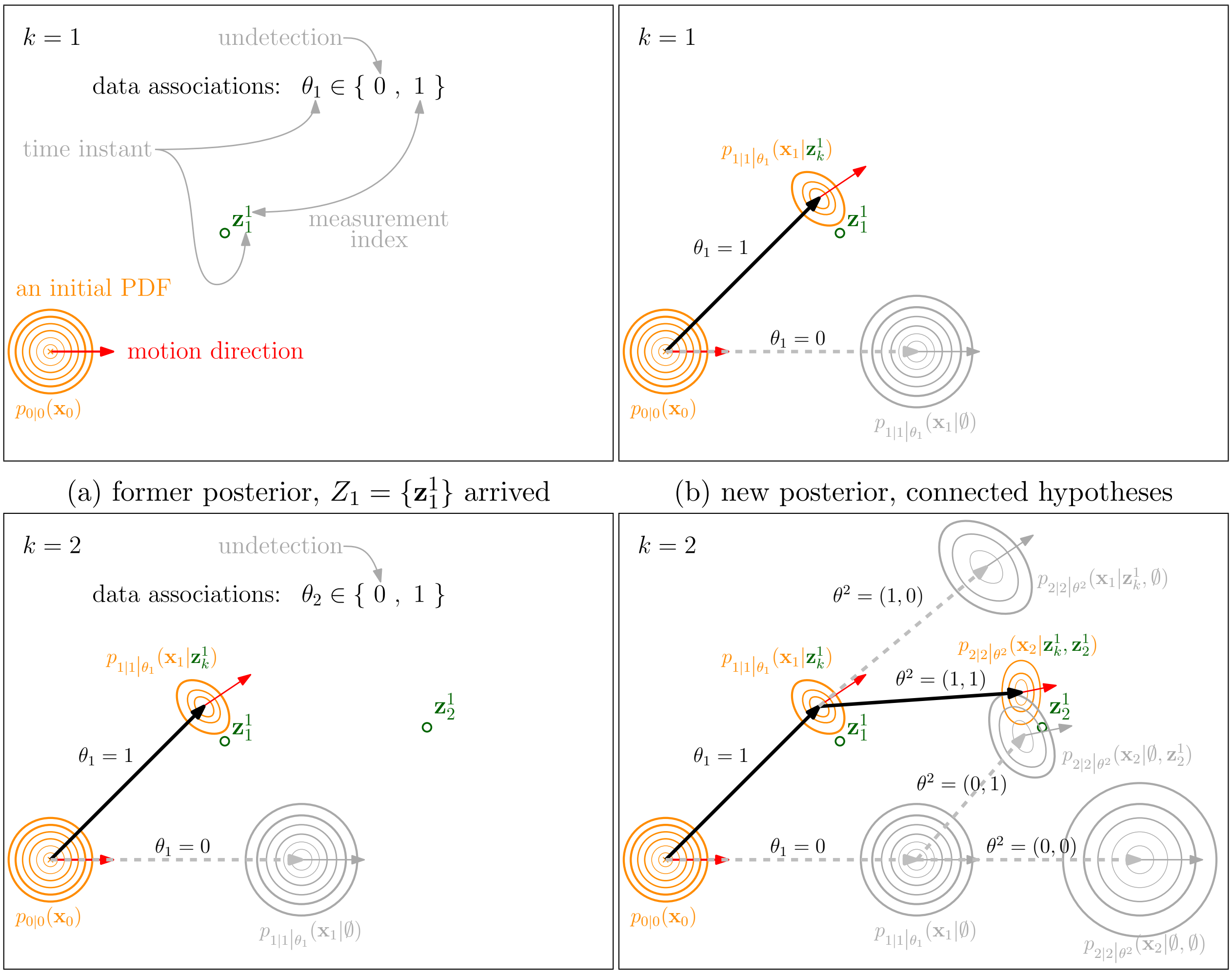


Fig. 3.1: Illustration of single object tracking in clutter.

3.1 Explicit Data Associations

The first problem in this thesis is solved with a rather intuitive approach. The measurements at any time instant k are modeled as an arbitrary ordered realization $\mathbf{Z}_k = (\mathbf{z}_k^1, \ldots, \mathbf{z}_k^{m_k})$ of a measurement point process. The approach used in this section, is to separately try every possibility, i.e. to use every measurement and a hypothesis of undetection to update the knowledge about the state. This can be understood as forming the so-called *data-to-track association hypotheses* denoted by θ, which leads to a special case of the *multiple hypothesis tracking* (MHT) algorithm for a single target case.

This approach is graphically illustrated in Fig. 3.2 for a special case when the target motion is two-dimensional and $\mathbf{z}_k^i$ are noisy position measurements. At the time step $k = 1$, assume a single measurement $\mathbf{z}_1^1$ arrived. The possibilities are such that the measurement was clutter-generated, or such that the measurement was generated by the object. In the first case, it is assumed that the object was undetected and so a predictive PDF is kept in the posterior, and in the second case the measurement is used to perform Bayes update. The corresponding data associations θ_1 are "0" for the undetection, and "1" for the detection by the first (and the only) measurement. This forms two possible hypotheses about the trajectory evolution expressed as conditional PDFs. Similarly, assume a single measurement $\mathbf{z}_2^1$ arrived in the time step $k = 2$. The corresponding data associations θ_2 are analogous, but they are used for both existing hypotheses, hence yielding four new hypotheses in total.



(a) former posterior, $Z_1 = \{\mathbf{z}_1^1\}$ arrived (b) new posterior, connected hypotheses

(c) former posterior, $Z_2 = \{\mathbf{z}_2^1\}$ arrived (d) new posterior, connected hypotheses

Fig. 3.2: Forming the data association hypotheses and trajectories.

Note, that MHT is usually used in multi-target tracking problems [1, 2, 20], where other difficulties arise. The following discussion, mostly inspired by the online course [28], should establish fundamentals of the MHT for a single target case, rather than to explain the general MHT. For other data association techniques, such as using the Viterbi algorithm, see e.g. [41], or [2, pp. 403-282] for others. Modeling without explicit data associations, arising from using sets, will be investigated later in Chapter 4, where a slightly more general assumption of possibly nonexistent object is discussed.

3.2 Derivation of the Filter

First, representation of posterior PDF is discussed. Motivated by the given illustration of MHT approach, namely from Fig. 3.2d, it can be seen that the final knowledge about the state at time k is expressed in terms of some posterior PDFs $p_{k|k|\theta^k}(\mathbf{x}_k|\mathbf{Z}^k, \theta^k)$ conditioned on both the measurement sequence $\mathbf{Z}^k = (\mathbf{Z}_1, \ldots, \mathbf{Z}_k)$ and the data association hypothesis $\theta^k = (\theta_1, \ldots, \theta_k)$. Moreover, assume there exists a probability $\Pr(\theta^k|\mathbf{Z}^k)$ of each hypothesis. Then, from the law of total probability and conditioning, the posterior PDF can be expressed as a mixture PDF

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) = \sum_{\theta^k} p_{k|k,\theta^k}(\mathbf{x}_k, \theta^k|\mathbf{Z}^k) = \sum_{\theta^k} \Pr(\theta^k|\mathbf{Z}^k) \cdot p_{k|k|\theta^k}(\mathbf{x}_k|\mathbf{Z}^k, \theta^k) \tag{3.1}$$

where the symbol $\sum_{\theta^k}$ formally stands for summation over all hypotheses $\sum_{\theta_1=0}^{m_1} \cdots \sum_{\theta_k=0}^{m_k}$. If we index them with some injective function $h_k : \theta^k \to \mathbb{N}$ and $\mathcal{H}_k$ being total number of existing hypotheses (such as $h_k = h_{k-1} + \mathcal{H}_{k-1}\theta_k$), it could be also viewed as $\sum_{h_k=1}^{\mathcal{H}_k}$.

To simplify the notation, abbreviate

$$w_{k|k}^{\theta^k} \triangleq \Pr(\theta^k|\mathbf{Z}^k), \qquad \text{or denoted by} \quad w_{k|k}^{h_k} \tag{3.2}$$

$$p_{k|k}^{\theta^k}(\mathbf{x}_k) \triangleq p_{k|k|\theta^k}(\mathbf{x}_k|\mathbf{Z}^k, \theta^k), \qquad \text{or denoted by} \quad p_{k|k}^{h_k}(\mathbf{x}_k) \tag{3.3}$$

where $w_{k|k}^{\theta^k}$ can be called weights. Hence Eq. (3.1) can be written as

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) = \sum_{\theta^k} w_{k|k}^{\theta^k} p_{k|k}^{\theta^k}(\mathbf{x}_k) = \sum_{h_k=1}^{\mathcal{H}_k} w_{k|k}^{h_k} p_{k|k}^{h_k}(\mathbf{x}_k) \tag{3.4}$$

Now, the time and measurement update will be discussed. Following the discussion in Appendix A, about the generalizations of the Chapman-Kolomogorov equation and Bayes rule for modeling with sets, one can deduce these equations apply in analogous sense to this case as well.

3.2.1 Time-update

Assume that the target motion is described by Markov transition PDF $\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$, then the Chapman-Kolomogorov equation can be used straightforwardly,

$$p_{k|k-1}(\mathbf{x}_k|\mathbf{Z}^{k-1}) = \int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p_{k-1|k-1}(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1})d\mathbf{x}_{k-1} \tag{3.5a}$$

$$= \sum_{\theta^{k-1}} w_{k-1|k-1}^{\theta^{k-1}} \int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p_{k-1|k-1}^{\theta^{k-1}}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \tag{3.5b}$$

$$= \sum_{\theta^{k-1}} w_{k|k-1}^{\theta^{k-1}} p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) = \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} w_{k|k-1}^{h_{k-1}} p_{k|k-1}^{h_{k-1}}(\mathbf{x}_k) \tag{3.5c}$$

Hence for each hypothesis, the time-update of corresponding weights and associated PDFs is given by

$$w_{k|k-1}^{\theta^{k-1}} = w_{k-1|k-1}^{\theta^{k-1}}, \qquad \text{or denoted by} \quad w_{k|k-1}^{h_{k-1}} \tag{3.6}$$

$$p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) = \int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p_{k-1|k-1}^{\theta^{k-1}}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1}, \qquad \text{or denoted by} \quad p_{k|k-1}^{h_{k-1}}(\mathbf{x}_k) \tag{3.7}$$

and the number of hypotheses in (3.5c) remains unchanged.

3.2.2 Bayes-update

For the Bayes-update, we first discuss the measurement model, then we derive a (generalized) measurement likelihood function and the Bayes update from the proportional relation such as in Eq. (2.10,2.26).

Measurement Model

Modeling the underlying point processes with RFSs[1], we can say that the measurement process $\Sigma_k|\mathbf{x}_k$ conditioned on $\mathbf{x}_k$, is modeled as a union of the object-generated measurement-

[1]Regarding the original MHT approach where no sets appear explicitly, modeling measurements as sets is little obscure. We use the RFS approach to emphasize the underlying phenomenological intuition of the measurements to be unordered in reality, and also because the results can be used straightforwardly in the next Chapter.

set $O_k|\mathbf{x}_k$, and the clutter C_k. The RFS $O_k|\mathbf{x}_k$ is assumed to be either empty (undetection), or singleton (detection), such that it is Bernoulli with the PDF

$$p_{O_k|\mathbf{x}_k}(Z_k^o|\mathbf{x}_k) = \begin{cases} 1 - P_D(\mathbf{x}_k), & \text{if } Z_k^o = \emptyset \\ P_D(\mathbf{x}_k)l(\mathbf{z}|\mathbf{x}_k), & \text{if } Z_k^o = \{\mathbf{z}\} \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

where $P_D(\mathbf{x}_k)$ is the probability of detection, $l(\mathbf{z}|\mathbf{x}_k)$ is the measurement likelihood function of the single object, and Z_k^o is the realization of $O_k|\mathbf{x}_k$.

For the clutter-generated measurements, it is common to assume that they are identically distributed in space with some spatial PDF $c(\mathbf{z})$, and their number is Poisson with expected value $\lambda > 0$. In another words, C_k is Poisson with the PDF

$$p_{C_k}(Z_k^c) = e^{-\lambda} \prod_{\mathbf{z} \in Z_k^c} \lambda c(\mathbf{z}) \tag{3.9}$$

where Z_k^c the realization of C_k.

Likelihood Function

Now, a measurement likelihood function, which is a PDF of $\Sigma_k|\mathbf{x}_k = O_k|\mathbf{x}_k \cup C_k$, can be found using multiple ways. Naturally, the measurement set can be assumed ordered, and a simple intuition for the case $Z_k^o = \emptyset$ and $Z_k^o = \{\mathbf{z}\}$ can be used. Since this thesis aims to present these slightly more general approaches, as they are used in more advanced multi-target algorithms, the tools of FISST will be used. Ordering the measurement set into a tuple will be done afterwards.

From the FISST, a convolution formula for PDFs, or product formula for PGFLs can be used. Since for a single target the convolution formula is still tractable, it will be used now, leaving the PGFL version for a more general case later.

From the convolution formula (2.33), it follows that the measurement likelihood is

$$
\begin{aligned}
l(Z_k|\mathbf{x}_k) &\triangleq p_{\Sigma_k|\mathbf{x}_k}(Z_k|\mathbf{x}_k) && \text{(3.10a)}\\
&= \sum_{Z_k^o \uplus Z_k^c = Z_k} p_{O_k|\mathbf{x}_k}(Z_k^o|\mathbf{x}_k) \cdot p_{C_k}(Z_k^c) && \text{(3.10b)}\\
&= \sum_{\substack{Z_k^o \subseteq Z_k \\ |Z_k^o| \leq 1}} p_{O_k|\mathbf{x}_k}(Z_k^o|\mathbf{x}_k) \cdot p_{C_k}(Z_k \setminus Z_k^o) && \text{(3.10c)}\\
&= p_{O_k|\mathbf{x}_k}(\emptyset|\mathbf{x}_k) p_{C_k}(Z_k \setminus \emptyset) + \sum_{\mathbf{z} \in Z_k} p_{O_k|\mathbf{x}_k}(\{\mathbf{z}\}|\mathbf{x}_k) p_{C_k}(Z_k \setminus \{\mathbf{z}\}) && \text{(3.10d)}\\
&= \big(1 - P_D(\mathbf{x}_k)\big) p_{C_k}(Z_k) + \sum_{\mathbf{z} \in Z_k} P_D(\mathbf{x}_k) l(\mathbf{z}|\mathbf{x}_k) \overbrace{e^{-\lambda} \prod_{\mathbf{h} \in Z_k \setminus \{\mathbf{z}\}} \lambda c(\mathbf{h}) \cdot \underbrace{\left(\frac{\lambda c(\mathbf{z})}{\lambda c(\mathbf{z})}\right)}_{1}}^{p_{C_k}(Z_k)} && \text{(3.10e)}\\
&= \big(1 - P_D(\mathbf{x}_k)\big) p_{C_k}(Z_k) + \sum_{\mathbf{z} \in Z_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})} p_{C_k}(Z_k) && \text{(3.10f)}\\
&= p_{C_k}(Z_k) \cdot \left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z} \in Z_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right) && \text{(3.10g)}
\end{aligned}
$$

where $\setminus$ stands for set difference. Now, to integrate this result into the MHT approach, an arbitrary ordering of $Z_k = \{\mathbf{z}_k^1, \dots, \mathbf{z}_k^{m_k}\}$, denoted by $\mathbf{Z}_k = (\mathbf{z}_k^1, \dots, \mathbf{z}_k^{m_k}) = (\mathbf{z}_k^{\theta_k})_{\theta_k=1}^{m_k}$, is formed introducing a summing index $\theta_k = 0, 1, \dots, m_k$, where $\theta_k = 0$ stands for missing detection (modeling undetection of any target). Since m_k is known when the measurements arrive, we can write the resulting likelihood as a PDF of ordered Σ_k, as

$$
\begin{aligned}
l(\mathbf{Z}_k|\mathbf{x}_k) = p_{\mathbf{Z}_k|\mathbf{x}_k}(\mathbf{Z}_k|\mathbf{x}_k) &= \frac{1}{m_k!} \cdot p_{\Sigma_k|\mathbf{x}_k}(Z_k|\mathbf{x}_k) && \text{(3.11a)}\\
&= \left(\frac{e^{-\lambda}}{m_k!} \prod_{\theta_k=1}^{m_k} \lambda c(\mathbf{z}_k^{\theta_k})\right) \cdot \left(\underbrace{1 - P_D(\mathbf{x}_k)}_{\text{summand for } \theta_k = 0} + \sum_{\theta_k=1}^{m_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})}\right) && \text{(3.11b)}
\end{aligned}
$$

Filtering Equation

The general Bayes update for this case in the proportional relation form is

$$
p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) \propto l(\mathbf{Z}_k|\mathbf{x}_k) p_{k|k-1}(\mathbf{x}_k|\mathbf{Z}^{k-1}) \tag{3.12}
$$

Since the term $\frac{e^{-\lambda}}{m_k!}\prod_{\theta_k=1}^{m_k}\lambda c(\mathbf{z}_k^{\theta_k})$ in Eq. (3.11b) is constant with respect to $\mathbf{x}_k$, it can be left to the proportional constant (i.e. omitted). Using the particular form of the prior PDF (3.5c), the resulting proportional relation becomes

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) \propto \left(1 - P_D(\mathbf{x}_k) + \sum_{\theta_k=1}^{m_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})}\right) p_{k|k-1}(\mathbf{x}_k|\mathbf{Z}^{k-1}) \tag{3.13}$$

$$\propto \sum_{\theta^{k-1}} w_{k|k-1}^{\theta^{k-1}} \left(1 - P_D(\mathbf{x}_k) + \sum_{\theta_k=1}^{m_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})}\right) p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) \tag{3.14}$$

and since the posterior has to integrate to one, the resulting equality relation is

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) = \frac{1}{1-\Delta_k^{HT}} \sum_{\theta^{k-1}} w_{k|k-1}^{\theta^{k-1}} \left(1 - P_D(\mathbf{x}_k) + \sum_{\theta_k=1}^{m_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})}\right) p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) \tag{3.15}$$

where

$$\Delta_k^{HT} = \int_{\mathcal{X}} P_D(\mathbf{x}_k) \left(1 - \sum_{\theta_k=1}^{m_k} \frac{l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})}\right) p_{k|k-1}(\mathbf{x}_k|\mathbf{Z}^{k-1}) d\mathbf{x}_k \tag{3.16}$$

Note, that the superscript HT stands for *hypothesis tracking*. It might not seem reasonable to write the normalization constant as $1-\Delta_k^{HT}$, but it is especially useful when compared to the Bernoulli filter later. For the following derivation, we use the proportional relation (3.13).

It can be seen that every existing hypothesis θ^{k-1} is endowed with a data association, expressed by the index θ_k, to yield a new hypothesis $\theta^k = (\theta^{k-1}, \theta_k)$. This process can be understood as trying every measurement to update every existing component of the prior PDF, measuring its probability, and storing it for the future. To close the recursion, let the posterior be written as a mixture PDF as in Eq. (3.4). First, rewrite the proportional relation (3.13) such that every summand is a PDF (and so it integrates to one), multiplied by some constant. For each existing hypothesis θ^{k-1} and for each value of θ_k we define

$$\theta_k = 0 : \begin{cases} c_k^{\theta^k} \triangleq \int_{\mathcal{X}} \big(1 - P_D(\mathbf{x}_k)\big) p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) d\mathbf{x}_k & \text{(3.17a)} \\ p_{k|k}^{\theta^k}(\mathbf{x}_k) = \frac{1}{c_k^{\theta^k}} \big(1 - P_D(\mathbf{x}_k)\big) p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) & \text{(3.17b)} \end{cases}$$

$$\theta_k \geq 1: \begin{cases} c_k^{\theta^k} \triangleq \int_{\mathcal{X}} \dfrac{P_D(\mathbf{x}_k) l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})} p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) d\mathbf{x}_k & \text{(3.18a)} \\ p_{k|k}^{\theta^k}(\mathbf{x}_k) = \dfrac{1}{c_k^{\theta^k}} \dfrac{P_D(\mathbf{x}_k) l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k)}{\lambda c(\mathbf{z}_k^{\theta_k})} p_{k|k-1}^{\theta^{k-1}}(\mathbf{x}_k) & \text{(3.18b)} \end{cases}$$

where c^{θ^k} $\forall \theta^k$ are normalization constants such that the densities $p_{k|k}^{\theta^k}(\mathbf{x}_k)$ integrate to one. This connects every existing hypothesis θ^{k-1} with the data association θ_k. Thus, we get the relation in a compact form

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) \propto \sum_{\theta^k} w_{k|k-1}^{\theta^{k-1}} c_k^{\theta^k} p_{k|k}^{\theta^k}(\mathbf{x}_k) \tag{3.19}$$

After normalizing the mixture (see Section 2.3.3), we get the general updated PDF,

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) = \sum_{\theta^k} w_{k|k}^{\theta^k} p_{k|k}^{\theta^k}(\mathbf{x}_k), \quad \text{where} \quad w_{k|k}^{\theta^k} = \frac{c_k^{\theta^k} w_{k|k-1}^{\theta^{k-1}}}{\sum_{\theta^k} c_k^{\theta^k} w_{k|k-1}^{\theta^{k-1}}} \tag{3.20}$$

or more conveniently using indexing with h_k,

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) = \sum_{h_k=1}^{\mathcal{H}_k} w_{k|k}^{h_k} p_{k|k}^{h_k}(\mathbf{x}_k), \quad \text{where} \quad w_{k|k}^{h_k} = \frac{c_k^{h_k} w_{k|k-1}^{h_{k-1}}}{\sum_{h_k=1}^{\mathcal{H}_k} c_k^{h_k} w_{k|k-1}^{h_{k-1}}} \tag{3.21}$$

where the total number of hypotheses in the posterior PDF is $\mathcal{H}_k = \mathcal{H}_{k-1}(1 + m_k)$. Note, that the normalization constant, $1 - \Delta_k^{HT} = \sum_{h_k=1}^{\mathcal{H}_k} c_k^{h_k} w_{k|k-1}^{h_{k-1}}$.

3.3 Gaussian Sum Filter Implementation

Assume the transition PDF, measurement likelihood, and initial condition are such as in the linear-Gaussian case (2.11-2.13). Also assume constant probability of detection $P_D = P_D(\mathbf{x}_k) \in \mathbb{R}$. There are no additional assumptions for clutter parameters λ and $c(\mathbf{z})$. Then, using the results reviewed in Section 2.1, and some indexing h_k, both the prior and posterior PDFs are Gaussian mixtures, denoted by

$$p_{k|k-1}(\mathbf{x}_k|\mathbf{Z}^{k-1}) = \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} w_{k|k-1}^{h_{k-1}} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^{h_{k-1}}, \mathbf{P}_{k|k-1}^{h_{k-1}}) \tag{3.22}$$

$$p_{k|k}(\mathbf{x}_k|\mathbf{Z}^{k}) = \sum_{h_{k}=1}^{\mathcal{H}_{k}} w_{k|k}^{h_{k}} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^{h_{k}}, \mathbf{P}_{k|k}^{h_{k}}) \tag{3.23}$$

From (3.6,3.7), and the KF equations (2.16a,2.16b), it is obvious that the time-update for each hypothesis become

$$w_{k|k-1}^{h_{k-1}} = w_{k-1|k-1}^{h_{k-1}} \tag{3.24a}$$

$$p_{k|k-1}^{h_{k-1}}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^{h_{k-1}}, \mathbf{P}_{k|k-1}^{h_{k-1}}) \tag{3.24b}$$

$$\text{where} \qquad \mathbf{m}_{k|k-1}^{h_{k-1}} = \mathbf{F}\mathbf{m}_{k-1|k-1}^{h_{k-1}} \tag{3.24c}$$

$$\mathbf{P}_{k|k-1}^{h_{k-1}} = \mathbf{F}\mathbf{P}_{k-1|k-1}^{h_{k-1}}\mathbf{F}^T + \mathbf{Q} \tag{3.24d}$$

For the Bayes-update, from (3.17,3.18), and the KF equations (2.16c-2.16e), the connection between each old and new hypothesis is given by

$$\theta_k = 0 : \begin{cases} c_k^{h_k} = 1 - P_D & \text{(3.25a)} \\ p_{k|k}^{h_k}(\mathbf{x}_k) = \dfrac{1 - P_D}{c_k^{h_k}} \cdot p_{k|k-1}^{h_{k-1}}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^{h_k}, \mathbf{P}_{k|k}^{h_k}) & \text{(3.25b)} \\ \text{where} \quad \mathbf{m}_{k|k}^{h_k} = \mathbf{m}_{k|k-1}^{h_{k-1}} & \text{(3.25c)} \\ \mathbf{P}_{k|k}^{h_k} = \mathbf{P}_{k|k-1}^{h_{k-1}} & \text{(3.25d)} \end{cases}$$

$$
\theta_k \geq 1 : \begin{cases}
c_k^{h_k} \triangleq \dfrac{P_D}{\lambda c(\mathbf{z}_k^{\theta_k})} \displaystyle\int_{\mathcal{X}} l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k) p_{k|k-1}^{h_{k-1}}(\mathbf{x}_k) d\mathbf{x}_k & \text{(3.26a)} \\
\quad = \dfrac{P_D}{\lambda c(\mathbf{z}_k^{\theta_k})} \cdot \mathcal{N}(\mathbf{z}_k^{\theta_k}; \hat{\mathbf{z}}_{k|k-1}^{h_{k-1}}, \mathbf{P}_{k|k-1}^{\hat{\mathbf{z}},h_{k-1}}) & \text{(3.26b)} \\
\text{where} \quad \hat{\mathbf{z}}_{k|k-1}^{h_{k-1}} = \mathbf{H}\mathbf{m}_{k|k-1}^{h_{k-1}} & \text{(3.26c)} \\
\mathbf{P}_{k|k-1}^{\hat{\mathbf{z}},h_{k-1}} = \mathbf{H}\mathbf{P}_{k|k-1}^{h_{k-1}}\mathbf{H}^T + \mathbf{R} & \text{(3.26d)} \\
p_{k|k}^{h_k}(\mathbf{x}_k) = \dfrac{P_D}{c_k^{\theta_{1:k}} \lambda c(\mathbf{z}_k^{\theta_k})} l(\mathbf{z}_k^{\theta_k}|\mathbf{x}_k) p_{k|k-1}^{h_{k-1}}(\mathbf{x}_k) & \text{(3.26e)} \\
\quad = \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^{h_k}, \mathbf{P}_{k|k}^{h_k}) & \text{(3.26f)} \\
\text{where} \quad \mathbf{m}_{k|k}^{h_k} = \mathbf{m}_{k|k-1}^{h_{k-1}} + \mathbf{K}_k(\mathbf{z}_k^{\theta_k} - \mathbf{H}\mathbf{m}_{k|k-1}^{h_{k-1}}) & \text{(3.26g)} \\
\mathbf{P}_{k|k}^{h_k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1}^{h_{k-1}} & \text{(3.26h)} \\
\text{with} \quad \mathbf{K}_k = \mathbf{P}_{k|k-1}^{h_{k-1}}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k|k-1}^{h_{k-1}}\mathbf{H}^T + \mathbf{R})^{-1} & \text{(3.26i)}
\end{cases}
$$

and so the relation (3.19) becomes

$$
\begin{aligned}
p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k) \propto & \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} w_{k|k-1}^{h_{k-1}}(1-P_D)\mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^{h_{k-1}}, \mathbf{P}_{k|k-1}^{h_{k-1}}) + \\
& + \sum_{h_{k-1}=1}^{\mathcal{H}_{k-1}} \sum_{\theta_k=1}^{m_k} w_{k|k-1}^{h_{k-1}} \frac{P_D \mathcal{N}(\mathbf{z}_k^{\theta_k}; \hat{\mathbf{z}}_{k|k-1}^{h_{k-1}}, \mathbf{P}_{k|k-1}^{\hat{\mathbf{z}},h_{k-1}})}{\lambda c(\mathbf{z}_k^{\theta_k})} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^{h_k}, \mathbf{P}_{k|k}^{h_k})
\end{aligned} \qquad (3.27)
$$

and the final equality is given by normalizing the weights (3.21). Note that the hypotheses indexing h_k is arbitrary since injective. Recall the possibility of using $h_k = h_{k-1} + \mathcal{H}_{k-1}\theta_k$.

Since the total number of hypotheses increases over time, it is necessary to use some of the tools presented in Section 2.3. When computing the measurement prediction to form the constant $c_k^{h_k}$ in Eq. (3.26b), the gating can be performed. It should be noted that the probability of each hypothesis depends also on the clutter spatial PDF $c(\mathbf{z})$. If the clutter is not uniform, i.e. $c(\mathbf{z})$ is not constant, then the gating, as presented in Section 2.3, may not be plausible to decide which measurements to use. The pruning, merging and capping can be done after the update, followed by normalizing the reduced posterior.

Two algorithms arise as a special reduction strategies [1, 2, 3, 28], that are the *nearest neighbor* (NN) filter [42] and the *probabilistic data association* (PDA) filter [43, 44]. Both of them keep only a single hypothesis in the posterior PDF. Also, in both of them, the gating can be performed. In the NN filter, only the measurement which is the closest to the predictive one is used, i.e. the one with the greatest value[2] of $d^2(\mathbf{z})$, if any, and all the other measurements are left without notice. In the PDA filter, all the resulting hypotheses, in the sense of GM components, are fused using merging. The strategies are graphically compared in Fig. 3.3. Note, that in general (when not using the GSF implementation), the NN and PDA are viewed as standalone filters, and should not be addressed as special cases of MHT, see e.g. [5].

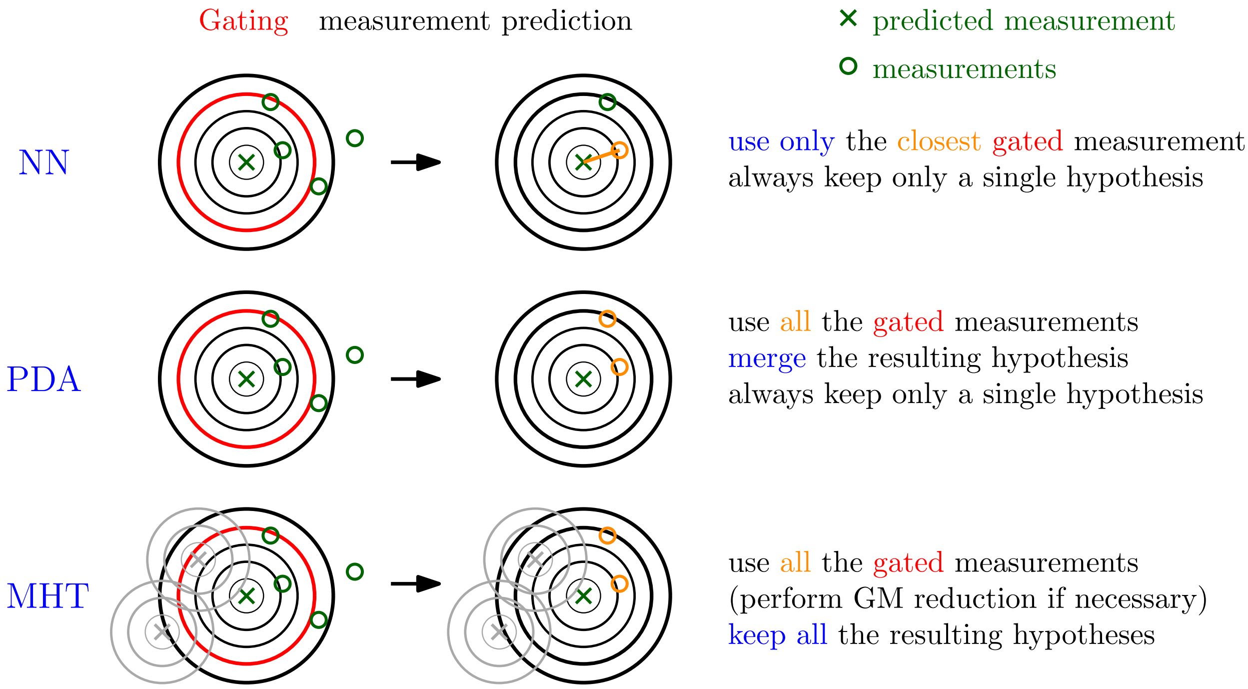


Fig. 3.3: Relation among NN, PDA, and MHT for single target.

In order to form easy-to-understand results, tracking filters are desired to yield estimates as points. In this case, the *minimum mean square error* (MMSE) estimate, and the *most probable hypothesis* (MPH) estimate can be formed. In the GSF framework, the MMSE estimate is the mean of Gaussian PDF, which arises when all hypotheses are merged. The MPH estimate is simply the mean of the component with the largest weight. Since the probabilities of each hypothesis vary in time, it is not granted that every single estimate from a sequence of MPH estimates result from a single "true" hypothesis. That is, there may appear some "jumps" in the resulting trajectory estimate, unreasonable from the motion model point of view. A possible, yet computationally demanding solution would be to store every MPH trajectory estimate through time and to switch between those trajectories.

[2] See Eq. (2.47).

CHAPTER 4

Bernoulli Object Tracking in Clutter

This chapter is devoted to tracking an object, that randomly switches *on/off* in the sense of its existence and that appears to be in presence of undetections and clutter. Moreover, it is assumed that the initial probabilistic description may not be given. A graphical illustration of a possible scenario is given in Fig. 4.1.

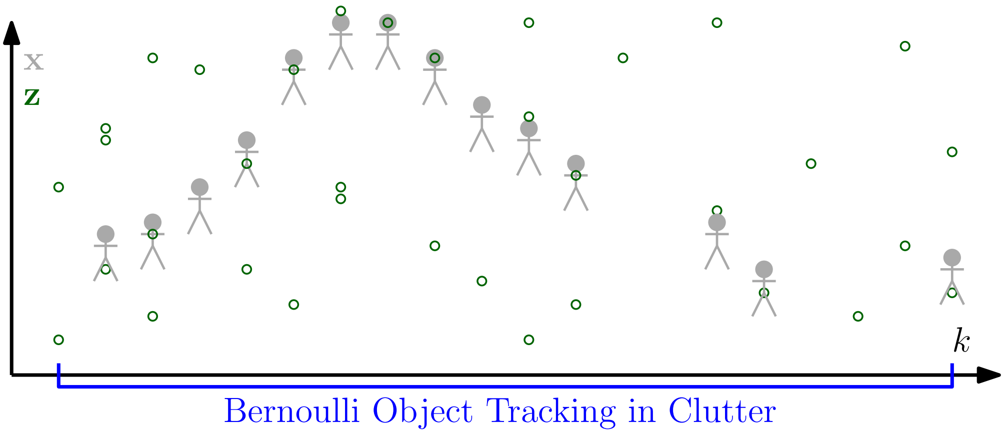


Fig. 4.1: Illustration of Bernoulli object tracking in clutter.

4.1 Avoiding Explicit Data Associations

It would be possible to continue modeling within the MHT approach, however, forming explicit data associations has been criticized due to multiple reasons [4, 35]. We briefly review some of those.

From Eq. (3.1), it is seen that the data association hypothesis θ^k is treated as random variable, which is to be jointly estimated with the system state $\mathbf{x}_k$, i.e. the PDF $p_{k|k,\theta^k}(\mathbf{x}_k, \theta^k|\mathbf{Z}^k)$ is introduced. Since, by an intuition, it is assumed that a point object has a single trajectory, we would expect there exists a particular data association and hypothesis such that it is the true one in reality, i.e. θ^k must be a nonrandom sequence. Since the Bayes filtering framework assumes the estimated entities to be random, the Bayes-optimalilty statement is theoretically violated. Also, noting $\theta_k \in \{0, \dots, m_k\}$, the number of measurements is implicitly included into the estimated values, which is theoretically problematic as well. Moreover, the question is, whether forming data association hypotheses even makes sense in the case of closely-spaced targets, or mainly, in case of other types of measurements, such as pixelized image, which leads to the *track before detect* (TBD) tracking approach. This should serve as a motivation to leave modeling with explicit data associations and turn to some other general approach. That is, in this Chapter using point processes modeled with RFSs will be considered. The Bernoulli filter, also called the joint target detection and tracking (JoTT) filter, is derived in [4, 45], and is a basis of many modern MTT filters.

4.2 Derivation of the Filter

First, the posterior uncertainty representation of the object is discussed. As stated in Section 2.2.9, a target that randomly switches *off* and *on* can be modeled with a Bernoulli RFS which is either empty (object missing), or singleton (object present). Consider the posterior PDF, now denoted with the symbol f, to be

$$f_{k|k}(X_k|Z^k) = \begin{cases} 1 - q_{k|k}, & \text{if } X_k = \emptyset \\ q_{k|k} p_{k|k}(\mathbf{x}_k), & \text{if } X_k = \{\mathbf{x}_k\} \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where $q_{k|k} = \Pr(|X_k| = 1 \big| Z^k)$ is the posterior probability of existence, and $p_{k|k}(\mathbf{x}_k)$ is the posterior spatial PDF of the single target. Since those are just parameters of the actual posterior PDF, there si no need to write any conditioning on Z^k explicitly. The subscripts $k|k$ should just indicate to which filtering stage is a particular parameter related to. In the following, the goal is to find out how these parameters propagate through the motion and measurement models.

4.2.1 Time-update

For the time-update, we will form a motion model, that is a multi-target Markov transition density $\psi_{k|k-1}$, which will be substituted into the multi-target Chapman Kolomogorov Eq. (2.25) to yield a multi-target prior PDF.

Motion Model

For a maximally single object being present, the sets X_k and X_{k-1}, which are inputs of $\psi_{k|k-1}$, can be both empty or singleton. If a target was not present at time $k-1$, then it may appear with state $\mathbf{x}_k$ with probability of birth P_B with the spatial PDF $p_B(\mathbf{x}_k)$, or stay not present with the probability $1 - P_B$. On the other hand, if the object was present, then it may survive to the next time step k with probability P_S and move with respect to the single-target Markov density $\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$, or disappear with probability $1 - P_S$. This intuition gives rise to the multi-target Markov transition PDF of a form

$$\psi_{k|k-1}(X_k|X_{k-1}) = \begin{cases} 1 - P_B, & \text{if } X_k = \emptyset \quad \& \ X_{k-1} = \emptyset \\ P_B \cdot p_B(\mathbf{x}_k), & \text{if } X_k = \{\mathbf{x}_k\} \ \& \ X_{k-1} = \emptyset \\ 1 - P_S, & \text{if } X_k = \emptyset \quad \& \ X_{k-1} = \{\mathbf{x}_{k-1}\} \\ P_S \cdot \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}), & \text{if } X_k = \{\mathbf{x}_k\} \ \& \ X_{k-1} = \{\mathbf{x}_{k-1}\} \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

Whether $\psi_{k|k-1}$ refers to single-target or multi-target transition PDF should be clear from the context, or noting that its arguments are sets.

Prediction Equation

Substitute (4.1) and (4.2) into (2.25). From the definition of the set integral, we get

$$f_{k|k-1}(X_k|Z^{k-1}) = \int_{\mathcal{X}} \psi_{k|k-1}(X_k|X_{k-1}) f_{k-1|k-1}(X_{k-1}|Z^{k-1}) \delta X_{k-1} \tag{4.3a}$$

$$= \psi_{k|k-1}(X_k|\emptyset) f_{k-1|k-1}(\emptyset|Z^{k-1}) + + \int_{\mathcal{X}} \psi_{k|k-1}(X_k|\{\mathbf{x}_k\}) f_{k-1|k-1}(\{\mathbf{x}_k\}|Z^{k-1}) d\mathbf{x}_{k-1} + 0 \tag{4.3b}$$

X_k is assumed to be either empty or singleton. For the case $X_k = \emptyset$ we have

$$f_{k|k-1}(\emptyset|Z^{k-1}) = (1-P_B)(1-q_{k-1|k-1}) + \int_{\mathcal{X}} (1-P_S)q_{k-1|k-1}p_{k-1|k-1}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \tag{4.4a}$$

$$= (1-P_B)(1-q_{k-1|k-1}) + (1-P_S)q_{k-1|k-1} \tag{4.4b}$$

$$= 1 - \underbrace{\big(P_B(1-q_{k-1|k-1}) + P_S q_{k-1|k-1}\big)}_{q_{k|k-1}\triangleq} \tag{4.4c}$$

and for $X_k = \{\mathbf{x}_k\}$ we have

$$\begin{aligned} f_{k|k-1}(\{\mathbf{x}_k\}|Z^{k-1}) &= P_B \cdot p_B(\mathbf{x}_k)(1-q_{k-1|k-1}) + \\ &\quad + \int_{\mathcal{X}} P_S \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})q_{k-1|k-1}p_{k-1|k-1}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \qquad (4.5a) \\ &= P_B(1-q_{k-1|k-1})p_B(\mathbf{x}_k) + \\ &\quad + P_S q_{k-1|k-1}\int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p_{k-1|k-1}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \qquad (4.5b) \end{aligned}$$

which can be written as a constant multiplied by a function of $\mathbf{x}_k$. One could verify that the resulting PDF $f_{k|k-1}(X_k|Z^{k-1})$ integrates to one in terms of the set integral, and hence it is a multi-target PDF. This implies that it can be rewritten into the form of the Bernoulli PDF

$$f_{k|k-1}(X_k|Z^{k-1}) = \begin{cases} 1 - q_{k|k-1}, & \text{if } X_k = \emptyset \\ q_{k|k-1}p_{k|k-1}(\mathbf{x}_k), & \text{if } X_k = \{\mathbf{x}_k\} \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

where the parameters are

$$q_{k|k-1} = P_B(1-q_{k-1|k-1}) + P_S q_{k-1|k-1} \tag{4.7}$$

and

$$
\begin{aligned}
q_{k|k-1}p_{k|k-1}(\mathbf{x}_k) &\stackrel{!}{=} P_B(1-q_{k-1|k-1})p_B(\mathbf{x}_k)+ \\
&\quad + P_S q_{k-1|k-1}\int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p_{k-1|k-1}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \qquad (4.8a)\\
\Rightarrow \quad p_{k|k-1}(\mathbf{x}_k) &= \frac{P_B(1-q_{k-1|k-1})}{q_{k|k-1}}p_B(\mathbf{x}_k)+ \\
&\quad + \frac{P_S q_{k-1|k-1}}{q_{k|k-1}}\int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})p_{k-1|k-1}(\mathbf{x}_{k-1})d\mathbf{x}_{k-1} \qquad (4.8b)
\end{aligned}
$$

4.2.2 Bayes-update

For the Bayes-update, the measurement model will be easily adopted from the one derived in previous Chapter 3. Then, the formal Bayes-update will be derived without making any data association hypotheses.

Measurement Model

Now, the measurement process $\Sigma_k|\Xi_k$ is conditioned on both the empty, and singleton state-set,

$$
\Sigma_k|\Xi_k = \begin{cases} C_k, & \text{if } \Xi_k = \emptyset \\ O_k|\mathbf{x}_k \cup C_k, & \text{if } \Xi_k = \{\mathbf{x}_k\} \\ \text{undefined}, & \text{otherwise} \end{cases} \qquad (4.9)
$$

since $O_k|\mathbf{x}_k \cup C_k = \Sigma_k|\mathbf{x}_k$ was already investigated in previous Chapter 3, we can readily write the measurement likelihood function,

$$
\begin{aligned}
l(Z_k|X_k) &\triangleq p_{\Sigma_k|\Xi_k}(Z_k|X_k) \qquad (4.10a)\\
&= \begin{cases} p_{C_k}(Z_k), & \text{if } X_k = \emptyset \\ p_{C_k}(Z_k)\left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z}\in Z_k} \frac{P_D(\mathbf{x}_k)l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right), & \text{if } X_k = \{\mathbf{x}_k\} \\ 0, & \text{otherwise} \end{cases} \qquad (4.10b)
\end{aligned}
$$

Again, whether l refers to single-target or multi-target likelihood function should be clear from the context, or noting that its arguments are sets.

Filtering Equation

First derive the normalization constant of the multi-target Bayes rule (2.26),

$$
\begin{aligned}
K &\triangleq \int_{\mathcal{X}} l(Z_k|X_k) f_{k|k-1}(X_k|Z^{k-1})\delta X_k && \text{(4.11a)}\\
&= l(Z_k|\emptyset) f_{k|k-1}(\emptyset|Z^{k-1}) + \int_{\mathcal{X}} l(Z_k|\{\mathbf{x}_k\}) f_{k|k-1}(\{\mathbf{x}_k\}|Z^{k-1}) d\mathbf{x}_k + 0 && \text{(4.11b)}\\
&= p_{C_k}(Z_k)(1-q_{k|k-1}) + \\
&\qquad + p_{C_k}(Z_k) q_{k|k-1} \int_{\mathcal{X}} \left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z}\in Z_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right) p_{k|k-1}(\mathbf{x}_k) d\mathbf{x}_k && \text{(4.11c)}\\
&= p_{C_k}(Z_k)\Bigg[1 - q_{k|k-1} \underbrace{\int_{\mathcal{X}} P_D(\mathbf{x}_k)\left(1 - \sum_{\mathbf{z}\in Z_k} \frac{l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right) p_{k|k-1}(\mathbf{x}_k) d\mathbf{x}_k}_{\Delta_k^{BT}\triangleq}\Bigg] && \text{(4.11d)}\\
&= p_{C_k}(Z_k)\left(1 - q_{k|k-1}\Delta_k^{BT}\right) && \text{(4.11e)}
\end{aligned}
$$

where the superscript BT stands for *Bernoulli tracking*. Using this result, the final multi-target posterior PDF is given by

$$
\begin{aligned}
f_{k|k}(X_k|Z^k) &= \frac{l(Z_k|X_k) f_{k|k-1}(X_k|Z^{k-1})}{K} && \text{(4.12a)}\\
&= \begin{cases}
\frac{\cancel{p_{C_k}(Z_k)}(1-q_{k|k-1})}{\cancel{p_{C_k}(Z_k)}\left(1-q_{k|k-1}\Delta_k^{BT}\right)}, & \text{if } X_k = \emptyset\\
\frac{\cancel{p_{C_k}(Z_k)}\left(1-P_D(\mathbf{x}_k)+\sum_{\mathbf{z}\in Z_k}\frac{P_D(\mathbf{x}_k)l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right)q_{k|k-1}p_{k|k-1}(\mathbf{x}_k)}{\cancel{p_{C_k}(Z_k)}\left(1-q_{k|k-1}\Delta_k^{BT}\right)}, & \text{if } X_k = \{\mathbf{x}_k\}\\
0, & \text{otherwise}
\end{cases} && \text{(4.12b)}
\end{aligned}
$$

which can be rewritten in the form such that it is Bernoulli[1] with the parameters

$$
1 - q_{k|k} \stackrel{!}{=} \frac{1 - q_{k|k-1}}{1 - q_{k|k-1}\Delta_k^{BT}} \quad \Rightarrow \quad q_{k|k} = \frac{1-\Delta_k^{BT}}{1 - q_{k|k-1}\Delta_k^{BT}} q_{k|k-1} \tag{4.13}
$$

[1]We can now see, that the Bernoulli PDF is a multi-target conjugate prior with respect to the given likelihood function.

and

$$q_{k|k}p_{k|k}(\mathbf{x}_k) \stackrel{!}{=} \frac{\left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z} \in Z_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right) q_{k|k-1} p_{k|k-1}(\mathbf{x}_k)}{1 - q_{k|k-1}\Delta_k^{BT}} \tag{4.14a}$$

$$\Rightarrow \quad p_{k|k}(\mathbf{x}_k) = \frac{1}{1 - \Delta_k^{BT}} \left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z} \in Z_k} \frac{P_D(\mathbf{x}_k) l(\mathbf{z}|\mathbf{x}_k)}{\lambda c(\mathbf{z})}\right) p_{k|k-1}(\mathbf{x}_k) \tag{4.14b}$$

To sum up, we got a recursion for both the probability of existence (4.7,4.13) and the object spatial PDF (4.8b,4.14b), which define the prior and posterior multi-target Bernoulli PDFs.

4.3 Gaussian Sum Filter Implementation

As well as in the MHT case, assume the single-target transition PDF, and measurement likelihood are such as in the linear-Gaussian case (2.11,2.12). Also assume constant probability of detection $P_D = P_D(\mathbf{x}_k)$. Again, there are no additional assumptions for the clutter parameters λ and $c(\mathbf{z})$. The initial PDF from (2.13) can be used as $p_{0|0}(\mathbf{x}_k)$, and if so, the initial probability of existence $q_{0|0}$ has to be given as well. For the multi-target transition model, consider the target birth PDF to be a GM of the form

$$p_B(\mathbf{x}_k) = \sum_{i=1}^{N_{B,k}} w_{B,k}^i \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{B,k}^i, \mathbf{Q}_{B,k}^i) \tag{4.15}$$

where $N_{B,k}$ is the number of birth-originated components at time k, and assume P_B and P_S are specified. If the initial uncertainty about the target state is not known, it could be left undefined, starting the recursion from the prediction step, with the target birth description understood as the predicted initial PDF. Also, heuristic or ad-hoc logic-based methods of track formation exist, see [3, pp. 322-329][1, pp. 153-168].

Such assumptions result in the spatial PDFs of prior and posterior Bernoulli RFSs being Gaussian mixtures, denoted by

$$p_{k|k-1}(\mathbf{x}_k) = \sum_{h=1}^{\mathcal{H}_{k|k-1}} w_{k|k-1}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^h, \mathbf{P}_{k|k-1}^h) \tag{4.16}$$

$$p_{k|k}(\mathbf{x}_k) = \sum_{h=1}^{\mathcal{H}_{k|k}} w_{k|k}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h) \tag{4.17}$$

where $\mathcal{H}_{k|k-1}$, and $\mathcal{H}_{k|k}$ are total number of components in the prior and posterior PDFs, respectively.

Assume the posterior probability of existence from the last time instant is $q_{k-1|k-1}$. From (4.8b), and the KF equations (2.16a,2.16b), the time-update of the spatial PDF is

$$p_{k|k-1}(\mathbf{x}_k) = \frac{P_B(1-q_{k-1|k-1})}{q_{k|k-1}} \sum_{i=1}^{N_{B,k}} w_{B,k}^i \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{B,k}^i, \mathbf{Q}_{B,k}^i) + \\ + \frac{P_S q_{k-1|k-1}}{q_{k|k-1}} \sum_{h=1}^{\mathcal{H}_{k-1|k-1}} w_{k-1|k-1}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^h, \mathbf{P}_{k|k-1}^h) \tag{4.18a}$$

$$\text{where} \quad \mathbf{m}_{k|k-1}^h = \mathbf{F}\mathbf{m}_{k-1|k-1}^h \tag{4.18b}$$

$$\mathbf{P}_{k|k-1}^h = \mathbf{F}\mathbf{P}_{k-1|k-1}^h \mathbf{F}^T + \mathbf{Q} \tag{4.18c}$$

where the time-updated probability of existence $q_{k|k-1}$ is simply given by Eq. (4.7). The prior spatial PDF (4.18a) can be easily rewritten into a GM of the form (4.16), where $\mathcal{H}_{k|k-1} = \mathcal{H}_{k-1|k-1} + N_{B,k}$ is the prior number of GM components.

The Bernoulli posterior spatial PDF is computed in the same manner as the MHT posterior PDF (3.25,3.26) in Section 3.3, because:

- here, the GM components are not essentially hypotheses, but can be indexed by h as well as h_k, and so they can be interpreted as hypotheses,
- the summation $\sum_{\mathbf{z}\in Z_k}$ can be rewritten using arbitrary indexing of the elements of Z_k, such as with θ_k as $\sum_{\theta_k=1}^{|Z_k|}$, only the index should not be called association hypothesis,
- the constant Δ_k^{BT} (4.11d) is therefore formally the same as Δ_k^{HT} (3.16),
- update Eq. (4.14b) is then formally equivalent to (3.15), and so, schematically

$$p_{k|k}(\mathbf{x}_k) \propto \sum_{h=1}^{\mathcal{H}_{k|k-1}} w_{k|k-1}^h (1-P_D) \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^h, \mathbf{P}_{k|k-1}^h) + \\ + \sum_{h=1}^{\mathcal{H}_{k|k-1}} \sum_{\theta_k=1}^{|Z_k|} w_{k|k-1}^h \frac{P_D \mathcal{N}(\mathbf{z}_k^{\theta_k}; \text{"prediction"})}{\lambda c(\mathbf{z}_k^{\theta_k})} \mathcal{N}(\mathbf{x}_k; \text{"update with } \mathbf{z}_k^{\theta_k}\text{"}) \tag{4.19}$$

but note, that since the time-update differs, each algorithm outputs different results for the spatial PDF. Precisely speaking, the Bernoulli parameter $p_{k|k}(\mathbf{x}_k)$ is not equal to $p_{k|k}(\mathbf{x}_k|\mathbf{Z}^k)$

from Chapter 3, as the time evolves.

In order to decrease the computational costs, the GM reduction discussed in Section 3.3 should also be adopted by the Bernoulli filter. It is to say that in order to compute the updated probability of existence $q_{k|k}$, the value Δ_k^{BT} is needed. In theory, this value follows from the normalization of the mixture (4.19), but can also be computed from this mixture being already reduced too. From the implementation point of view, it is a matter of choice on which information do we want to base the estimate of $q_{k|k}$. One or the other, assume the spatial posterior PDF (4.19) being rewritten as

$$p_{k|k}(\mathbf{x}_k) \propto \sum_{h=1}^{\mathcal{H}_{k|k}} \tilde{w}_{k|k}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h) \tag{4.20}$$

where $\mathcal{H}_{k|k}$ denotes the number of components of possibly reduced mixture, and $\tilde{w}_{k|k-1}^h$ represents the un-normalized weights. Then, the final spatial PDF is given by

$$p_{k|k}(\mathbf{x}_k) = \sum_{h=1}^{\mathcal{H}_k} w_{k|k}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h), \quad \text{where} \quad w_{k|k}^h = \frac{\tilde{w}_{k|k}^h}{\sum_{h=1}^{\mathcal{H}_{k|k}} \tilde{w}_{k|k}^h} \tag{4.21}$$

and $q_{k|k}$ is readily given by Eq. (4.13) using $\Delta_k^{BT} = 1 - \sum_{h=1}^{\mathcal{H}_{k|k}} \tilde{w}_{k|k}^h$.

Both the MMSE and MPH estimates can be adopted to the Bernoulli filter, but we argue that MPH estimate should be called something like "strongest component estimate". The resulting quantity of the Bernoulli filter, which describes the object spatially, is the *whole* spatial PDF. It is not expected, that there should exist one and only "true" component in reality, called hypothesis, as in the MHT case.

If we had continued with the explicit data association approach, the resulting algorithms [3] would be similar to the Bernoulli filter. They do not model the target birth in the prediction, but rather deal with a track formation logic. If the PDA reduction strategy is used, the resulting algorithm is commonly referred to as *integrated* PDA (IPDA) filter [3]. However, to the author's best knowledge, nothing like *integrated* NN (INN) filter appears in the literature.

CHAPTER 5

Multiple Objects Tracking in Clutter

This chapter is devoted to tracking multiple moving objects in presence of undetections and clutter. Each of the objects can randomly appear and disappear, and its initial probabilistic description may not be given. The goal is to jointly estimate the number of objects and their states (or also trajectories). A graphical illustration is in Fig. 5.1.

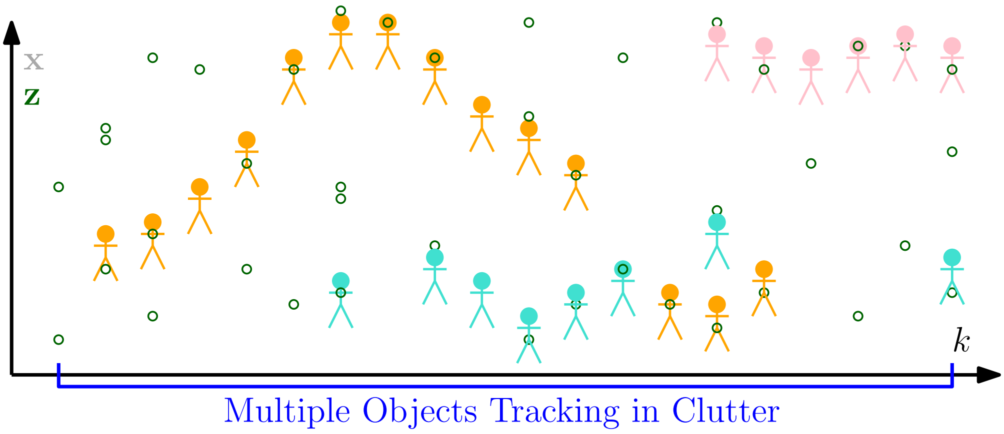


Fig. 5.1: Illustration of multiple objects tracking in clutter.

If we consider that the only data given to the algorithm, is a set of measurements at a time, we argue that such problem is challenging. Moreover, even if we knew true positions of the objects (imagine the characters not colored), deducing what the trajectories were, is ambiguous. In reality, circumstances might be better disposed than they are in Fig. 5.1, but it should be noted that the problem is generally assumed to be of such nature.

5.1 Approaches to MTT

In the previous chapters, we investigated two conceptually different approaches to single-target tracking; in general, forming explicit data association hypotheses or avoiding them. The algorithms dealing with the MTT problem can be also based on those concepts. However, using RFSs does not generally imply avoiding data associations.

In this thesis, we are interested in a specific assumption or approximation, such that the moving objects are i.i.d.. This assumption leads to mathematically elegant algorithms, that are similar to the single-target tracking case, namely with respect to the computational costs. To provide a context and to motivate the approach, we first give a quick overview and taxonomy of the MTT algorithms.

5.1.1 MTT Outlook

The most recent, yet already not up-to-date, survey of MTT can be found in [5]. For the MHT approach survey see [20], and for RFS see [22]. For complementary information and other approaches see e.g. [46, 47], and of course [1, 2, 3].

Having in mind, that we limit ourselves to the special case of the problem, we could categorize object tracking algorithms with respect to the object/s appearance, and a general approach used (omitting many other aspects). From this point of view, a naive and over-simplified mind map is depicted in Fig. 5.2. Algorithms, that are more deeply discussed in this thesis, are depicted in red.

As was already noted, the general MHT could be understood as a generalization of the algorithm presented in Chapter 3. For convenience, we define a *hypothesis-based approach* to put the NN and PDA based algorithms together with MHT. These are namely the *global* NN (GNN) [2], and *joint* PDA (JPDA) filters for the case of having known the number of objects, and the *joint integrated* PDA (JIPDA) filter [3] for the general scenario. It should be noted that there exists a variety of specific types of solutions to the MHT approach [1, 2, 20], which are outside the scope of this thesis.

For the RFS approach, we first discuss the algorithms that do not assume the targets to be i.i.d. An RFS-based counterpart to the "non-i.i.d." hypothesis-based approach could be called *multi Bernoulli approximation strategies* approach. Apart from the already discussed Bernoulli (or JoTT) filter, some of emerging solutions designed for the general scenario are the multi-target multi Bernoulli (MeMBr) filter [4, pp. 655-682], delta-generalized labeled multi Bernoulli (δ-GLMB) filter [48, 49], and to complement the survey [5], the Poisson multi Bernoulli mixture (PMBM) filter [50, 51] and filtering using sets of trajectories [52]. It should be noted, that these filters do not generally avoid data associations, and have a strong relation to the MHT, see e.g. [53].

We want to point out, that the "non-i.i.d." approaches try to model the moving objects comprehensively, trying to store information (such as a unique spatial probability distribu-

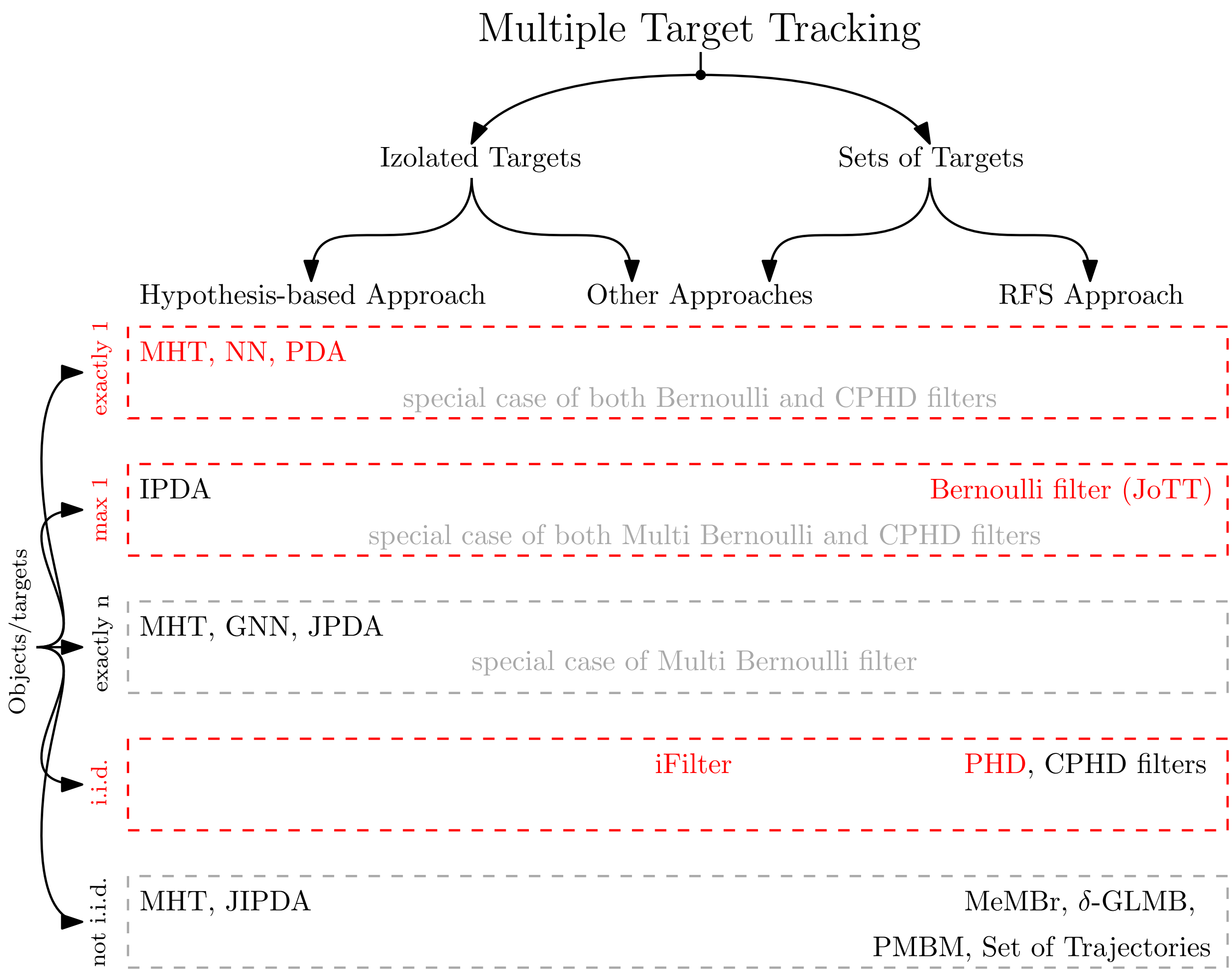


Fig. 5.2: Simplified mind map of (possibly multiple) target/object tracking.

tion) for each target rather separately. In case of the RFS approach, we could imagine it as filling a posterior multi-target PDF with a lot of data. This increases the accuracy, but also the computational cost, which is many times higher than that of single target tracking algorithms. In order to decrease the computational costs, drastic, but principled approximation strategies can be made, assuming the targets are i.i.d.

The term *other approaches* refers to the measure theoretic approach [46, 34, 26] unifying the point processes approach [47] or possible ad-hoc non-Bayesian solutions and others.

5.1.2 i.i.d. Approximation to the Objects

If the targets were i.i.d., each one of them would be spatially described by the same (probably multi-modal) spatial PDF. Using such approximation to the moving objects sounds computationally friendly, but there are some major problems (more deeply discussed later). The density of i.i.d. objects is not a conjugate prior with respect to a usual measurement likelihood, and hence the posterior is not from the same family of distributions as the prior. To ensure recursification, this brings us to introduce other approximations. Now, we discuss the RFS approach of Mahler and the point process approach of Streit.

Using the RFSs with the FISST theory, the general i.i.d. targets are described by the i.i.d. cluster process, which is similar to the Poisson RFS, but the cardinality distribution is arbitrary. Using such target description when performing the Bayes update, i.e. not necessarily during the whole recursion, the cardinalized probability hypothesis density (CPHD) filter was derived [4, pp. 632-653][54, 55]. Its special case[1] when the state-set cardinality is assumed to be Poisson[2] for the Bayes-update, is called the PHD filter [4, pp. 587-632][56, 38]. Both the PHD and CPHD filters share the same idea of derivation. Since the PHD filter has much simpler notation, we are only interested in the PHD filter in this thesis.

Using Poisson point processes (PPPs), another direction can be used to derive a filter very similar to the PHD filter, called the intensity filter (or iFilter since 2010) [23]. Both the PHD and intensity filters can be written such that they have the same resulting equations, but the derivation procedure is different. To continue investigating various approaches, the following section deals with the PPP derivation of the intensity filter, mainly inspired by [23]. The PHD filter is derived[3] in Appendix B using PGFLs, where various useful notes are given to understand the original derivation in [38] easier. A few alternatives to the derivation of the PHD (or intensity) filter exist [56, 38, 23, 57, 58], and a relationship with the MHT has been established also [59].

[1]Strictly speaking, the PHD filter is not a direct special case of the classical CPHD filter from [54], since this CPHD filter does not model target "spawning" within the motion model, but the PHD filter does.

[2]It can be shown [38, Theorem 4], that the Poisson approximation using only a PHD of an arbitrary RFS to describe the RFS, is the best in an informational sense (Kullback-Leibler divergence is minimal).

[3]Note that the PHD filter derivation using PGFLs can be understood as a fundamental approach for deriving more advanced filters, such as CPHD, and PMBM filters mentioned before. This fact establishes the reason for including this derivation into this thesis.

5.2 Derivation of the Intensity Filter

In this section, the derivation is inspired by [23]. The random state-set Ξ_k is treated as a PPP rather than a Poisson RFS (i.e. without the FISST), although the notation is kept similar for convenience. Unlike the original PHD filter derivation, the intensity filter derivation can be understood in "essentially elementary terms" [25, 27]. In order to tackle the clutter-generated measurements, the so-called *endogenous model of clutter* is introduced, defining the *clutter target*, or *target missing hypothesis* state[4]. Those clutter targets randomly "haunt" the measurement space $\mathcal{Z}$ with their clutter measurements. The augmented single-target state-space denoted with $\mathcal{X}^A$ is defined as a union of the single-target state-space $\mathcal{X}$ and the clutter target state-space ϕ, and so $\mathbf{x}_k \in \mathcal{X}^A \triangleq \phi \cup \mathcal{X}$. The clutter target state-space ϕ is defined to be a subset of $\mathbb{R}^{n_x}$ with nonzero volume, such that $\mathcal{X} \cap \mathcal{X}^A = \emptyset$, and so the augmented space $\mathcal{X}^A$ is a disconnected subset of $\mathbb{R}^{n_x}$ [25]. To separate the notation for functions[5] (and set functions such as integrals) defined on $\mathcal{X}$, ϕ and $\mathcal{X}^A$, superscripts will be used as follows.

- If a function is defined on $\mathcal{X}$, no superscript is used.
- If a function is defined on ϕ, superscript ϕ is used.
- If a function is defined on $\mathcal{X}^A$, superscript A, standing for *augmented* is used.

Now, the point process realizations $X_k = \{\mathbf{x}_k^1, \dots, \mathbf{x}_k^{n_k}\}$ can incorporate multiple "non-existing" distinct state-vectors $\mathbf{x}_k^i \in \phi$ for some $i \in \{1, \dots, n_k\}$. The posterior PDF of the point process Ξ_k modeling the moving objects, is

$$f_{k|k}(X_k|Z^k) = e^{-\overline{N}^A_{k|k}} \prod_{\mathbf{x}\in X_k} D^A_{k|k}(\mathbf{x}), \qquad \overline{N}^A_{k|k} = \int_{\mathcal{X}^A} D^A_{k|k}(\mathbf{x}) d\mathbf{x} \tag{5.1}$$

where $Z^k = (Z_1, \dots, Z_k)$ is the measurement-set sequence, $\overline{N}^A_{k|k}$ is the expected number of targets, and $D^A_{k|k}(\mathbf{x})$ is the intensity, or PHD. Moreover, denote $p^A_{k|k}(\mathbf{x})$ as the spatial PDF

[4]Note that in general, the introduction of clutter targets also enables to estimate target birth and measurement clutter process as a part of the algorithm. For more discussion, see [25].

[5]Note, that the notation for the point process Ξ_k, its realizations and individual vectors will not be explicitly distinguished.

of the targets such that

$$D^A_{k|k}(\mathbf{x}) = \overline{N}^A_{k|k} \cdot p^A_{k|k}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X}^A \tag{5.2}$$

Note that the augmented space brings a notion such that

$$\overline{N}^A_{k|k} = \int_{\mathcal{X}^A} D^A_{k|k}(\mathbf{x})d\mathbf{x} = \underbrace{\int_{\mathcal{X}} D_{k|k}(\mathbf{x})d\mathbf{x}}_{\overline{N}_{k|k}} + \underbrace{\int_{\phi} D^{\phi}_{k|k}(\mathbf{x})d\mathbf{x}}_{\overline{N}^{\phi}_{k|k}} \tag{5.3}$$

where $\overline{N}_{k|k}$ is expected number of true targets, and $\overline{N}^{\phi}_{k|k}$ is expected number of clutter targets. The goal is to find out how all these parameters propagate through the motion and measurement models.

5.2.1 Time-update

Define[6] the augmented Markov transition[7] PDF for $\mathbf{x}_k, \mathbf{x}_{k-1} \in \mathcal{X}^A$ as

$$\psi^A_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = \begin{cases} 1 - P_B, & \text{if } \mathbf{x}_k \in \phi, \ \& \ \mathbf{x}_{k-1} \in \phi \\ P_B \cdot p_B(\mathbf{x}_k), & \text{if } \mathbf{x}_k \in \mathcal{X}, \& \ \mathbf{x}_{k-1} \in \phi \\ 1 - P_S, & \text{if } \mathbf{x}_k \in \phi, \ \& \ \mathbf{x}_{k-1} \in \mathcal{X} \\ P_S \cdot \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}), & \text{if } \mathbf{x}_k \in \mathcal{X}, \& \ \mathbf{x}_{k-1} \in \mathcal{X} \end{cases} \tag{5.4}$$

where P_B is the probability of birth, $p_B(\mathbf{x}_k)$ is the birth spatial PDF, and P_S is the probability of survival. The term $\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ stands for the single-target Markov transition PDF (2.6). Note, that $p_B(\mathbf{x})$ does not necessarily have to be the same for each time instant, and we could generally define P_S to be state-dependent.

Assume each of the targets in $\Xi_{k-1} = \{\mathbf{x}^1_{k-1}, \dots, \mathbf{x}^{n_k}_{k-1}\}$ translates independently one by one with respect to (5.4), so the cardinality of realizations retains $|X_k| = |X_{k-1}|$. This implies that the random cardinality $N_k = |\Xi_k|$ retains Poisson with the expected number $\overline{N}^A_{k|k-1}$. Since we assumed Ξ_{k-1} to be i.i.d., Ξ_k is also i.i.d., and hence Ξ_k is also a PPP. Using this

[6] In [23], the Markov transition density was assumed general and not specified for the cases of $\mathbf{x}_k$ being from ϕ or not. Results provided in this thesis are therefore easier to use.

[7] Note that spawning of targets, contrary to the PHD filter, is not modeled within intensity filter.

idea, we can easily form the prior PDF from the product of transitioned spatial PDFs as

$$f_{k|k-1}(X_k|Z^k) = \overbrace{\frac{e^{-\overline{N}^A_{k-1|k-1}}\left(\overline{N}^A_{k-1|k-1}\right)^{n_k}}{n_k!}}^{\text{cardinality distribution}} \cdot \overbrace{n_k! \prod_{i=1}^{|X_k|} \int_{\mathcal{X}^A} \psi^A_{k|k-1}(\mathbf{x}^i_k|\mathbf{x}^i_{k-1}) p^A_{k-1|k-1}(\mathbf{x}^i_{k-1}) d\mathbf{x}^i_{k-1}}^{\text{spatial distribution conditioned on cardinality}} \tag{5.5a}$$

$$= e^{-\overline{N}^A_{k-1|k-1}} \underbrace{\int_{\mathcal{X}^A} \psi^A_{k|k-1}(\mathbf{x}^i_k|\mathbf{x}^i_{k-1}) \overbrace{\overline{N}^A_{k-1|k-1} p^A_{k-1|k-1}(\mathbf{x}^i_{k-1})}^{D^A_{k-1|k-1}(\mathbf{x}^i_{k-1})} d\mathbf{x}^i_{k-1}}_{D^A_{k|k-1}(\mathbf{x}^i_k)} \tag{5.5b}$$

$$= e^{-\overline{N}^A_{k|k-1}} \prod_{\mathbf{x} \in X_k} D^A_{k|k-1}(\mathbf{x}) \tag{5.5c}$$

Note that generally $\overline{N}_{k|k-1} \neq \overline{N}_{k-1|k-1}$, but rather

$$\overline{N}^A_{k|k-1} = \overline{N}_{k|k-1} + \overline{N}^\phi_{k|k-1} = \overline{N}_{k-1|k-1} + \overline{N}^\phi_{k-1|k-1} = \overline{N}^A_{k-1|k-1} \tag{5.6}$$

The posterior intensity over the augmented space is

$$D^A_{k|k-1}(\mathbf{x}_k) = \int_{\mathcal{X}^A} \psi^A_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) D^A_{k-1|k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1} \tag{5.7a}$$

$$= \begin{cases} (1-P_B)\overline{N}^\phi_{k-1|k-1} + (1-P_S)\overline{N}_{k-1|k-1}, & \text{if } \mathbf{x}_k \in \phi \\ P_B \cdot p_B(\mathbf{x}_k)\overline{N}^\phi_{k-1|k-1} + & \\ \qquad + P_S \cdot \int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) D_{k-1|k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}, & \text{if } \mathbf{x}_k \in \mathcal{X} \end{cases} \tag{5.7b}$$

This equation can be understood as the final time-update equation for the intensity filter. The original derivation from [23] stops at equation similar to (5.5c), hence this is a special case when the Markov transition PDF is of the form (5.4).

Results For $\mathbf{x}_k \in \mathcal{X}$

We are mainly interested in predicting true targets, rather than the clutter targets. For convenience, assume $\overline{N}^\phi_{k-1|k-1}$ is known at time k and denote $N_B = P_B \cdot \overline{N}^\phi_{k-1|k-1}$, which is obviously the expected number of target births. Then for $\mathbf{x}_k \in \mathcal{X}$ we get

$$D_{k|k-1}(\mathbf{x}_k) = N_B \cdot p_B(\mathbf{x}_k) + P_S \cdot \int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) D_{k-1|k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1} \tag{5.8}$$

which can be expressed using the prior expected number of targets and their spatial PDF,

$$\overline{N}_{k|k-1} = N_B + P_S \cdot \overline{N}_{k-1|k-1} \tag{5.9}$$

$$p_{k|k-1}(\mathbf{x}_k) = \frac{N_B}{\overline{N}_{k|k-1}} \cdot p_B(\mathbf{x}_k) + \frac{P_S \overline{N}_{k-1|k-1}}{\overline{N}_{k|k-1}} \cdot \int_{\mathcal{X}} \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) p_{k-1|k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1} \tag{5.10}$$

We can observe, that Eq. (5.10) is very similar to Eq. (4.8b) of the Bernoulli filter, only the coefficients scaling the birth PDF and the transitioned PDF differ. Instead of dividing by the prior probability of existence, we have the prior expected number of targets. Instead of multiplying the birth PDF with the probability of birth and probability of non-existence, we multiply by the expected number of target births. Finally, since the probability of existence is not computed, the second part is multiplied by the probability of survival and the posterior expected number of targets. To conclude, both filters have similar time-update schemes, although each of them tackles a different problem.

Initialization of Targets

Define the *birth process* to be a PPP with intensity $N^B_{k|k-1} \cdot p_B(\mathbf{x}_k)$. One can observe there is a superposition (see Section 2.2.10) in Eq. (5.8), of the birth process and a PPP describing the survived targets.

An initial spatial PDF $p_{0|0}(\mathbf{x}_0)$, together with expected number of targets $\overline{N}_{0|0}$, can be given at the initial time step, but it is not necessary since a birth process is modeled. Moreover, at any time step, if an initial PDF and probability of existence for any newly born target is known, it[8] can be easily incorporated into the birth spatial PDF $p_B(\mathbf{x}_k)$ yielding a mixture PDF. Note that when initializing targets by adding their initial prior knowledge into $p_B(\mathbf{x}_k)$, the effect is the same as adding their initial posterior knowledge straightly into the posterior PPP at $k-1$, but the target is not even present at that time step. It should be noted, that in MTT filters, which model the birth processes, the targets are assumed to be initialized using prior knowledge in the prediction step. This was already seen in the Bernoulli filter in Chapter 4. In theory, this is obviously different from the single-target case, that is usually assumed to start from a known posterior initial PDF at time $k=0$. As noted in Chapter 4, other methods of track formation exist, see e.g. [3, pp. 322-329][1, pp.153 168].

[8] After its single-target prediction.

5.2.2 Bayes-update

The Bayes-update is based on the idea of splitting the state-set to two subsets, where one consists of undetected targets, and the other one of the detected ones. Those state-sets are constructed using the Bernoulli thinning (see Section 2.2.10) with an augmented probability of detection $P_D^A(\mathbf{x}_k)$. Both sets are therefore PPPs with some intensity. The intensity of undetected targets is left unchanged, and the other one is Bayes-updated. As has been said before, the Bayes rule does not result into a posterior being a PPP under the usual measurement likelihood and the prior being a PPP. Therefore, an approximation is made to get the Bayes-updated part to be a PPP. In the end, both PPPs are superposed to yield the final posterior. Using the notation yet to be defined, the derivation procedure is depicted in Fig. 5.3.

Bernoulli thinning

$$\overset{\text{input}}{D^A_{k|k-1}(\mathbf{x}_k)} = \underbrace{\overset{\text{Undetected}}{\big(1 - P_D^A(\mathbf{x}_k)\big) D^A_{k|k-1}(\mathbf{x}_k)}} + \underbrace{\overset{\text{Detected}}{P_D^A(\mathbf{x}_k) D^A_{k|k-1}(\mathbf{x}_k)}}$$

PPP approximation of Bayes update with Z_k

$$\underset{\text{output}}{D^A_{k|k}(\mathbf{x}_k)} = \underbrace{\overbrace{\big(1 - P_D^A(\mathbf{x}_k)\big) D^A_{k|k-1}(\mathbf{x}_k)} + \overbrace{\sum_{\mathbf{z} \in Z_k} \frac{P_D^A(\mathbf{x}_k) l^A(\mathbf{z}|\mathbf{x}_k)}{D_m(\mathbf{z})} D^A_{k|k-1}(\mathbf{x}_k)}}_{\text{Superposition}}$$

Fig. 5.3: Illustration of Bayes-update procedure, $\mathbf{x}_k \in \mathcal{X}^A$.

Splitting the Prior

Note, that Bernoulli thinning is an operation performed for the whole point process, rather then for some specific outcome. As a result, when performed with parameter $P_D^A(\mathbf{x}_k)$ it yields two PPPs with intensities $\big(1 - P_D^A(\mathbf{x}_k)\big) D^A_{k|k-1}(\mathbf{x}_k)$ for the undetected process, and $D^{DA}_{k|k-1}(\mathbf{x}_k) \triangleq P_D^A(\mathbf{x}_k) D^A_{k|k-1}(\mathbf{x}_k)$ for the detected process. For the undetected process, assuming there is no measurement available to use for the update, its intensity is retained in the posterior.

The outcomes of the thinned processes, being just a part of the final posterior, should be denoted with some other symbol. However, the individual vectors have the same meaning.

Since we do not need to work with the undetected targets explicitly, the terms Ξ_k^D, $N_k^D = |\Xi_k^D|$ and $X_k^D = \{\mathbf{x}_k^1, \ldots, \mathbf{x}_k^{n_k}\}$ can be used to denote the detected PPP, its cardinality and its outcomes, respectively. Moreover, its PDF will be denoted by $f_{k|k-1}^D(X_k^D|Z^{k-1})$ and its expected number of targets and spatial PDF are

$$\overline{N}_{k|k-1}^{DA} = \int_{\mathcal{X}^A} D_{k|k-1}^{DA}(\mathbf{x}_k) d\mathbf{x}_k = \overline{N}_{k|k-1}^{A} \int_{\mathcal{X}^A} P_D^A(\mathbf{x}_k) p_{k|k-1}^A(\mathbf{x}_k) d\mathbf{x}_k \tag{5.11}$$

$$p_{k|k-1}^{DA}(\mathbf{x}_k) = \frac{1}{\overline{N}_{k|k-1}^{DA}} D_{k|k-1}^{DA}(\mathbf{x}_k) = \frac{P_D^A(\mathbf{x}_k) p_{k|k-1}^A(\mathbf{x}_k)}{\int_{\mathcal{X}^A} P_D^A(\mathbf{x}_k) p_{k|k-1}^A(\mathbf{x}_k) d\mathbf{x}_k} \tag{5.12}$$

Complete Update of the Detected Process

Similarly to the single-target tracking case, the measurements $Z_k = \{\mathbf{z}_k^1, \ldots, \mathbf{z}_k^{m_k}\}$ can be obtained due to true targets, or due to clutter. Since the clutter targets are modeled, every single measurement from Z_k is assumed to be generated from one of the detected targets. Therefore, when Z_k arrives, the number of detected targets is known to be m_k.

The first step is finding the posterior PDF of the detected targets. Define[9] the single-target augmented measurement likelihood function for $\mathbf{x}_k \in \mathcal{X}^A$ as

$$l^A(\mathbf{z}|\mathbf{x}_k) = \begin{cases} c(\mathbf{z}), & \text{if } \mathbf{x}_k \in \phi, \\ l(\mathbf{z}|\mathbf{x}_k), & \text{if } \mathbf{x}_k \in \mathcal{X}, \end{cases} \tag{5.13}$$

where $c(\mathbf{z})$ is the clutter spatial PDF and $l(\mathbf{z}|\mathbf{x}_k)$ is the single-target measurement likelihood function (2.7). Since every single measurement in $Z_k = \{\mathbf{z}_k^1, \ldots, \mathbf{z}_k^{m_k}\}$ was generated by a target in X_k^D, the PDF of the measurement process $\Sigma_k|\Sigma^{k-1}$, can be derived using analogous ideas as in derivation of the prior PDF. It is a PPP with the PDF

$$f_{\Sigma_k|\Sigma^{k-1}}(Z_k|Z^{k-1}) = \overbrace{\frac{e^{-\overline{N}_{k|k-1}^{DA}} \left(\overline{N}_{k|k-1}^{DA}\right)^{n_k}}{n_k!}}^{\text{card. distribution}} \cdot \overbrace{n_k! \prod_{i=1}^{n_k} \int_{\mathcal{X}^A} l^A(\mathbf{z}_k^i|\mathbf{x}_k^i) \frac{P_D^A(\mathbf{x}_k) p_{k|k-1}^A(\mathbf{x}_k)}{\int_{\mathcal{X}^A} P_D^A(\mathbf{x}_k) p_{k|k-1}^A(\mathbf{x}_k) d\mathbf{x}_k} d\mathbf{x}_k^i}^{\text{spatial distribution conditioned on cardinality}} \tag{5.14a}$$

$$= e^{-\overline{N}_{k|k-1}^{DA}} \prod_{i=1}^{m_k} \underbrace{\int_{\mathcal{X}^A} l^A(\mathbf{z}_k^i|\mathbf{x}_k^i) \overbrace{\overline{N}_{k|k-1}^{A} P_D^A(\mathbf{x}_k^i) p_{k|k-1}^A(\mathbf{x}_k^i)}^{D_{k|k-1}^{DA}(\mathbf{x}_k^i)} d\mathbf{x}_k^i}_{D_m(\mathbf{z}_k^i) \triangleq} \tag{5.14b}$$

$$= e^{-\overline{N}_{k|k-1}^{DA}} \prod_{\mathbf{z} \in Z_k} D_m(\mathbf{z}) \tag{5.14c}$$

[9]Similarly to the time-update, in [23] the augmented measurement likelihood was not specified as herein.

where the measurement intensity is

$$D_m(\mathbf{z}_k) = \int_{\mathcal{X}^A} l^A(\mathbf{z}_k|\mathbf{x}_k) P_D^A(\mathbf{x}_k) D_{k|k-1}^A(\mathbf{x}_k) d\mathbf{x}_k \tag{5.15a}$$

$$= c(\mathbf{z}_k)\overline{N}_{k|k-1}^{\phi} + \int_{\mathcal{X}} l(\mathbf{z}_k|\mathbf{x}_k) P_D(\mathbf{x}_k) D_{k|k-1}(\mathbf{x}_k) d\mathbf{x}_k \tag{5.15b}$$

Note, that the (prior) expected number of clutter targets/measurements $\overline{N}_{k|k-1}^{\phi}$, which is an estimate, is usually denoted λ and assumed known.

Now, the measurement likelihood will be derived. Since we do not know, which target generated which measurement, the likelihood is given by

$$l^A(Z_k|X_k^D) = \begin{cases} \sum_{\sigma\in\mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i), & \text{if } m_k = n_k \\ 0, & \text{otherwise} \end{cases} \tag{5.16}$$

where $\mathrm{Sym}(m_k)$ represents the set of all permutations of integers up to m_k. Notation for l^A is abused again, and should be obvious from the context. The Bayes updated PDF is then nonzero only if $m_k = n_k$ (i.e. if $|Z_k| = |X_k^D|$), and for such case it is

$$f_{k|k}^D(X_k^D|Z^k) = \frac{l^A(Z_k|X_k^D) f_{k|k-1}^D(X_k^D|Z^{k-1})}{f_{\Sigma_k|\Sigma^k}(Z_k|Z^k)} \tag{5.17a}$$

$$= \frac{\left(\sum_{\sigma\in\mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i)\right)\left(\cancel{e^{-\overline{N}_{k|k-1}^{DA}}} \prod_{\mathbf{x}\in X_k^D} P_D^A(\mathbf{x}_k) D_{k|k-1}^A(\mathbf{x})\right)}{\cancel{e^{-\overline{N}_{k|k-1}^{DA}}} \prod_{\mathbf{z}\in Z_k} D_m(\mathbf{z})} \tag{5.17b}$$

$$= \sum_{\sigma\in\mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} \frac{l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i) P_D^A(\mathbf{x}_k^i) D_{k|k-1}^A(\mathbf{x}_k^i)}{D_m(\mathbf{z}_k^{\sigma(i)})} \tag{5.17c}$$

$$= \sum_{\sigma\in\mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} \frac{l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i) D_{k|k-1}^{DA}(\mathbf{x}_k^i)}{D_m(\mathbf{z}_k^{\sigma(i)})} \tag{5.17d}$$

where the equality between (5.17b) and (5.17d) holds because each of the products can be expressed using the same index. The PDF in (5.17d) is obviously not a PDF of some PPP.

From this point, there are several ways to continue with the derivation. We consider only the following two possibilities, that were highlighted in [23, 58]. First, we could integrate the PDF (5.17d) over all, except one state variable to find a single-variate function and then find the expected number of targets. The found values would be used to form an approximating PPP of the updated point process [23, pp. 239-243]. This procedure can be made more rigorous using the newly proven result for PHDs from Section 2.2.4, see Appendix A.5.1.

Second, we could establish a connection with the positron emission tomography (PET), noting that the desired intensity filter update is formally equivalent to a single step of the Shepp-Vardi algorithm with the time-of-flight data (when made continuous as a chosen volume partition approaches zero), originally derived using the expectation-maximization (EM) method. At the expense of possible convergence problems of a continuous version of the EM method [60, 61] we argue that this can be made more rigorous avoiding the possibly questionable limiting process of the original derivation in [23, pp. 112-118][58].

Both possibilities provide some valuable insight into the problem. In both of the derivations, the PDF (5.17d) is integrated over all, except one state, but the circumstances differ. We present both possibilities in this thesis, contributing with the mentioned improvements.

Straightforward Derivation of the Posterior Intensity Function

For simplicity, explicit conditioning on the measurement process Σ^k is omitted and so Ξ_k^D denotes the posterior point process. Remind that the point processes discussed in this thesis are actually RFSs and they can be equivalently modeled with FISST. The PHD defined in Section 2.2 with the FISST set integral is therefore defined for the updated point process with the PDF (5.17d) as well. The idea is to find the PHD of this point process and parameterize a PPP with it. In terms of FISST, this induces an approximation which can be shown to be the best in an informational sense, see [38, Theorem 4]. In Section 2.2.4, the PHD was said to be equal to a product of the expected number of targets in the whole single-target state-space and a function that is a normalized sum of marginal PDFs of ordered $\Xi_k^D|N_k^D$ being a singleton set. Since the resulting PHD is used as an intensity of an approximating PPP, the function becomes the spatial distribution parameter of a PPP and hence can be denoted $p_{k|k}^D(\mathbf{x}_k)$, for more detail see Appendix A.5.1. Since the PDF before the Poisson approximation is known to be (5.17d), we are able to form the PHD using the aforementioned new result.

Since the PDF (5.17d) is nonzero only if $m_k = n_k$, its cardinality is given by the Kronecker delta function centered at m_k, $f_{N_k^D}(n_k) = \delta_{m_k}^{\mathrm{Kron}}(n_k)$ and so is actually nonrandom. This simplifies the computation significantly.

- The expected value of the cardinality distribution, which is the desired expected number of targets in the whole single-target state-space is m_k.

- The normalized sum of marginal PDFs of ordered $\Xi_k^D|N_k^D$ being singleton (see Appendix A.5.1), denoted[10] with $p_{k|k}^D(\mathbf{x}_k^l)$ has a single summand and is given by

$$
\begin{aligned}
p_{k|k}^D(\mathbf{x}_k^l) &= \frac{\sum_{n=1}^{+\infty} n \cdot f_{N_k}(n) \cdot f_{\mathbf{X}_k^D|N_k^D}(\mathbf{x}_k^l|n)}{\sum_{n=1}^{+\infty} n \cdot f_{N_k^D}(n)} = \frac{\cancel{m_k} \cdot f_{\mathbf{X}_k^D|N_k^D}(\mathbf{x}_k^l|m_k)}{\cancel{m_k}} && \text{(5.18a)}\\
&= \int_{(\mathcal{X}^A)^{m_k-1}} f_{\mathbf{X}_k^D|N_k^D}(\mathbf{x}_k^1,\dots,\mathbf{x}_k^l,\dots,\mathbf{x}_k^{m_k}|m_k)d\mathbf{x}_k^1 \cdots d\mathbf{x}_k^{l-1} d\mathbf{x}_k^{l+1}\cdots d\mathbf{x}_k^{m_k} && \text{(5.18b)}\\
&= \frac{1}{m_k!}\int_{(\mathcal{X}^A)^{m_k-1}} \underbrace{f_{\Xi_k^D|N_k^D}(\{\mathbf{x}_k^1,\dots,\mathbf{x}_k^l,\dots,\mathbf{x}_k^{m_k}\}|m_k)}_{f_{k|k}^D(\{\mathbf{x}_k^1,\dots,\mathbf{x}_k^l,\dots,\mathbf{x}_k^{m_k}\}|Z^k)} \prod_{\substack{j=1\\ j\neq l}}^{m_k} d\mathbf{x}^j && \text{(5.18c)}\\
&= \frac{1}{m_k!}\int_{(\mathcal{X}^A)^{m_k-1}} \sum_{\sigma\in\mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} \frac{l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i)D_{k|k-1}^{DA}(\mathbf{x}_k^i)}{D_m(\mathbf{z}_k^{\sigma(i)})} \prod_{\substack{j=1\\ j\neq l}}^{m_k} d\mathbf{x}^j && \text{(5.18d)}\\
&= \frac{1}{m_k!} \underbrace{\sum_{\sigma\in\mathrm{Sym}(m_k)} \frac{l^A(\mathbf{z}_k^{\sigma(l)}|\mathbf{x}_k^l)D_{k|k-1}^{DA}(\mathbf{x}_k^l)}{D_m(\mathbf{z}_k^{\sigma(l)})}}_{\substack{\text{this expression has } m_k! \text{ summands, which can be}\\ \text{rearranged into } (m_k-1)! \text{ sums over meaurements}}} \underbrace{\prod_{\substack{i=1\\ i\neq l}}^{m_k} \frac{\overbrace{\int_{\mathcal{X}^A} l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i)D_{k|k-1}^{DA}(\mathbf{x}_k^i)d\mathbf{x}_k^i}^{\cancel{D_m(\mathbf{z}_k^{\sigma(i)})}}}{\cancel{D_m(\mathbf{z}_k^{\sigma(i)})}}}_{=1} && \text{(5.18e)}\\
&= \frac{(m_k-1)!}{m_k!}\sum_{i=1}^{m_k} \frac{l^A(\mathbf{z}_k^i|\mathbf{x}_k^l)D_{k|k-1}^{DA}(\mathbf{x}_k^l)}{D_m(\mathbf{z}_k^i)} = \frac{1}{m_k}\sum_{i=1}^{m_k} \frac{l^A(\mathbf{z}_k^i|\mathbf{x}_k^l)P_D^A(\mathbf{x}_k^l)D_{k|k-1}^{A}(\mathbf{x}_k^l)}{D_m(\mathbf{z}_k^i)} && \text{(5.18f)}
\end{aligned}
$$

The desired PHD is then given by

$$
D_{k|k}^{DA} = m_k \cdot p_{k|k}^D(\mathbf{x}_k) = \sum_{i=1}^{m_k} \frac{l^A(\mathbf{z}_k^i|\mathbf{x}_k^l)P_D^A(\mathbf{x}_k^l)D_{k|k-1}^{A}(\mathbf{x}_k^l)}{D_m(\mathbf{z}_k^i)} \tag{5.19}
$$

Taking this function to be the intensity of the approximating PPP finishes the derivation.

Likelihood Maximization Using EM Method

This part of the intensity filter update can be shown to be a single maximization step of the EM algorithm, designed for a specific continuous problem to maximize the measurement likelihood. The derivation in [23, pp. 112-118][58] focused on the positron emission tomography (PET) with the time-of-flight data leading to the Shepp-Vardi algorithm. One step of this algorithm, when "made continuous", was shown to be equivalent to the desired ap-

[10]Note, that introduction of the index l is just for simplicity when computing the integrals.

proximate Bayes-update of detected process in the intensity filter. However, in the original derivation a spatially discrete problem was solved, and the results were made continuous taking a limit as a chosen volume partition approaches zero. We argue, that there is no reason for breaking the single-target state-space into discrete pieces of volume to derive the desired update. Therefore, we propose a derivation verifying that this step in the intensity filter is one step of a continuous version of the imaging algorithm, which however generally does not converge.

Note, that the EM method was originally defined for estimating discrete variables [62]. A general EM method for a continuous problem like this exists [63], but it can be ill-posed and thus does not generally converge [60, 61]. However, we are interested only in a single M-step. The version of the EM method used in the following can be found in [23, pp. 223-228].

Let the estimated parameter[11] θ, be an intensity function $\theta(\mathbf{x}_k) \in \Theta$, where Θ is a set of all valid intensity functions defined on $\mathcal{X}^A$. Assume $\theta^{k-1}(\mathbf{x}_k)$ is the intensity of the detected process $D^{DA}_{k|k-1}(\mathbf{x}_k)$. Then $\theta^k(\mathbf{x}_k)$ will be taken to be the intensity after the desired approximate Bayes-update, denoted $D^{DA}_{k|k}(\mathbf{x}_k)$. Assuming Z_k arrived, the objective is to find $\theta^k(\mathbf{x}_k)$ maximizing the likelihood defined as

$$p(Z_k|\theta^k) \triangleq m_k! \prod_{\mathbf{z}\in Z_k} \int_{\mathcal{X}^A} l^A(\mathbf{z}|\mathbf{x}_k)\theta^k(\mathbf{x}_k)d\mathbf{x}_k \tag{5.20}$$

Let the detected state-set $X^D_k = \{\mathbf{x}^1_k, \ldots, \mathbf{x}^{m_k}_k\}$ be the set of missing data. The E-step (just a formal part) is then given by evaluating the auxiliary function, which can be defined as the following expectation over the general posterior,

$$\begin{aligned} Q(\theta, \theta^{k-1}) &= \mathrm{E}_{\mathbf{X}^D_k|\Sigma_k,\theta^{k-1}}[\log f(X^D_k, Z_k|\theta)] && (5.21\text{a}) \\ &= \int_{(\mathcal{X}^A)^{m_k}} \left(\log f(X^D_k, Z_k|\theta)\right) \frac{f^D_{k|k}(X^D_k|Z_k, \theta^{k-1})}{m_k!} d\mathbf{x}^1_k \ldots d\mathbf{x}^{m_k}_k && (5.21\text{b}) \end{aligned}$$

where the term $\frac{1}{m_k!}$ could be interpreted as taking an ordered tuple $\mathbf{X}^D_k$ of Ξ^D_k for simplicity[12],

[11] Note that θ in this section, should not be confused with a data association hypothesis. We use θ to respect the standard notation in EM methods.

[12] Since m_k is known, the expatiation in (5.21b) can be understood as expectation for usual joint PDFs.

or ordering a differential $d\{\mathbf{x}_k^1, \dots, \mathbf{x}_k^{m_k}\}$. The joint PDF of $\Xi_k^D, \Sigma_k|\theta$ is

$$f(X_k^D, Z_k|\theta) \triangleq \left[l(Z_k|X_k^D) f_{k|k-1}^D(X_k^D) \right]_{D_{k|k-1}^A = \theta} \tag{5.22a}$$

$$= \left(\sum_{\sigma \in \mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i) \right) \left(e^{-\int_{\mathcal{X}^A} \theta(\mathbf{x}) d\mathbf{x}} \prod_{j=1}^{m_k} \theta(\mathbf{x}_k^j) \right) \tag{5.22b}$$

The M-step is given by maximizing $Q(\theta, \theta^{k-1})$ with $\theta \in \Theta$,

$$\theta^k(\mathbf{x}) := \arg\max_{\theta \in \Theta} Q(\theta, \theta^{k-1}) \tag{5.23}$$

To continue, the logarithm of (5.22b) is given by

$$\log f(X_k^D, Z_k|\theta) = \cancel{\log \left(\sum_{\sigma \in \mathrm{Sym}(m_k)} \prod_{i=1}^{m_k} l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i) \right)} - \int_{\mathcal{X}^A} \theta(\mathbf{x}) d\mathbf{x} + \sum_{j=1}^{m_k} \log \theta(\mathbf{x}_k^j) \tag{5.24}$$

where the first part can be neglected for the minimization, since it is independent of the optimal argument θ. Substituting (5.24) and (5.17d given θ^{k-1}) into $Q(\theta, \theta^{k-1})$, and noting that $D_m(\mathbf{z}) = \int_{\mathcal{X}^A} l^A(\mathbf{z}|\mathbf{x}_k^j) \theta^{k-1}(\mathbf{x}_k^j) d\mathbf{x}_k^j$, gives

$$Q(\theta, \theta^{k-1}) =$$

$$= \int_{(\mathcal{X}^A)^{m_k}} \left(- \int_{\mathcal{X}^A} \theta(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^{m_k} \log \theta(\mathbf{x}_k^i) \right) \frac{1}{m_k!} \sum_{\sigma \in \mathrm{Sym}(m_k)} \prod_{j=1}^{m_k} \frac{l^A(\mathbf{z}_k^{\sigma(j)}|\mathbf{x}_k^j) \theta^{k-1}(\mathbf{x}_k^j)}{D_m(\mathbf{z}_k^{\sigma(j)})} d\mathbf{x}_k^1 \dots d\mathbf{x}_k^{m_k} \tag{5.25a}$$

$$= - \int_{\mathcal{X}^A} \theta(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^{m_k} \int_{(\mathcal{X}^A)^{m_k}} \log \theta(\mathbf{x}_k^i) \frac{1}{m_k!} \sum_{\sigma \in \mathrm{Sym}(m_k)} \prod_{j=1}^{m_k} \frac{l^A(\mathbf{z}_k^{\sigma(j)}|\mathbf{x}_k^j) \theta^{k-1}(\mathbf{x}_k^j)}{D_m(\mathbf{z}_k^{\sigma(j)})} d\mathbf{x}_k^j \tag{5.25b}$$

$$= - \int_{\mathcal{X}^A} \theta(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^{m_k} \int_{\mathcal{X}^A} \log \theta(\mathbf{x}_k^i) \frac{1}{m_k!} \sum_{\sigma \in \mathrm{Sym}(m_k)} \prod_{\substack{j=1 \\ j \neq i}}^{m_k} \frac{l^A(\mathbf{z}_k^{\sigma(i)}|\mathbf{x}_k^i) \theta^{k-1}(\mathbf{x}_k^i) \cancel{D_m(\mathbf{z}_k^{\sigma(j)})}}{D_m(\mathbf{z}_k^{\sigma(i)}) \cancel{D_m(\mathbf{z}_k^{\sigma(j)})}} d\mathbf{x}_k^i \tag{5.25c}$$

$$= - \int_{\mathcal{X}^A} \theta(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^{m_k} \int_{\mathcal{X}^A} \log \theta(\mathbf{x}_k^i) \frac{l^A(\mathbf{z}_k^i|\mathbf{x}_k^i) \theta^{k-1}(\mathbf{x}_k^i)}{D_m(\mathbf{z}_k^i)} d\mathbf{x}_k^i \tag{5.25d}$$

To minimize $Q(\theta, \theta^{k-1})$ with respect to θ, we could take a functional derivative (assuming θ^{k-1} is a parameter) and set it to zero everywhere in $\mathcal{X}^A$, see e.g. [64] to get an intuition. Note, that since the free variable is from the augmented space $\mathcal{X}^A$, this step might be questionable. Assuming $\mathcal{X}^A$ is defined such that it is possible, we get

$$\frac{\delta Q(\theta,\theta^{k-1})}{\delta \mathbf{x}} = -1 + \frac{1}{\theta(\mathbf{x})}\sum_{i=1}^{m_k}\frac{l^A(\mathbf{z}_k^i|\mathbf{x})\theta^{k-1}(\mathbf{x})}{D_m(\mathbf{z}_k^i)} \overset{!}{=} 0 \qquad \forall \mathbf{x}\in\mathcal{X}^A \tag{5.26}$$

$$\frac{\delta}{\delta \mathbf{x}}\left(\frac{\delta Q(\theta,\theta^{k-1})}{\delta \mathbf{x}}\right) = -\frac{1}{\big(\theta(\mathbf{x})\big)^2}\sum_{i=1}^{m_k}\frac{l^A(\mathbf{z}_k^i|\mathbf{x})\theta^{k-1}(\mathbf{x})}{D_m(\mathbf{z}_k^i)} \leq 0 \qquad \forall \mathbf{x}\in\mathcal{X}^A,\ \forall\theta\in\Theta \tag{5.27}$$

And therefore we get

$$\theta^k(\mathbf{x}) = \sum_{i=1}^{m_k}\frac{l^A(\mathbf{z}_k^i|\mathbf{x})\theta^{k-1}(\mathbf{x})}{D_m(\mathbf{z}_k^i)} \tag{5.28}$$

which is due to (5.27) the desired argument of maxima. Remind that θ stands for the intensity function. The desired update is then given by

$$D_{k|k}^{DA}(\mathbf{x}_k) = \sum_{i=1}^{m_k}\frac{l^A(\mathbf{z}_k^i|\mathbf{x}_k)D_{k|k-1}^{DA}(\mathbf{x}_k)}{D_m(\mathbf{z}_k^i)} = \sum_{\mathbf{z}\in Z_k}\frac{l^A(\mathbf{z}|\mathbf{x}_k)P_D^A(\mathbf{x}_k)}{D_m(\mathbf{z})}D_{k|k-1}^A(\mathbf{x}_k) \tag{5.29}$$

Superposing the Updated Processes

The final updated intensity is given by a superposition of detected and undetected processes (both PPPs now), which yields the final intensity

$$D_{k|k}^A(\mathbf{x}_k) = \left(1-P_D^A(\mathbf{x}_k)\right)D_{k|k-1}^A(\mathbf{x}_k) + \sum_{\mathbf{z}\in Z_k}\frac{l^A(\mathbf{z}|\mathbf{x}_k)P_D^A(\mathbf{x}_k)}{D_m(\mathbf{z})}D_{k|k-1}^A(\mathbf{x}_k) \tag{5.30a}$$

$$= \left(1-P_D^A(\mathbf{x}_k) + \sum_{\mathbf{z}\in Z_k}\frac{l^A(\mathbf{z}|\mathbf{x}_k)P_D^A(\mathbf{x}_k)}{D_m(\mathbf{z})}\right)D_{k|k-1}^A(\mathbf{x}_k) \tag{5.30b}$$

$$= \begin{cases}\left(1-P_D^\phi(\mathbf{x}_k)+\sum_{\mathbf{z}\in Z_k}\frac{c(\mathbf{z})P_D^\phi(\mathbf{x}_k)}{D_m(\mathbf{z})}\right)D_{k|k-1}^\phi(\mathbf{x}_k), & \text{if } \mathbf{x}_k\in\phi,\\ \left(1-P_D(\mathbf{x}_k)+\sum_{\mathbf{z}\in Z_k}\frac{l(\mathbf{z}|\mathbf{x}_k)P_D(\mathbf{x}_k)}{D_m(\mathbf{z})}\right)D_{k|k-1}(\mathbf{x}_k), & \text{if } \mathbf{x}_k\in\mathcal{X}\end{cases} \tag{5.30c}$$

We can see that it is generally possible to estimate the intensity of clutter jointly with the posterior intensity of true targets. However, the probability of detection $P_D^\phi(\phi)$ of the clutter targets must be given. Vaguely speaking, how could we model a value, interpreted as a probability of detecting a non-existing target? Therefore, we will omit further discussion about the case when $\mathbf{x}_k\in\phi$, and understand clutter targets as a useful theoretical construct for the derivation of the filter.

Results For $\mathbf{x}_k \in \mathcal{X}$

Denote the expected number of clutter measurements $\lambda = \overline{N}^{\phi}_{k|k-1}$ (it appears in $D_m(\mathbf{z})$) and assume it is known. Then, the final intensity can be written as

$$D_{k|k}(\mathbf{x}_k) = \left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z}\in Z_k} \frac{l(\mathbf{z}|\mathbf{x}_k)P_D(\mathbf{x}_k)}{\lambda c(\mathbf{z}) + \int_{\mathcal{X}} l(\mathbf{z}|\mathbf{x}_k)P_D(\mathbf{x}_k)D_{k|k-1}(\mathbf{x}_k)d\mathbf{x}_k}\right) D_{k|k-1}(\mathbf{x}_k) \tag{5.31}$$

The integral in (5.31) is (in case of the PHD filter) usually denoted by

$$D_{k|k-1}[l(\mathbf{z}|\cdot)P_D] = \int_{\mathcal{X}} l(\mathbf{z}|\mathbf{x}_k)P_D(\mathbf{x}_k)D_{k|k-1}(\mathbf{x}_k)d\mathbf{x}_k \tag{5.32}$$

For the expected number of targets and the spatial PDF we can generally get nothing but

$$\overline{N}_{k|k} = \int_{\mathcal{X}} D_{k|k}(\mathbf{x}_k)d\mathbf{x}_k, \qquad p_{k|k}(\mathbf{x}_k) = \frac{1}{\overline{N}_{k|k}} D_{k|k}(\mathbf{x}_k) \tag{5.33}$$

but in case of constant probability of detection P_D, it becomes

$$\overline{N}_{k|k} = (1 - P_D)\overline{N}_{k|k-1} + \sum_{\mathbf{z}\in Z_k} \frac{P_D\overline{N}_{k|k-1}p_{\mathbf{z}_k}(\mathbf{z})}{\lambda c(\mathbf{z}) + P_D\overline{N}_{k|k-1}p_{\mathbf{z}_k}(\mathbf{z})} \tag{5.34}$$

$$p_{k|k}(\mathbf{x}_k) = \frac{\overline{N}_{k|k-1}}{\overline{N}_{k|k}} \left(1 - P_D + \sum_{\mathbf{z}\in Z_k} \frac{l(\mathbf{z}|\mathbf{x}_k)P_D}{\lambda c(\mathbf{z}) + P_D\overline{N}_{k|k-1}p_{\mathbf{z}_k}(\mathbf{z})}\right) p_{k|k-1}(\mathbf{x}_k) \tag{5.35}$$

where

$$p_{\mathbf{z}_k}(\mathbf{z}) \triangleq \int_{\mathcal{X}} l(\mathbf{z}|\mathbf{x}_k)p_{k|k-1}(\mathbf{x}_k)d\mathbf{x}_k \tag{5.36}$$

is the single-target measurement prediction, which is the same for all the targets.

For the single-target case, the Bayes-update of the spatial PDFs was formally equivalent for both MHT and Bernoulli filters. For the Poisson approximation to the multi-target case, we observe that the main difference lies in the additive weighting of each measurement.

5.3 Gaussian Sum Filter Implementation

The crucial update equations (5.8,5.31) were derived in this thesis such that they are formally the same for both the intensity and PHD filter (without target spawning and constant probability of survival, see Appendix B). Thus the following hold for both filters. The GSF implementation of the general PHD filter (i.e. with spawning and state-dependent probability of survival) is commonly known as the Gaussian Mixture PHD filter [65].

Remind that the targets are assumed to be independent, and their number is unknown and varies in time. If any target is present, its single-target transition PDF and measurement likelihood are assumed to be such as in the linear-Gaussian case (2.11,2.12). Assume constant probability of detection $P_D = P_D(\mathbf{x}_k)$. No additional assumptions for the clutter parameters λ and $c(\mathbf{z})$ are needed. The initial PDF, if given, is a GM of the form

$$p_{0|0}(\mathbf{x}_0) = \sum_{h=1}^{\mathcal{H}_{0|0}} w_{0|0}^h \mathcal{N}(\mathbf{x}_0; \mathbf{m}_{0|0}^h, \mathbf{P}_{0|0}^h) \tag{5.37}$$

where $\mathcal{H}_{0|0}$ is the number of targets that are assumed present at the initial time, and $w_{0|0}^h$, $\mathbf{m}_{0|0}^h$, and $\mathbf{P}_{0|0}^h$ are the weight, mean, and covariance matrix of the h-th target, respectively. Note, that if the initial probability of existence associated with the h-th target $q_{0|0}^h$ is given, the weight can be computed as $w_{0|0}^h = \frac{q_{0|0}^h}{\mathcal{H}_{0|0}}$. Consider the expected number of births N_B is given, and the birth PDF $p_B(\mathbf{x}_k)$ is a GM of the form (4.15). As discussed in Section 5.2.1, new targets can be easily initialized during the run using the birth PDF, however, we assume $p_B(\mathbf{x}_k)$ to be time-independent in this thesis. Finally, assume the probability of survival P_S is given.

Similarly to the Bernoulli filter, such assumptions result into the spatial PDFs of the prior and posterior PPPs being Gaussian mixtures, denoted by

$$p_{k|k-1}(\mathbf{x}_k) = \sum_{h=1}^{\mathcal{H}_{k|k-1}} w_{k|k-1}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k-1}^h, \mathbf{P}_{k|k-1}^h) \tag{5.38}$$

$$p_{k|k}(\mathbf{x}_k) = \sum_{h=1}^{\mathcal{H}_{k|k}} w_{k|k}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h) \tag{5.39}$$

where $\mathcal{H}_{k|k-1}$, and $\mathcal{H}_{k|k}$ are total number of components in the prior and posterior PDFs at time k, respectively.

Assume the posterior expected number of targets from the last time instant is $\overline{N}_{k-1|k-1}$. From (5.10) and the KF equations (2.16a,2.16b), the time-update of the spatial PDF is

$$p_{k|k-1}(\mathbf{x}_k) = \frac{N_B}{\overline{N}_{k|k-1}} \sum_{i=1}^{N_{B,k}} w^i_{B,k} \mathcal{N}(\mathbf{x}_k; \mathbf{m}^i_{B,k}, \mathbf{Q}^i_{B,k}) + \\ + \frac{P_S \overline{N}_{k-1|k-1}}{\overline{N}_{k|k-1}} \sum_{h=1}^{\mathcal{H}_{k-1|k-1}} w^h_{k-1|k-1} \mathcal{N}(\mathbf{x}_k; \mathbf{m}^h_{k|k-1}, \mathbf{P}^h_{k|k-1}) \tag{5.40a}$$

$$\text{where} \quad \mathbf{m}^h_{k|k-1} = \mathbf{F}\mathbf{m}^h_{k-1|k-1} \tag{5.40b}$$

$$\mathbf{P}^h_{k|k-1} = \mathbf{F}\mathbf{P}^h_{k-1|k-1}\mathbf{F}^T + \mathbf{Q} \tag{5.40c}$$

where the time-updated expected number of targets $\overline{N}_{k|k-1}$ is simply given by Eq. (5.9). Eq. (5.40a) can be rewritten into a GM of a form (5.38), with $\mathcal{H}_{k|k-1} = \mathcal{H}_{k-1|k-1} + N_{B,k}$ being the prior number of GM components.

For the Bayes-update, the formula is slightly different from the single target tracking case. We first discuss the update of the spatial PDF. In order to compute the integral (5.32) from the denominator for each measurement, compute the probabilities (5.36). From the single-target measurement prediction (2.17), it follows that

$$p_{\mathbf{z}_k}(\mathbf{z}) = \int_{\mathcal{X}} l(\mathbf{z}|\mathbf{x}_k) p_{k|k-1}(\mathbf{x}_k) d\mathbf{x}_k = \sum_{h=1}^{\mathcal{H}_{k|k-1}} w^h_{k|k-1} \mathcal{N}(\mathbf{z}_k; \hat{\mathbf{z}}^h_{k|k-1}, \mathbf{P}^{\hat{\mathbf{z}},h}_{k|k-1}) \tag{5.41a}$$

$$\text{where} \quad \hat{\mathbf{z}}^h_{k|k-1} = \mathbf{H}\mathbf{m}^h_{k|k-1} \tag{5.41b}$$

$$\mathbf{P}^{\hat{\mathbf{z}},h}_{k|k-1} = \mathbf{H}\mathbf{P}^h_{k|k-1}\mathbf{H}^T + \mathbf{R} \tag{5.41c}$$

Then, form the unnormalised spatial posterior PDF

$$p_{k|k}(\mathbf{x}_k) \propto \left(1 - P_D + \sum_{\mathbf{z} \in Z_k} \frac{l(\mathbf{z}|\mathbf{x}_k) P_D}{\lambda c(\mathbf{z}) + P_D \overline{N}_{k|k-1} p_{\mathbf{z}_k}(\mathbf{z})}\right) p_{k|k-1}(\mathbf{x}_k) \tag{5.42a}$$

$$\propto \sum_{h=1}^{\mathcal{H}_{k|k-1}} w^h_{k|k-1}(1 - P_D) \mathcal{N}(\mathbf{x}_k; \mathbf{m}^h_{k|k-1}, \mathbf{P}^h_{k|k-1}) + \\ + \sum_{h=1}^{\mathcal{H}_{k|k-1}} \sum_{i=1}^{m_k} w^h_{k|k-1} \frac{P_D \mathcal{N}(\mathbf{z}^i_k; \hat{\mathbf{z}}_{k|k-1}, \mathbf{P}^{\hat{\mathbf{z}},h}_{k|k-1})}{\lambda c(\mathbf{z}^i_k) + P_D \overline{N}_{k|k-1} p_{\mathbf{z}_k}(\mathbf{z}^i_k)} \mathcal{N}(\mathbf{x}_k; \mathbf{m}^h_{k|k}, \mathbf{P}^h_{k|k}) \tag{5.42b}$$

where $\mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h$ are given by the KF update equations (2.16a-2.16e)

$$\mathbf{m}_{k|k}^h = \mathbf{m}_{k|k-1}^h + \mathbf{K}_k(\mathbf{z}_k^i - \mathbf{H}\mathbf{m}_{k|k-1}^h) \tag{5.43}$$

$$\mathbf{P}_{k|k}^h = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_{k|k-1}^h \tag{5.44}$$

$$\text{with} \qquad \mathbf{K}_k = \mathbf{P}_{k|k-1}^h\mathbf{H}^T(\mathbf{H}\mathbf{P}_{k|k-1}^h\mathbf{H}^T + \mathbf{R})^{-1} \tag{5.45}$$

Observe, that the unnormalised posterior (5.42b) can be written as a GM, and denote it

$$p_{k|k}(\mathbf{x}_k) \propto \sum_{h=1}^{\mathcal{H}_{k|k}} \tilde{w}_{k|k}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h) \tag{5.46}$$

Where $\mathcal{H}_{k|k} = \mathcal{H}_{k|k-1}(1 + m_k)$. The final step si normalising the posterior. Noting the proportional constant is equal to $\frac{\overline{N}_{k|k-1}}{\overline{N}_{k|k}}$ and $\overline{N}_{k|k-1}$ is known, the final update is given by

$$p_{k|k}(\mathbf{x}_k) = \sum_{h=1}^{\mathcal{H}_{k|k}} w_{k|k}^h \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k|k}^h, \mathbf{P}_{k|k}^h), \quad \text{where } w_{k|k}^h = \frac{\tilde{w}_{k|k}}{\sum_{h=1}^{\mathcal{H}_{k|k}} \tilde{w}_{k|k}} \tag{5.47}$$

$$\overline{N}_{k|k} = \overline{N}_{k|k-1} \sum_{h=1}^{\mathcal{H}_{k|k}} \tilde{w}_{k|k} \tag{5.48}$$

In order to moderate computational costs, the pruning and capping (see Section 2.3) can be readily used. The gating is discussed later.

Obviously, the MMSE and MPH estimates do not make sense for the intensity (PHD) filter. Note, that when dealing with a multi-target PDF, the classical maximum-aposteriori (MAP) and expected-aposteriori (EAP) state estimators are not even defined in general [4, pp. 494-508]. Since we deal with a GM representation of the intensity function, the means of the posterior GM components are its local maxima (if they are not too closely spaced). We can take simply $\overline{N}_{k|k}$ means corresponding to Gaussian densities with the largest weights, $\overline{N}_{k|k}$ arguments of highest peaks, or all components with a weight greater than a threshold.

Note, that the GM components in the prior or posterior PDFs cannot be interpreted as hypotheses similarly to the single-target sense. They contain information about a possible evolution of all the targets, without any specification which component corresponds to which target evolution. Because of this interpretation difficulties, we argue that gating should not be performed, unless whenever any new target appears to the scenario, its initial spatial PDF is known and properly incorporated into the birth spatial PDF.

The GM components represent the whole intensity function (or PHD), i.e. a joint information for all the targets. From the filter derivation it can be seen that no target trajectories are modeled. Forcing the filter to yield trajectory estimates can be done [66, 67, 68] but it is rather heuristic[13]. From the very recent theory of filtering with sets of trajectories, it is possible to form the trajectory PHD filter [69, 70], as an approximate propagation of the so-called multi-trajectory density through (an appropriate) BRRs.

Other Notes on Intensity/PHD Filter

Now we discuss how i.i.d. approximation to the objects might be poor. First, consider sampling. Imagine taking two independent samples from a multi-modal PDF with two local maxima. If the peaks represent targets, then from the independent property, the two samples can happen to be generated both near one of those peaks. Such situation cannot happen e.g. in case of modeling with multi-Bernoulli distribution. Second, consider the Poisson cardinality distribution. In case of having a known number of targets present, for example one, then even if the expected number of the targets was equal to one, the probability of having exactly one target is $p_N(1) = \frac{e^{-1}1^1}{1!} = e^{-1} \approx 0.37$, which is considerably low value.

Assume, that exactly one object is present. From the additional weighting of each measurement, it is obvious that even in case of setting the expected number of targets to one, the PHD filter behaves differently, than the single-target filters discussed in previous chapters. However, it can be shown that all the presented filters (at least with respect to their behavior) can be straightly derived from the CPHD filter [4, pp. 641-642] [54, 71].

Other drawbacks and notes about the PHD filter can be found e.g. in [4]

[13] PHD filter, especially with its particle implementation, can be combined with other MTT filters as a "declutter" algorithm (followed by clustering), or to perform "global gating", see [68],[4, pp. 596-599].

CHAPTER 6

Practical Comparison

The previous chapters dealt with the theoretical basis of the presented filtering algorithms including their theoretical comparison. Since each of the presented algorithms is designed for different scenario and outputs various kinds of estimate, precise numerical comparison is meaningless. This chapter is thus devoted to a practical comparison only. Particularly, we address a specific real-world application, which is tracking of moving objects from video data (visual tracking), illustrated in Fig. 6.1.

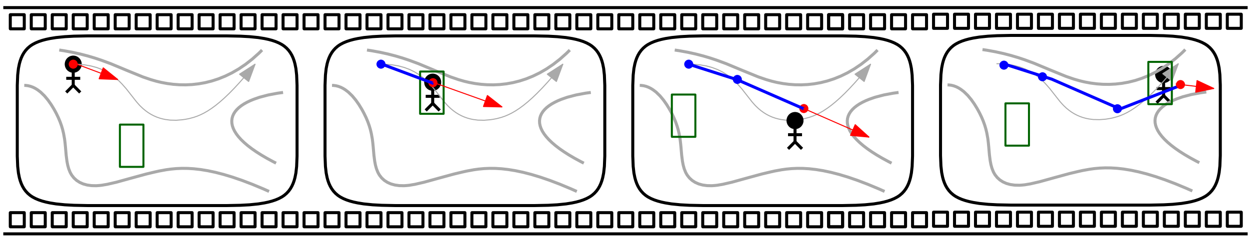


Fig. 6.1: Illustration of tracking a moving object from video data.

In the computer learning community, various techniques are used for such task [72, 73], largely involving ad-hoc non-Bayesian methods, focusing on different kinds of objectives and concrete applications, such as video surveillance and autonomous driving. In this thesis, we consider especially tracking of pedestrians and cars. Using a few easy-to-understand examples, we compare the presented algorithms with respect to their general behavior and practical use, rather than numerical performance.

In the following, we first discuss how the measurement-sets are created. Then we form a motion and measurement models, followed by the results and discussion.

6.1 Data Acquisition

The video data consist of a time-sequence of images (or frames), where each image consists of pixels. Since the presented algorithms assume to be given a set (or tuple) of measurements, that are points in some vector space $\mathcal{Z}$ at a time, we must use an additional algorithm, that takes each image as its input, and returns such measurements at its output. Such additional algorithm is commonly addressed as image object detector and comes from the area of computer vision, see [74, 75] for a survey. There are various types of detectors, such as those focusing on detecting features, contours, faces, body parts, cars, or other objects, as well as those focusing on detection of multiple classes of objects at once. We chose to use the so-called *you only look once* version 3 (YOLOv3) image detector [76] with the Darknet implementation in C [77, 78]. This algorithm makes the use of neural networks [73] and is known especially for its good ratio of time demands and accuracy. To push our examples closer to real-world computational constraints, we use an alternative set of weights for the neural network, referred to as *tiny* weights, which are especially designed for constrained environments. Therefore, the algorithm output is worse than that commonly claimed for YOLOv3, but the algorithm works much faster, being ready for constrained (e.g. on-line) use. We used the value 0.1 for the detection threshold.

For a given video frame, YOLOv3 outputs a set of so-called bounding boxes with each one being associated with some class of objects, such that there is e.g. a person, a car or a horse, inside the bounding box. For simplicity, we use only the center of each bounding box to fill in the measurement-set for the tracking algorithms. In another words, each measurement $\mathbf{z}_k$ represents a two-dimensional position $[x_k, y_k]^T$ of the object within the image coordinates (in pixels). No information about the object width, height, or a class to which it belongs to, is used for tracking the object. Therefore, the following results are rather illustrative, leaving any improvements to a possible consequent work. Since we focus on a general behavior of the tracking algorithms, we do not present any specific evaluation metrics, used to compare tracking performance numerically. It should be noted, that there exist various evaluation techniques, specific for either the visual object tracking field (see e.g. [73, pp. 64-65][79, pp. 4-5]), and the general MTT [80, 81].

Above we discussed how the measurements are generated from the image observation. We now make an assumption about the dynamic model of the objects, i.e. the Markov transition density $\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the measurement likelihood function $l(\mathbf{z}_k|\mathbf{x}_k)$.

6.2 Model of the Objects

For simplicity, we model the individual targets to be 4-dimensional random vectors $\mathbf{x}_k = [x, v_x, y, v_y]^T$, such that both their position x, y and velocity v_x, v_y are 2D as if they moved on the surface of the images, rather than "inside them" in 3D. Therefore, the measurements $\mathbf{z}_k$ are position observations. In order to use the GM implementation without any further approximations, we use the *almost constant velocity* model [82], which is linear. It can be written in terms of (2.11,2.12), where the state transition matrix $\mathbf{F}$, state noise covariance matrix $\mathbf{Q}$, measurement matrix $\mathbf{H}$ and the measurement noise covariance matrix $\mathbf{R}$ are

$$\mathbf{F} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{Q} = q \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} & 0 & 0 \\ \frac{T^2}{2} & T & 0 & 0 \\ 0 & 0 & \frac{T^3}{3} & \frac{T^2}{2} \\ 0 & 0 & \frac{T^2}{2} & T \end{bmatrix} \tag{6.1}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad \mathbf{R} = R \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{6.2}$$

where $T > 0$ is the sampling period, and $q > 0, R > 0$ are parameters of the process and measurement noises. The values of these parameters differ within the following examples.

6.3 Demonstrations

There are many datasets for the video detection and tracking algorithms, see e.g. [79]. The videos used in this thesis were chosen from several different datasets, that are outlined in the following text. The complete results are saved as video files or figures on a CD attached to this thesis, and also can be found at [83]. Overall results are discussed here, providing a few sample photos and graphs to get an intuition. Note, that for the sake of clarity, some axes are omitted in the following figures. The trajectories, and state estimates are depicted in red, with contours of posterior spatial PDF for a given time step in red and green. In the figures presenting the estimated positions with respect to time k, the measurements are depicted in gray. The algorithm names have been abbreviated with simple **NN**, **PDA**, **MHT**, **B** (for Bernoulli filter), and **PHD** (for intensity/PHD filter). Note that MHT stands for the single-target MHT as presented in Section 3. For the MHT and Bernoulli filters, both the MMSE and MPH estimates were obtained, but only the results concerning the MMSE are presented in the following samples. For the Bernoulli filter, the trajectory estimate was drawn if $q_{k|k} \geq 0.5$, and for the PHD filter, the estimates of the states are means of components with the largest weights, and their number is rounded $\overline{N}_{k|k}$.

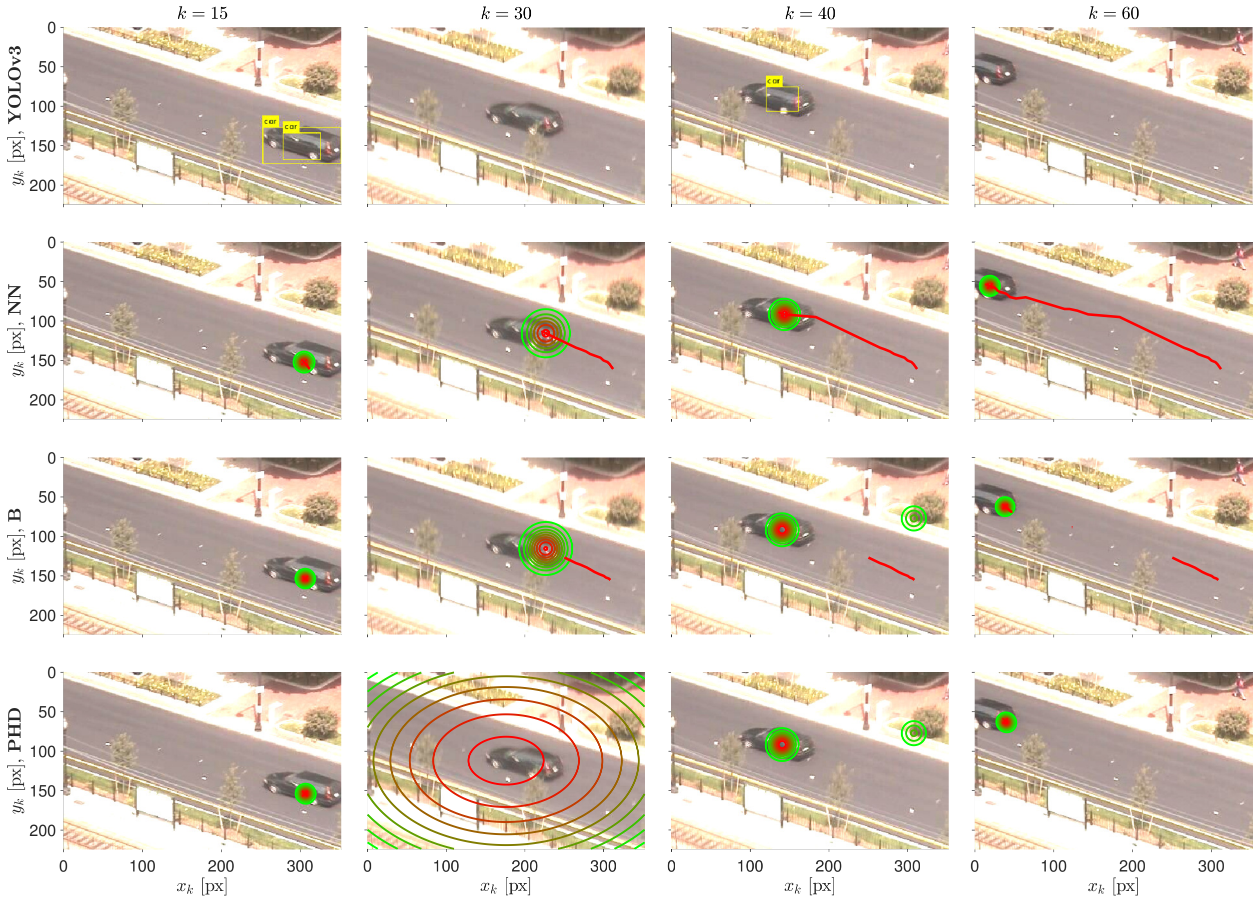


Fig. 6.2: Samples from *boulevard* video: YOLOv3, NN, Bernoulli, and PHD filters.

The values common for all the cases are as follows. The gating was performed with the gate $G = 2^2$. For the MHT, Bernoulli, and PHD filters, the pruning threshold was taken to be 0.01 and the capping maximum number of components was 100. The probability of birth $P_B = 0.01$ and of survival $P_S = 0.98$. The clutter was assumed to be uniformly distributed over the whole picture. The birth process was modeled as a single Gaussian, placed in the middle of the area with a "large" covariance. No spawning was modeled. The initial PDF was chosen heuristically. The algorithms were implemented in MATLAB® R2019a.

6.3.1 Tracking of a Car With a Static Camera

The video *boulevard* is taken from the SBMnet dataset [84]. A car passing a road is undetected for a considerably large number of frames when being in the middle of the scene. For the algorithms, following parameters were used: the process and measurement noise coeffi-

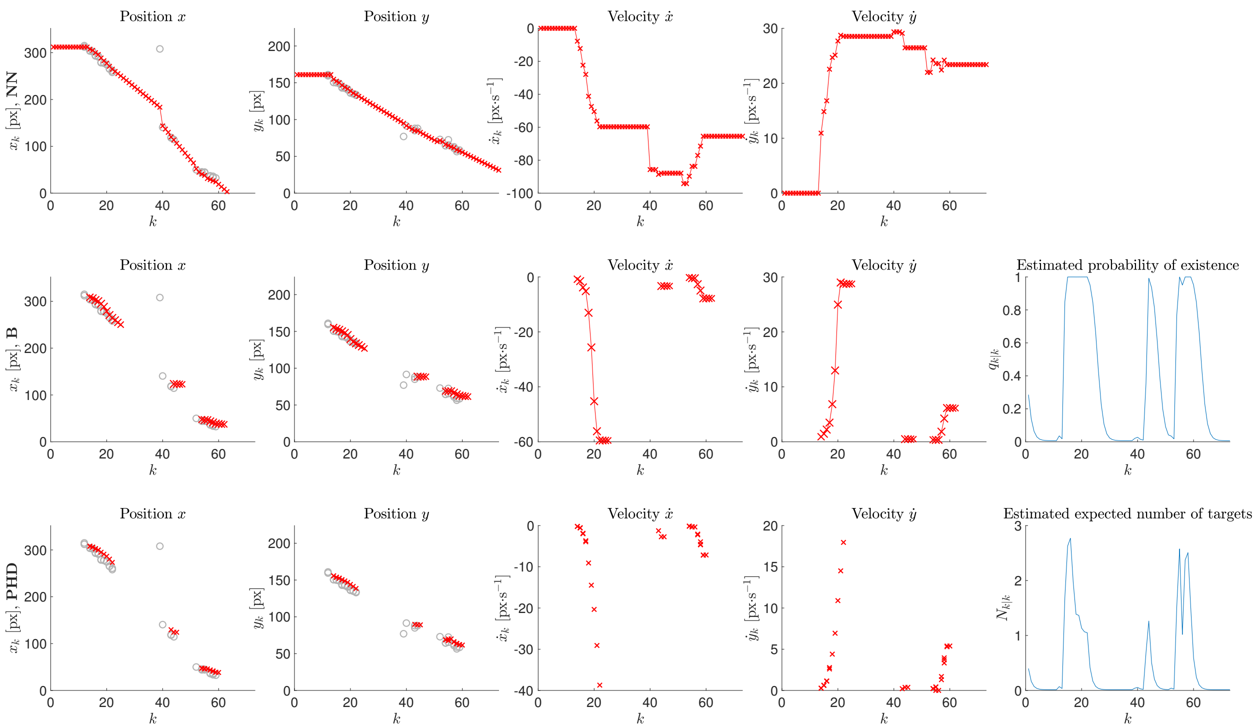


Fig. 6.3: Estimated values for a part of *boulevard* video: NN, Bernoulli, and PHD filters.

cients $q = 20^2$, $R = 10^2$, the probability of detection $P_D = 0.6$, the expected number of birth $N_B = 0.02$, and of clutter measurements $\lambda = 0.1$. Sample frames and resulting estimates from some of the algorithms are in Fig. 6.2 and 6.3, respectively.

The main differences among the algorithms can readily be seen, such as there are gaps within the estimated trajectory from the Bernoulli filter corresponding to time instants when $q < 0.5$. Due to occasional multiple detections, the PHD filter sometimes results into estimating multiple objects being present.

It can be noted, how the birth process influences the posterior spatial PDF especially for the PHD filter. The contours that tracked the object vanish after multiple undetections in a row (the birth process simultaneously gains large weight) and soon become even pruned off, i.e. there is only the birth PDF left in the middle of the frame, and the estimated expected number of objects is close to zero at time $k = 30$. Similar performance holds for the Bernoulli filter, but the vanishing process is slower. Since we know the car was present for the whole time horizon, we would naturally choose NN, PDA, or MHT. However, when looking solely at the measurements in Fig. 6.3 (gray circles in position subplots), one could guess the target was switching *off/on*, and so the Bernoulli filter might be more appropriate.

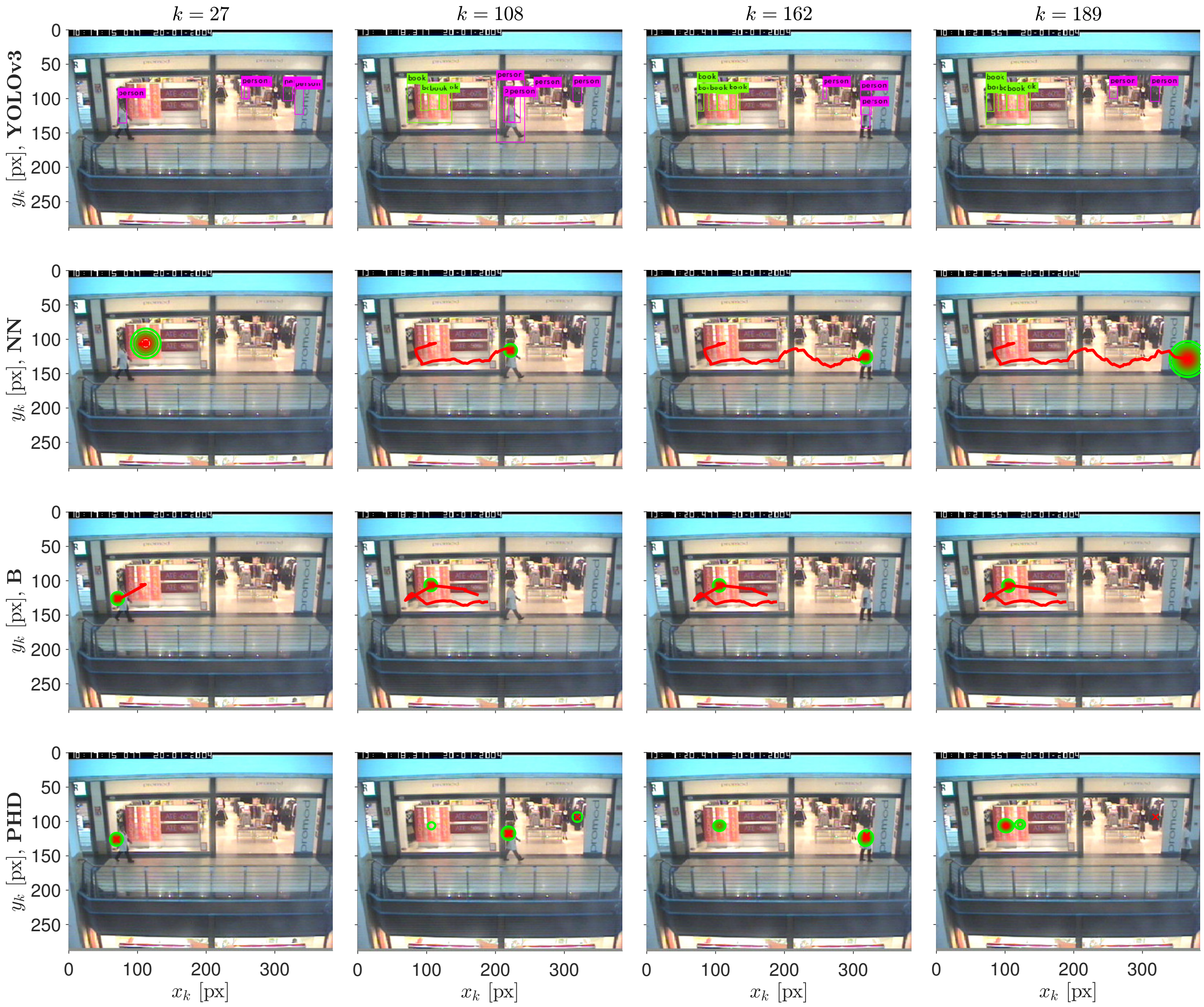


Fig. 6.4: Sample frames from *OneStopEnter1front_v1* video: YOLOv3, NN, Bernoulli, and PHD filters.

6.3.2 Tracking of a Customer in a Shopping Mall

The video *OneStopEnter1front_v1* is taken from the CAVIAR dataset [85]. A human passes by a shop in a shopping mall. Other objects that are not to be tracked were detected by the YOLOv3, and so the clutter does not behave according to its model (uniform distribution). For the algorithms, the same parameters were used as for the *boulevard* video. Sample frames and resulting estimates from some of the algorithms are in Fig. 6.4 and 6.5, respectively.

The performance of the NN filter seems to be superior. For the Bernoulli filter, when the target is undetected a few times in a row (around $k = 90$)[1], the birth process gives rise

[1]Note the gap around that time step in the estimated probability of existence, due to which there is a

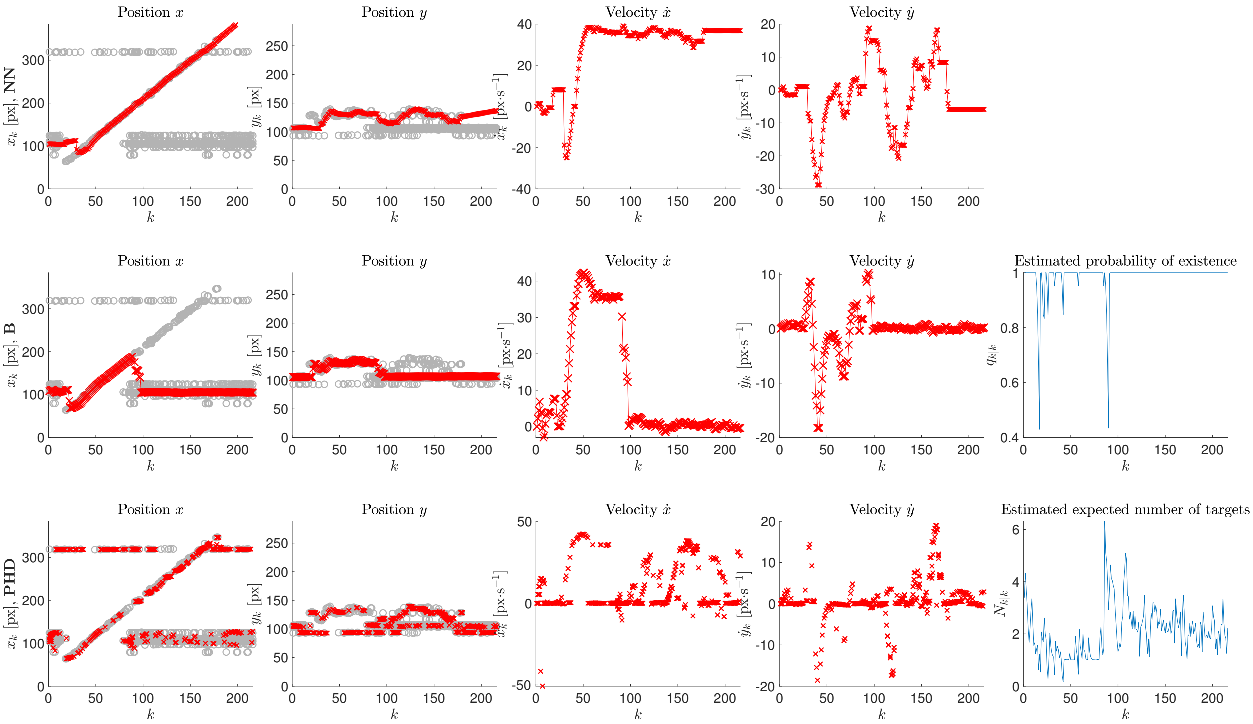


Fig. 6.5: Estimated values for a part of *OneStopEnter1front_v1* video: NN, Bernoulli, and PHD filters.

to track the static object behind the shop's window, which has been detected by YOLOv3, and hence the Bernoulli filter looses a track of the human.

Since the clutter measurements are actually not from clutter, but from static objects, the PHD filter tracks them (or finds their states). They behave according to the almost constant velocity motion model, with velocity being around zero.

gap in the estimated trajectory.

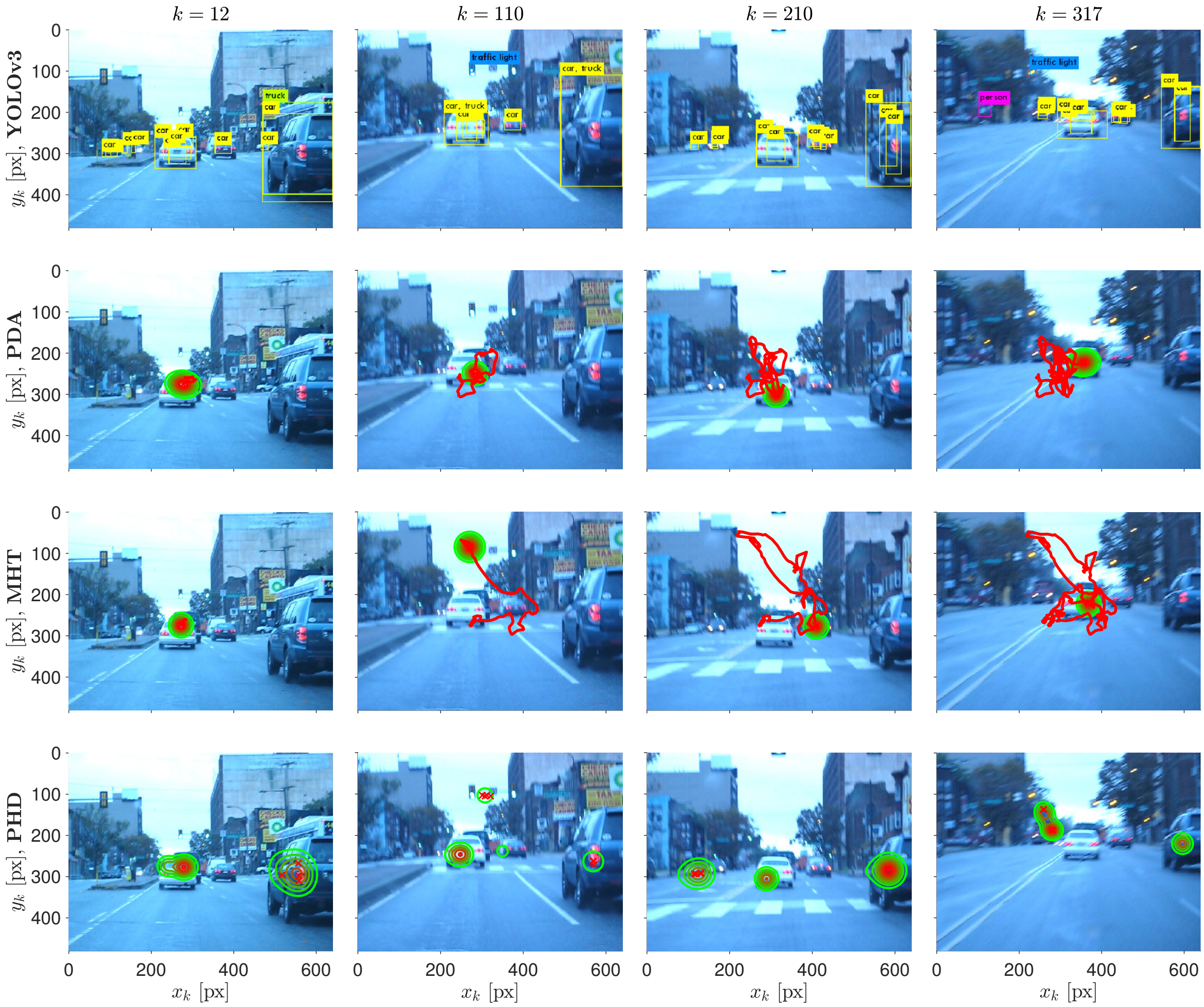


Fig. 6.6: Sample frames from *car3* video: YOLOv3, PDA, MHT, and PHD filters.

6.3.3 Automotive Tracking From Blurred and Tremulous Video

The video *car3* is taken from the BLUT dataset [86] and is also included in the visual tracker benchmark [87] dataset [88]. A camera, loosely attached probably to a windshield, shoots other cars while passing by a crossroad. For the algorithms, the following parameters were used: the process and measurement noise coefficients $q = 60^2$, $R = 35^2$, the probability of detection $P_D = 0.8$, the expected number of birth $N_B = 0.03$, and of clutter measurements $\lambda = 0.1$. Sample frames and resulting estimates from some of the algorithms are in Fig. 6.6 and 6.7, respectively.

If we consider tracking of the car, which rides in front of the windshield, we can say

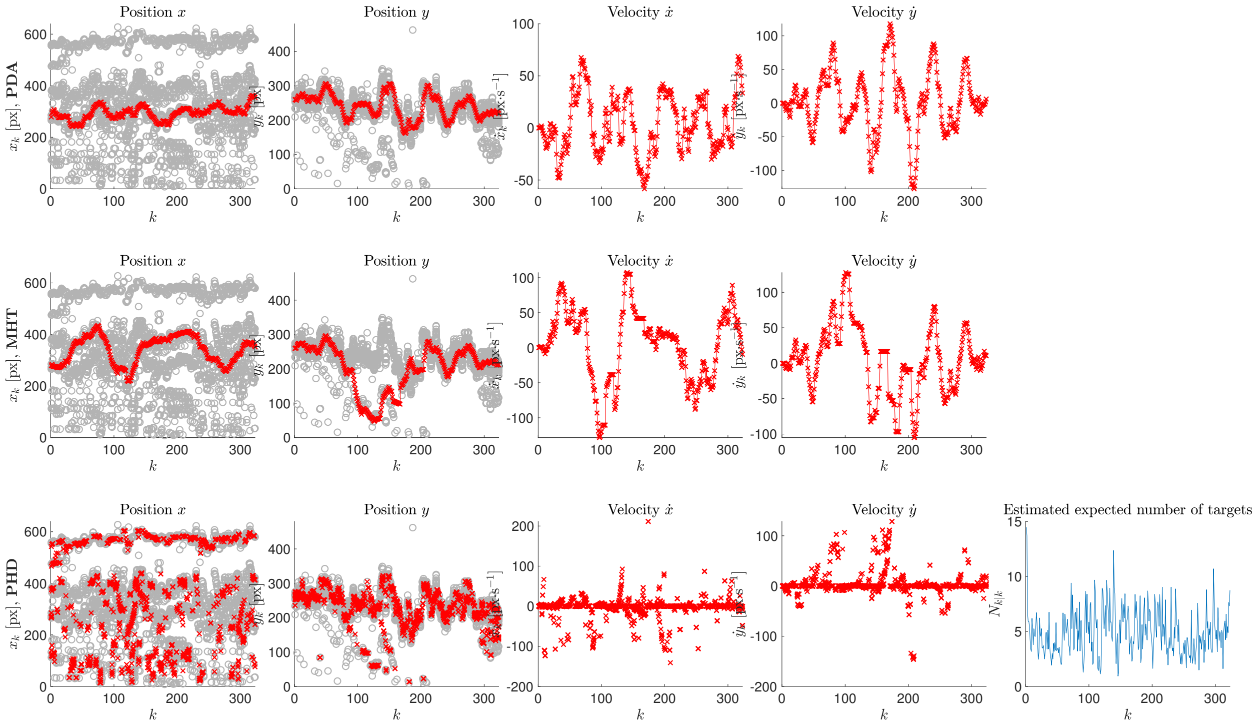


Fig. 6.7: Estimated values for a part of *car3* video: PDA, MHT, and PHD filters.

that the PDA filter significantly outperforms the other filters. When watching the resulting videos, the merging procedure can be straightforwardly seen as effective, while the MHT (and also both NN and Bernoulli filters) looses track of the car. The PHD filter can be used in case of tracking (or estimating actual states of) multiple existing objects, but it yields no trajectories at all.

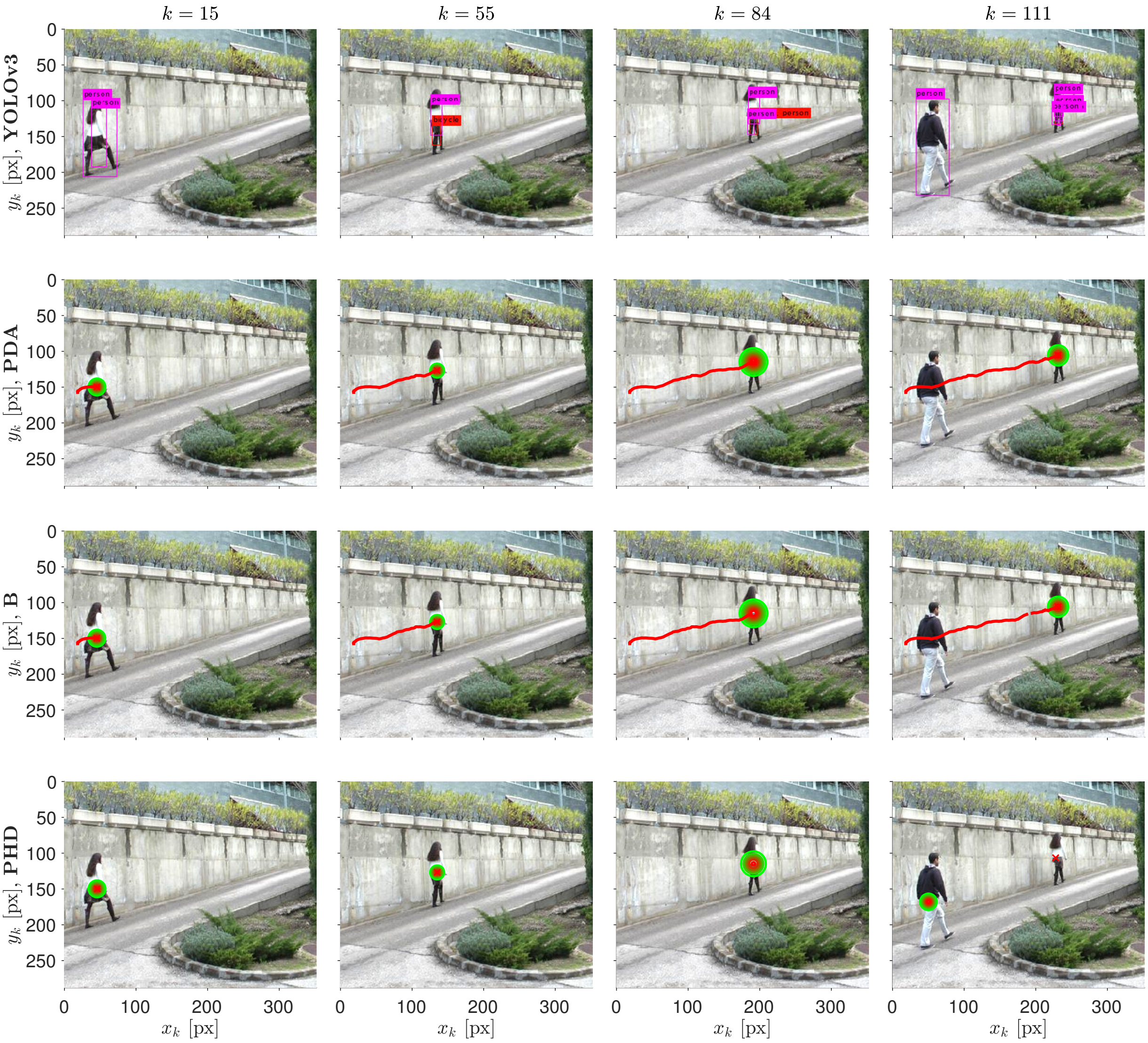


Fig. 6.8: Samples from *O_SM_04* video: YOLOv3, PDA, Bernoulli, and PHD filters.

6.3.4 Tracking of Pedestrians

The video *O_SM_04* is taken from the SBMnet dataset [84]. A pedestrian walks up a street and right before the end of the video, another pedestrian enters the scene. For the algorithms, the same parameters were used as for the *boulevard* video. The sample frames and resulting estimates from some of the algorithms are in Fig. 6.8 and 6.9, respectively.

If only the first pedestrian is to be tracked, since it is known to be present for the whole tracking time horizon, the NN, PDA, and MHT filters can be used efficiently. The Bernoulli filter might produce unwanted gaps in the trajectory estimates.

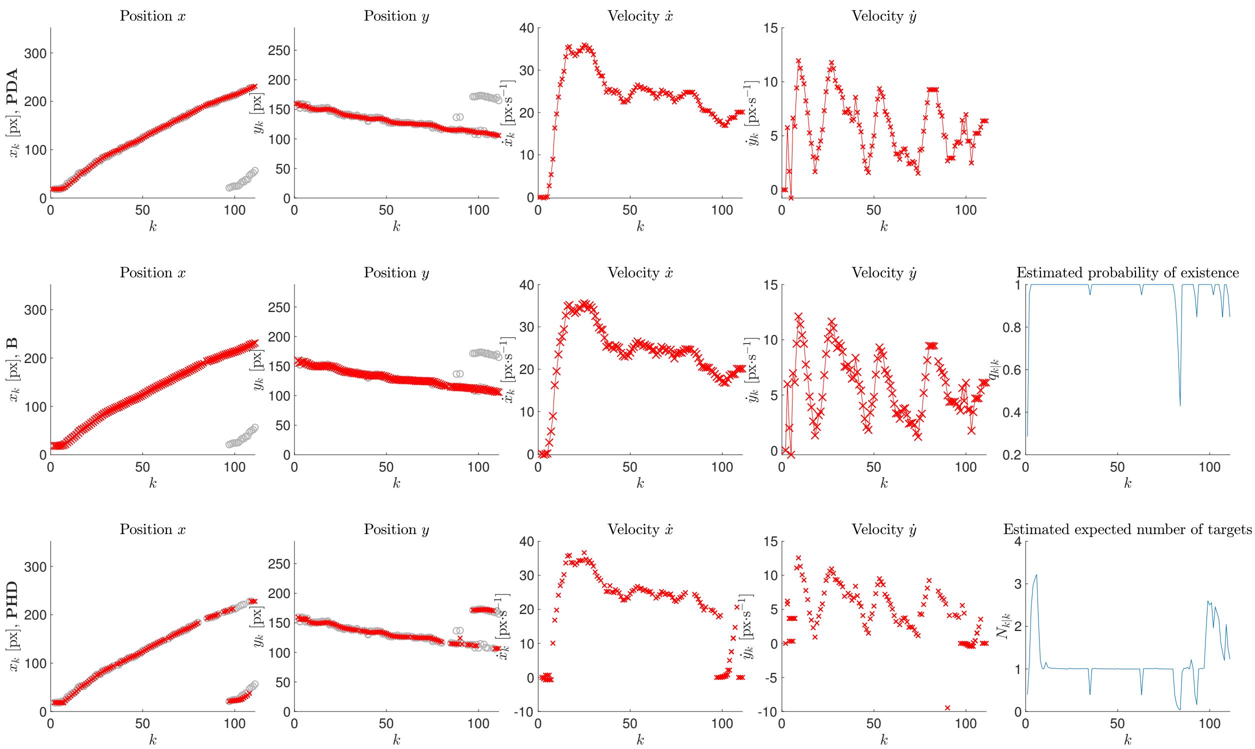


Fig. 6.9: Estimated values for a part of *O_SM_04* video for PDA, Bernoulli, and PHD filters.

For the PHD filter, an interesting behavior can be seen at the end of the video. When another pedestrian enters the scene, the filter "finds" him. However, even though the rounded $\overline{N}_{k|k}$ is equal to 2, resulting estimates are near to either one of the objects at a time. This phenomenon is closely related to an ambiguous state extraction from the posterior spatial PDF. If means of components with the largest weights are extracted (this case), the resulting estimated states are not necessarily the arguments of the highest peaks of the PDF, because the maximum of a Gaussian PDF depends on its covariance matrix. Therefore, we could say that the authoritative output from the PHD filter should be understood to be the whole posterior spatial PDF with the (not rounded) estimated expected number of components, not just the final estimated states themselves.

Fig. 6.10: Samples from *Jogging* video: YOLOv3, MHT, Bernoulli and PHD filters.

The video *Jogging* is taken from the dataset [88]. Two humans jogging on a sidewalk are captured by a moving camera. There is an occlusion of the jogging people caused by a traffic light pole, in the middle of the video. For the algorithms, the same parameters were used as for the *boulevard* video. The sample frames and resulting estimates from some of the algorithms are in Fig. 6.10 and 6.11, respectively.

If only the person on the right is to be tracked, then the MHT approach seems reasonable. However, strictly speaking, the person is nonexistent in the video during the frames when it is occluded by the traffic light pole. Therefore, the object is rather Bernoulli in fact.

If we realize that there is another person not being occluded by the time the other is, it

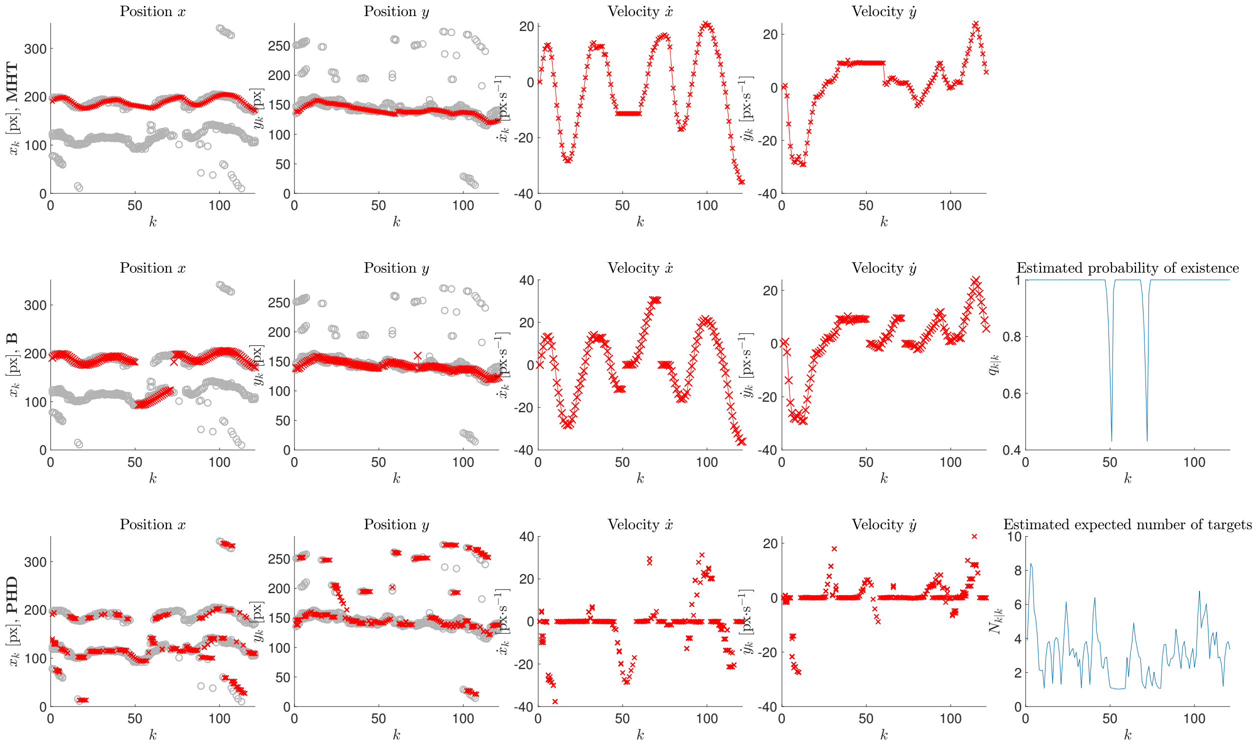


Fig. 6.11: Estimated values for a part of *Jogging* video: MHT, Bernoulli and PHD filters.

is natural that the Bernoulli filter tracks it. Therefore, we can say that the Bernoulli filter performs as expected.

For the PHD filter, the effect of ambiguous state extraction can be seen again e.g. at time step $k = 78$, when traffic lights gave rise to a GM component with large both the covariance and weight. Again, we should rely on the whole PHD filter output, rather than on the estimated states themselves.

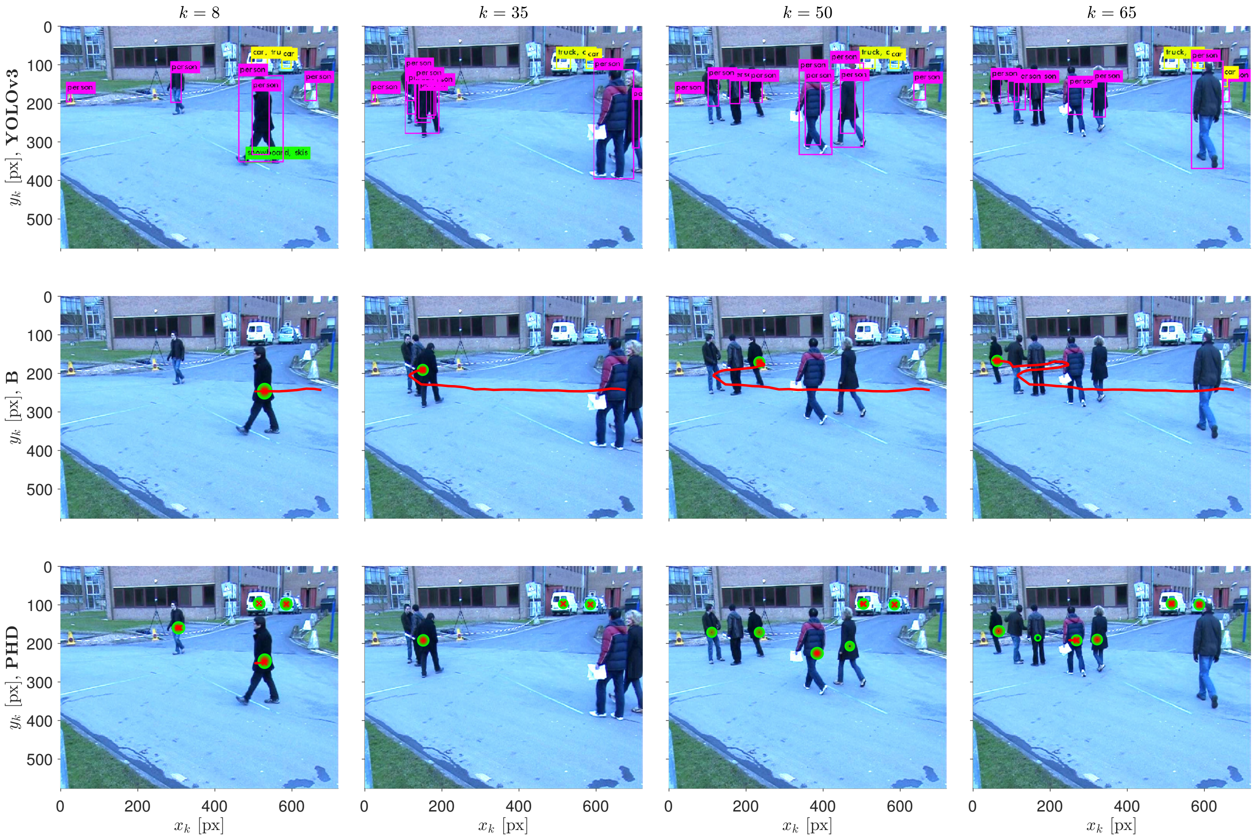


Fig. 6.12: Sample frames from *View_006_T12-T34* video: YOLOv3, Bernoulli and PHD filters.

The video *View_006_T12-T34* is taken from PETS 2009 dataset [89]. Multiple walking pedestrians are captured by a static camera. For the algorithms, the same parameters were used as for the *boulevard* video, except $\lambda = 0.2$. The sample frames and resulting estimates from some of the algorithms are in Fig. 6.12 and 6.13, respectively.

For single target tracking algorithms, consider tracking of the pedestrian that enters the scene on the right, follows to the left, then makes two switchbacks and walks out of the scene. All the single-target algorithms, such as the shown case of using the Bernoulli filter, track the mentioned pedestrian with considerable precision, even in the heavily cluttered environment, where the clutter obviously does not obey its model (which is a uniform distribution), but is rather generated from other real objects which are not considered to be tracked. Note, that such performance heavily depends on the chosen parameters, and also on the initial conditions. For instance, the initial condition was such that the moving object (pedestrian) has zero velocity. Since there are two static cars, parked in the right upper corner of the

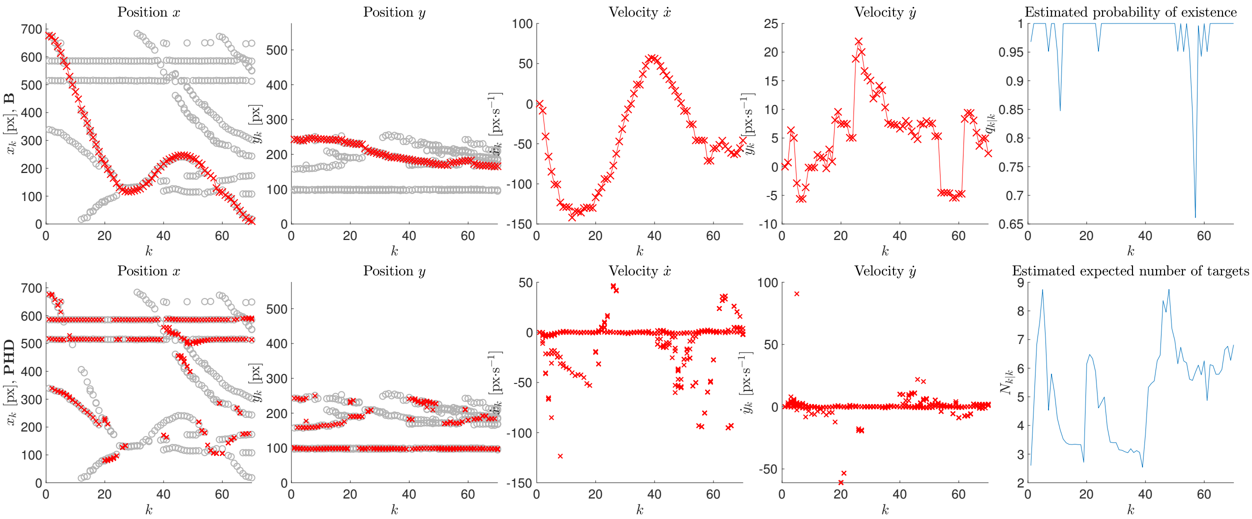


Fig. 6.13: Estimated values for a part of *View_006_T12-T34* video: Bernoulli and PHD filters.

scene next to the target's initial position, the filters could start tracking them instead of the pedestrian, since they have zero velocity (as the initial condition). However, this did not happen in the examples, mainly because of the performed gating.

In the case of the PHD filter, states of many objects are estimated, as they behave according to the single-target motion model. Again, the ambiguity of state extraction makes it harder to deduce the final states of the targets, hence the whole intensity function should be understood as the authoritative result.

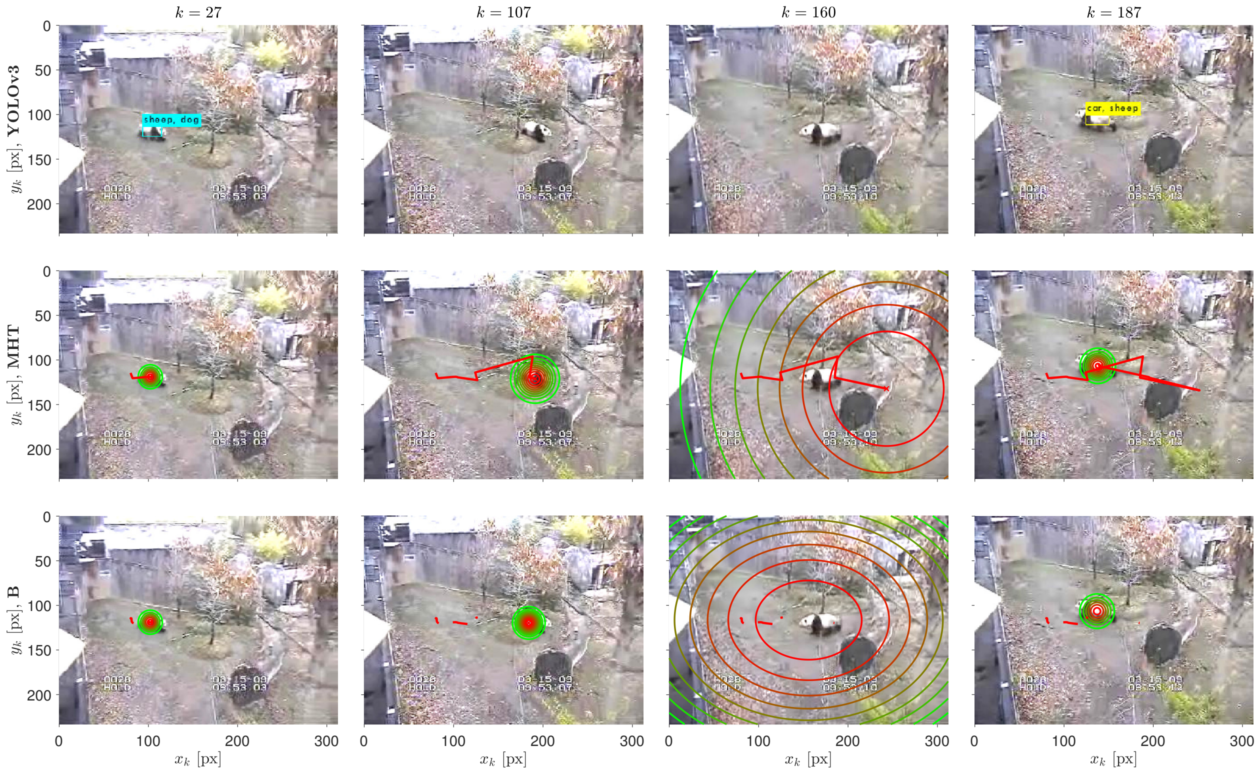


Fig. 6.14: Sample frames from *Panda* video: YOLOv3, MHT, and Bernoulli filters.

6.3.5 Tracking With Very Dense Measurements

The video *Panda* is taken from the dataset [88]. A panda takes a walk around a tree, probably in a zoo, and is captured by an occasionally moving camera. For the algorithms, the following parameters were used: the process and measurement noise coefficients $q = 20^2$, $R = 10^2$, the probability of detection $P_D = 0.3$, the expected number of birth $N_B = 0.02$, and of clutter measurements $\lambda = 0.1$. The sample frames and resulting estimates from some of the algorithms are in Fig. 6.14 and 6.15, respectively.

The neural network inside YOLOv3 was not probably trained to detect objects such as pandas, and thus struggles with detecting the panda. Only very few, and erroneously attributed detections are present. However, the tracking can be performed, but as it can be seen, the performance is considerably reduced.

According to the measurements, it is understandable that the MHT-based algorithms output trajectory with a few "jumps". When watching the video with the results for the Bernoulli filter, one can observe how the data-driven estimates of the panda vanishes when there are no measurements in a row and how the birth process takes place instead. Note,

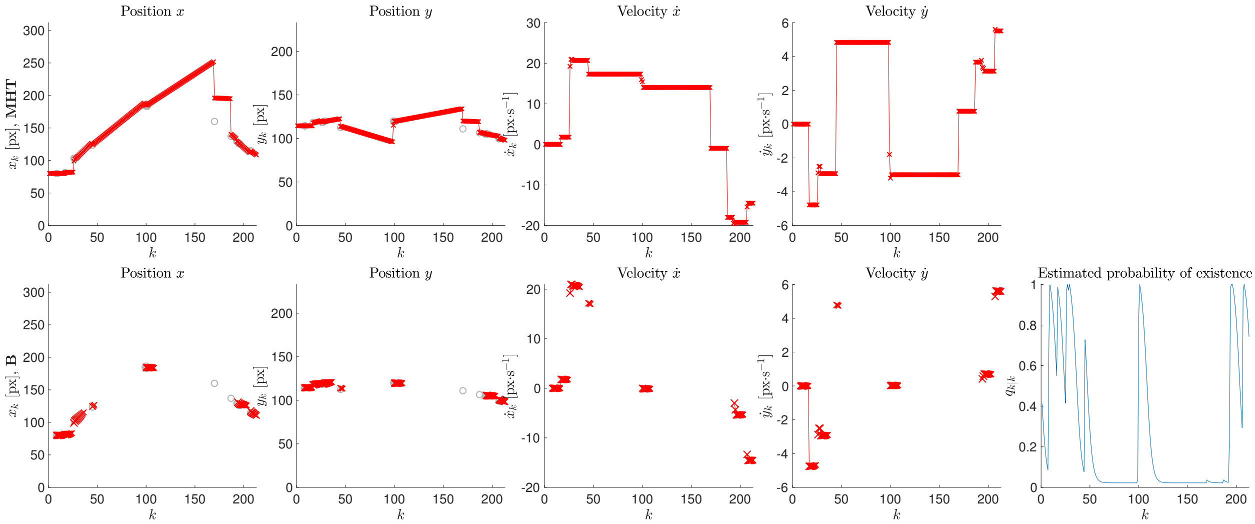


Fig. 6.15: Estimated values for a part of *Panda* video: MHT, and Bernoulli filters.

that $q_{k|k}$ does not become zero even if there is no measurement for a while, because the probability of birth is nonzero. To see this, assume there is no measurement at time k. Since P_D is constant, we get $\Delta^{HT} = P_D$, and then from the Bayes-update Eq. (4.13) we get

$$q_{k|k} = \frac{1 - P_D}{1 - q_{k|k-1} P_D} q_{k|k-1} \tag{6.3}$$

From the time-update Eq. (4.7), it then follows that

$$q_{k|k} = \frac{1 - P_D}{1 - \big(P_B(1 - q_{k-1|k-1}) + P_S q_{k-1|k-1}\big) P_D} \big(P_B(1 - q_{k-1|k-1}) + P_S q_{k-1|k-1}\big) \tag{6.4}$$

and in case there is no measurement for long time span enough, such that all the data-driven estimates are pruned out, a "steady state" of the probability of existence $q = q_{k|k} = q_{k-1|k-1}$ exists, such that it is a solution to the quadratic equation

$$(P_D P_B + P_D P_S) q^2 + \big(1 - P_D P_B + (1 - P_D) P_B - (1 - P_D) P_S\big) q - (1 - P_D) P_B = 0 \tag{6.5}$$

whose admissible solution for $P_D = 0.3, P_B = 0.01$, and $P_S = 0.98$ is approximately $q \approx 0.022$, which is the value that can be seen in Fig. 6.15. This concludes the examples.

CHAPTER 7

Conclusion

This thesis dealt with the tracking of moving object, focusing especially on fundamentals of the multi-target tracking (MTT) research direction. The assumptions of the classic filtering theory were being gradually relaxed starting with addition of an extraneous clutter measurements and the possibility of the target undetection. Then, a probability of the existence was introduced to model a possible nonexistence of the target for the whole tracking time horizon. Finally, the possibility of having multiple targets to track was discussed.

Three conceptually different approaches have been investigated. That is, the hypothesis-based approach that lead to the nearest neighbor (NN) filter, probabilistic data association (PDA) filter, and the special case of multiple hypothesis tracking (MHT) algorithm. Then, the random finite set (RFS) approach was studied, that lead to the Bernoulli filter and probability hypothesis density (PHD) filter (derived in Appendix B). Finally, especially for the multi-target case, the Poisson point processes approach was investigated, that lead to the intensity filter, which is known to be similar to the PHD filter without target spawning. The algorithms were carefully derived, reviewed, implemented with the Gaussian mixture filter technique, and both theoretically and practically compared.

A novel improvements to a derivation of a part of the Bayes-update in the intensity filter using a newly proven result and also using a continuous version of EM method was proposed.

The practical comparison pointed out various practical situations occurring in the real world problems of visual tracking, and showed how the filters behave in those situations, especially when their assumptions are not met. Since the algorithms provide different estimated values, such as probability of existence in the case of the Bernoulli filter or the expected number of objects in the case of the PHD filter, a precise numerical comparison was omitted.

Further theoretical aspects of modeling targets as sets, and also a more detailed version of the original derivation of the PHD filter were given in an appendix.

For the author's future work, emerging multi-target algorithms such as PMBM filter could be studied and implemented for further comparison. In general, the MTT is still under development. Overall future work in this field may include further investigation of modeling with sets of trajectories, measure-theoretic approaches, or some other approximation strategies to the multi-target Bayesian recursive relations or applying the theory to other practical situations. Also taking an inspiration from another fields, such as form quantum theory [90] was shown to be interesting. There are also many largely unexplored fields in MTT theory, such as adaptive filtering, identification, control, or tracking of continuously evolving targets.

APPENDIX A

Notes on Finite Set Modeling

In Section 2.2, we modeled both the moving objects and measurements as time-sequences of random finite sets (RFSs) or finite point processes. For the purpose of this thesis, the RFSs are viewed as a subclass of general point processes, that model the physical reality of unknown number of multiple targets as sets. Following the original construction of the RFSs with FISST theory [21, 4], they can be viewed as random variables, described by non-additive measures. In this chapter, we will give useful notes especially for the RFSs and FISST, providing a connection with the conventional probabilistic concepts that are used to deal with point processes. It should be noted, that the following steps might be slightly informal for the sake of analogy with the single-target case. For authoritative texts, see e.g. [91, 92, 93, 21, 4].

A.1 Deeper Introduction to the Modeling

Before introducing a probability measure, we discuss the underlying state space. For the point processes which are used to model targets, it can be defined as $E \triangleq \bigcup_{n=0}^{+\infty} \mathcal{X}^{(n)}$ [46], where $\mathcal{X}^{(n)}$ denotes the space of sets of a cardinality n with distinct members, with the notation $\mathcal{X}^{(0)} \triangleq \emptyset$. The space E can be called the abstract topological hyperspace [4, pp. 713], which is rather intractable to work with. Let $X \in E$ denote a realization and omitting technical details, $S \in E$ a measurable set. The probability $\Pr(\Xi \in S)$, can then be computed from the probability measure[1] $P_\Xi(S)$ using the formula

$$\Pr(\Xi \in S) = P_\Xi(S) = \int_S dP_\Xi(X) = \int_S p_\Xi(X) v(dX) \tag{A.1}$$

where v is the Lebesgue measure on E. The Radon-Nikodým derivative $p_\Xi(X)$, assuming it exists, can be called the multi-target PDF. In the case of the FISST, its direct analogy can be constructed from the belief mass measure using the set derivatives of a set integral [21,

[1]Note that this measure $P_\Xi(S)$ is additive, as in the single-target case.

pp. 144-151].

To make things less abstract, we now introduce the set functions

$$p_{\Xi|N}(X|n) \triangleq p_{\Xi|N}(\{\mathbf{x}^1, \dots \mathbf{x}^n\}|n) = n!\; p_{\mathbf{X}|N}(\mathbf{x}^1, \dots \mathbf{x}^n|n) \tag{A.2}$$

which for a given cardinality n are joint PDFs of the single-target states. A probability mass function (PMF) of a cardinality distribution $N \triangleq |\Xi|$ can be defined as $p_N(n) = \Pr(N = n)$. Then we can write the PDF of Ξ as

$$p_{\Xi}(X) \triangleq p_{\Xi}(X, n) = p_{\Xi}(\{\mathbf{x}^1, \dots, \mathbf{x}^n\}, n) = p_N(n) \cdot p_{\Xi}(X|n) \tag{A.3}$$

Now we can view the PDF as

$$p_{\Xi}(X) = \begin{cases} p_N(0) \cdot \overbrace{p_{\Xi|N}(\emptyset|0)}^{=1}, & \text{if } X = \emptyset \\ p_N(1) \cdot p_{\Xi|N}(\{\mathbf{x}^1\}|1), & \text{if } X = \{\mathbf{x}^1\} \\ p_N(2) \cdot p_{\Xi|N}(\{\mathbf{x}^1, \mathbf{x}^2\}|2), & \text{if } X = \{\mathbf{x}^1, \mathbf{x}^2\} \\ \quad\vdots & \quad\vdots \\ p_N(n) \cdot p_{\Xi|N}(\{\mathbf{x}^1, \dots, \mathbf{x}^n\}|n), & \text{if } X = \{\mathbf{x}^1, \dots, \mathbf{x}^n\} \\ \quad\vdots & \quad\vdots \end{cases} \tag{A.4}$$

It should be noted, that there exist other mathematical concepts used to deal with the point processes, such as Janossy and counting measures [91, 92, 46]. However, these are not essentially needed in this thesis. We have defined only the main concepts, which resemble the single-target case, especially its measure-theoretic fundamentals.

As well as in the single-target case, we need to be able to compute integrals, such as in the Chapman-Kolomogorov equation (2.9). The straightforward generalization into the multi-target case would involve the Lebesgue integral such as in Eq. (A.1), over the whole abstract space E. Now, we shall investigate the integral.

A.1.1 Connection Between Point Processes and FISST

To motivate the definition of the set integral and to comment on the correctness of modeling with the RFSs, we discuss the integral (A.1). Since S can be general, the measure v in (A.1) is counterintuitive. We could, however, take the integral with respect to some other measure, to be easier to understand [4, pp. 714-715],[37]. Define a generalized Lebesgue measure[2]

$$\mu_K(S) \triangleq \sum_{i=0}^{+\infty} \frac{v^i(S \cap \mathcal{X}^i)}{K^i \cdot i!} \tag{A.5}$$

where v^i is the Lebesgue measure on the space of ordered realizations $\mathcal{X}^i$ and K is the unit of measurement in $\mathcal{X}$. Note that K is introduced in order to guarantee that each summand in (A.5) is unitless and thus the sum is well defined.

The generalized Lebesgue integral, in order to motivate the definition[3] of set integral, will be taken over the whole $E = \bigcup_{n=0}^{+\infty} \mathcal{X}^{(n)}$. It is

$$\int_E p_\Xi(X)\mu_K(dX) = \sum_{n=0}^{+\infty} \int_{\mathcal{X}^{(n)}} p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\})\mu_K(d\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) \tag{A.6}$$

$$= \sum_{n=0}^{+\infty} \int_{\mathcal{X}^{(n)}} p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) \sum_{i=0}^{+\infty} \frac{v^i\big(d(\mathbf{x}^1, \ldots, \mathbf{x}^n) \cap \mathcal{X}^i\big)}{K^i \cdot i!} \tag{A.7}$$

The set $d(\mathbf{x}^1, \ldots, \mathbf{x}^n) \cap \mathcal{X}^i$ is nonempty and so its measure is nonzero only if $i = n$. Then,

$$\int_E p_\Xi(X)\mu_K(dX) = \ldots = \sum_{n=0}^{+\infty} \frac{1}{K^n \cdot n!} \int_{\mathcal{X}^n} p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\})v^n\big(d(\mathbf{x}^1, \ldots, \mathbf{x}^n) \cap \mathcal{X}^n\big) \tag{A.8}$$

$$= \sum_{n=0}^{+\infty} \frac{1}{K^n \cdot n!} \int_{\mathcal{X}^n} p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\})d\mathbf{x}^1 \cdots d\mathbf{x}^n \tag{A.9}$$

Now, by comparing the set integral of the set-valued function $f(X)$ over $\mathcal{S} \subseteq \mathcal{X}$ [4, 21]

$$\int_{\mathcal{S}} f(X)\delta X \triangleq \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{S}^n} f(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\})d\mathbf{x}^1 \cdots d\mathbf{x}^n \tag{A.10}$$

[2] A mapping between tuples $(\mathbf{x}^1, \ldots, \mathbf{x}^n) \in \mathcal{X}^n$ and sets $\{\mathbf{x}^1, \ldots, \mathbf{x}^n\} \in \mathcal{X}^{(n)}$ is for simplicity taken for granted and not written, i.e., whether $S \in \mathcal{X}^{(n)}$ or $S \in \mathcal{X}^n$ is assumed to be clear from the context.

[3] Note that usually the set integral is not defined through the classic probability theory, but rather using non-additive measures [21].

with the integral defined using the generalized Lebesgue measure (A.9), one can see that the set integral (A.10) is "obtained" when integrating over $S = \uplus_{n=0}^{+\infty}\mathcal{S}^{(n)}$, with $\mathcal{S} \subseteq \mathcal{X}$, and omitting the unit of measurements K. Therefore, the set integral is not a measure theoretic integral but it is much more intuitive and tractable for engineers.

In Section 2.2, the belief mass measure was defined in order to describe RFSs. We now observe its connection to the probability measure [94],

$$\beta_\Xi(\mathcal{S}) = \Pr(\Xi \in \uplus_{n=0}^{+\infty}\mathcal{S}^{(n)}) = P_\Xi(\uplus_{n=0}^{+\infty}\mathcal{S}^{(n)}) \tag{A.11}$$

Since the set integral (A.10) is defined to integrate only over the regions in the single-target state space[4], the belief mass measure is not additive. However, omitting the unit of measurement, the densities used in the FISST $p_{\text{"FISST"}}(X)$, resulting from the belief mass measures and the (measure-theoretic) true probability densities $p_{\text{"true"}}(X)$ are the same almost everywhere and they relate to each other such that

$$p_{\text{"true"}}(X) = K^{|X|} \cdot p_{\text{"FISST"}}(X) \tag{A.12}$$

Therefore, they are "in essence" the same thing[5]. Assuming $K = 1$, they are equal and hence as descriptions of targets they can be used interchangeably[6]. Without this notion, one may object the use of the multi-target BRRs with FISST (2.25-2.26) to be rather heuristic.

To sum up, using the FISST brings some tractability to the modeling. Note, that in general, the RFSs with FISST can be used to model concepts such as fuzzy sets, and others [4]. For further information on the connection between the approaches to RFSs, see [26].

A.2 Expectations

Defining expectations and statistical moments for the multi-target case cannot be straightforward generalization of the single-target case since the state space E, is not a vector space. The operations of adding (integrating) and scaling sets are undefined. Hence, to define the expectation it is required to find another structure having some "nice" features. The

[4]More precisely speaking, over sets from E, that are constructed based only on a given set $\mathcal{S}$ from $\mathcal{X}$.

[5]For more discussion and references, see [37, 46].

[6]The PHD computation using the newly proven result (2.30) then applies to the conventionally modeled RFSs. Hence the proposed derivation of the detected process update in the intensity filter 5.2.2.

expectations can be defined indirectly for functions of set arguments $\gamma(X)$ as

$$\mathrm{E}[\gamma(X)] \triangleq \int_{\mathcal{X}} \gamma(X)\, p_{\Xi}(X)\delta X \tag{A.13}$$

Choosing a particular function $\gamma(X)$ results in different structures.

A.3 Cardinality Distribution

The cardinality distribution $p_N(n) = \Pr(|\Xi| = n)$ can be computed as an expectation

$$\begin{aligned} p_N(n) = \mathrm{E}[\delta_n^{\mathrm{Kron}}(|\Xi|)] &= \int_{\mathcal{X}} \delta_n^{\mathrm{Kron}}(|X|)\, p_{\Xi}(X)\delta X && \text{(A.14)} \\ &= \sum_{i=1}^{+\infty} \frac{1}{i!} \int_{\mathcal{X}^i} \delta_n^{\mathrm{Kron}}(i)\, p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(A.15)} \\ &= \frac{1}{n!} \int_{\mathcal{X}^n} p_N(n) \cdot p_{\Xi|N}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}|n) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(A.16)} \end{aligned}$$

where δ_n^{Kron} is the Kronecker delta centered at $n \in \mathbb{N}$. The expected number of targets in the whole single-target state space $\mathcal{X}$ is then

$$\begin{aligned} \mathrm{E}_N[N] = \mathrm{E}_{\Xi}[|\Xi|] &= \sum_{n=0}^{+\infty} n\, p_N(n) && \text{(A.17a)} \\ &= \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} n\, p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(A.17b)} \\ &= \sum_{n=1}^{+\infty} \frac{1}{(n-1)!} \int_{\mathcal{X}^n} p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n && \text{(A.17c)} \\ &= \sum_{i=0}^{+\infty} \frac{1}{i!} \int_{\mathcal{X}^{i+1}} p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^{i+1}\}) d\mathbf{x}^1 \cdots d\mathbf{x}^{i+1} && \text{(A.17d)} \\ &= \sum_{i=0}^{+\infty} \frac{1}{i!} \int_{\mathcal{X}} \int_{\mathcal{X}^i} p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^i, \mathbf{w}\}) d\mathbf{x}^1 \cdots d\mathbf{x}^i d\mathbf{w} && \text{(A.17e)} \\ &= \int_{\mathcal{X}} \left(\sum_{i=0}^{+\infty} \frac{1}{i!} \int_{\mathcal{X}^i} p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^i\} \cup \{\mathbf{w}\}) d\mathbf{x}^1 \cdots d\mathbf{x}^i \right) d\mathbf{w} && \text{(A.17f)} \\ &= \int_{\mathcal{X}} \left(\int_{\mathcal{X}} p_{\Xi}(X \cup \{\mathbf{w}\}) \delta X \right) d\mathbf{w} && \text{(A.17g)} \end{aligned}$$

With a little effort, Eq. (A.17f) can be also expanded using the sum of Dirac deltas $\delta_{\mathbf{w}}(Y)$ centered at $\mathbf{w}$, as defined in (2.29),

$$\mathrm{E}_N[N] = \ldots = \int_{\mathcal{X}} \left(\sum_{i=0}^{+\infty} \frac{1}{i!} \int_{\mathcal{X}^i} \left(\sum_{\mathbf{y} \in Y} \delta_{\mathbf{w}}(\mathbf{y}) \right) p_{\Xi}(\{\mathbf{y}^1, \ldots, \mathbf{y}^i\}) d\mathbf{y}^1 \cdots d\mathbf{y}^i \right) d\mathbf{w} \tag{A.18a}$$

$$= \int_{\mathcal{X}} \left(\int_{\mathcal{X}} \delta_{\mathbf{w}}(X) p_{\Xi}(X) \delta X \right) d\mathbf{w} \tag{A.18b}$$

This idea can be generalized for $\mathcal{S} \subseteq \mathcal{X}$, i.e. the expected number of targets in a region $\mathcal{S}$, $\mathrm{E}[|\Xi \cap S|]$ with $S = \uplus_{n=0}^{+\infty} \mathcal{S}^{(n)}$ is

$$\mathrm{E}[|\Xi \cap \mathcal{S}|] \triangleq \mathrm{E}[|\Xi \cap S|] = \mathrm{E}[|\Xi| \cdot \mathbf{1}_S^{\Xi}] = \int_{\mathcal{S}} \left(\int_{\mathcal{S}} p_{\Xi}(X \cup \{\mathbf{w}\}) \delta X \right) d\mathbf{w} \tag{A.19a}$$

$$= \int_{\mathcal{S}} \left(\int_{\mathcal{S}} \delta_{\mathbf{w}}(X) p_{\Xi}(X) \delta X \right) d\mathbf{w} \tag{A.19b}$$

where the exponential notation (here used for $\mathbf{1}_S^X$) is taken from Eq. (2.32). The particular results of (A.19a) and (A.19b) leads to clever definition of the statistical moments.

A.4 Statistical Moments

Statistical moments cannot be generalized straightforwardly. Thus we define them as set-valued functions

$$D_{\Xi}(X) \triangleq \int_{\mathcal{X}} p_{\Xi}(X \cup W) \delta W \tag{A.20}$$

and call those the multi-target moment densities. For a fixed order n (the cardinality of X), we get

$$D_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}) = \int_{\mathcal{X}} p_{\Xi}(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\} \cup W) \delta W \tag{A.21}$$

An intuitive interpretation of the moment densities is such that they represent the probability density of n vectors in Ξ having the realizations $\mathbf{x}^1, \ldots, \mathbf{x}^n$ [38]. Fixing $n = 1$ we call $D_{\Xi}(\{\mathbf{x}\})$ the first moment density. From Eq. (A.19a), we can straightforwardly see its relationship with the expected number of targets and thus we can see that it is equal to the PHD (defined in Section 2.2) almost everywhere.

A.5 Notes on the PHD

A.5.1 Proof of Eq. (2.30)

In Section 2.2.4, it was stated that a PHD can be written as a product of the expected number of targets $\overline{N}$ and some PDF $p(\mathbf{x})$. We are about to rigorously specify the function $p(\mathbf{x})$, and prove the statement. Note, that this result should establish a connection between the PHD and the corresponding PDF given every single set of states.

$$\begin{aligned}
D_\Xi(\mathbf{x}) &= \int_{\mathcal{X}} p_\Xi(\{\mathbf{x}\} \cup W)\delta W && \text{(A.22a)}\\
&= \sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} p_\Xi(\{\mathbf{x}, \mathbf{w}^1, \ldots, \mathbf{w}^n\}) d\mathbf{w}^1 \cdots d\mathbf{w}^n && \text{(A.22b)}\\
&= \sum_{n=0}^{+\infty} \frac{n+1}{n+1} \cdot \frac{1}{n!} \int_{\mathcal{X}^n} p_N(n+1) \cdot p_{\Xi|N+1}(\{\mathbf{x}, \mathbf{w}^1, \ldots, \mathbf{w}^n\}|n+1) d\mathbf{w}^1 \cdots d\mathbf{w}^n && \text{(A.22c)}\\
&= \sum_{n=0}^{+\infty} (n+1) \cdot p_N(n+1) \int_{\mathcal{X}^n} p_{\mathbf{X}|N+1}(\mathbf{x}, \mathbf{w}^1, \ldots, \mathbf{w}^n|n+1) d\mathbf{w}^1 \cdots d\mathbf{w}^n && \text{(A.22d)}\\
&= \sum_{n=1}^{+\infty} n \cdot p_N(n) \int_{\mathcal{X}^n} p_{\mathbf{X}|N}(\mathbf{x}, \mathbf{w}^1, \ldots, \mathbf{w}^{n-1}|n) d\mathbf{w}^1 \cdots d\mathbf{w}^{n-1} && \text{(A.22e)}\\
&= \sum_{n=1}^{+\infty} n \cdot p_N(n) \cdot \underbrace{p_{\mathbf{X}|N}(\mathbf{x}|n)}_{\text{this is a PDF } \forall n} && \text{(A.22f)}
\end{aligned}$$

and from here we get two following results,

$$D_\Xi(\mathbf{x}) = \mathrm{E}_N[N \cdot p_{\mathbf{X}|N}(\mathbf{x}|N)] \tag{A.23}$$

$$D_\Xi(\mathbf{x}) = \underbrace{\mathrm{E}_N[N]}_{\overline{N}} \cdot \underbrace{\frac{\sum_{n=1}^{+\infty} n \cdot p_N(n) \cdot p_{\mathbf{X}|N}(\mathbf{x}|n)}{\sum_{n=1}^{+\infty} n \cdot p_N(n)}}_{\text{the PDF } p(\mathbf{x}) \triangleq} = \overline{N} \cdot p(\mathbf{x}) \tag{A.24}$$

where we see that $p(\mathbf{x})$ is the normalized (by $\overline{N}$) marginal PDF of the ordered Ξ for X being a singleton set. If the elements of Ξ were i.i.d, then $p_{\mathbf{X}|N}(\mathbf{x}|n)$ would not depend on the number of the targets. For the Poisson RFS (PPP), the PDF $p(\mathbf{x})$ is the spatial distribution.

A.5.2 Proof of Eq. (2.39), Recovery of the PHD

To see why Eq. (2.39) holds, expand it using linearity[7] of expectation

$$D_\Xi(\mathbf{x}) = D_\Xi(\{\mathbf{x}\}) = \left[\frac{\delta G_\Xi[h]}{\delta \mathbf{x}}\right]_{h=1} = \int_{\mathcal{X}} \left[\frac{\delta}{\delta \mathbf{x}} h^W\right]_{h=1} p_\Xi(W)\delta W \tag{A.25}$$

where h^W is also a functional. Therefore we can expand this term for every fixed $W = \{\mathbf{w}^1, \ldots, \mathbf{w}^n\}$ using the product rule as

$$\begin{aligned}
\left[\frac{\delta}{\delta \mathbf{x}} h^W\right]_{h=1} &= \left[\frac{\delta}{\delta \mathbf{x}}\Big(h(\mathbf{w}^1)\cdots h(\mathbf{w}^n)\Big)\right]_{h=1} && \text{(A.26a)}\\
&= \left[\sum_{i=1}^{n} h(\mathbf{w}^1)\cdots\left(\frac{\delta}{\delta \mathbf{x}} h(\mathbf{w}^i)\right)\cdots h(\mathbf{w}^n)\right]_{h=1} && \text{(A.26b)}\\
&= \left[\sum_{i=1}^{n} h(\mathbf{w}^1)\cdots\delta_{\mathbf{x}}(\mathbf{w}^i)\cdots h(\mathbf{w}^n)\right]_{h=1} && \text{(A.26c)}\\
&= \sum_{i=1}^{n} \delta_{\mathbf{x}}(\mathbf{w}^i) = \sum_{\mathbf{w}\in W} \delta_{\mathbf{x}}(\mathbf{w}) = \delta_{\mathbf{x}}(W) && \text{(A.26d)}
\end{aligned}$$

where $\delta_{\mathbf{x}}(W)$ is from the notation of Eq. (2.29). Thus we obtain

$$\int_{\mathcal{X}} \left[\frac{\delta}{\delta \mathbf{x}} h^W\right]_{h=1} p_\Xi(W)\delta W = \int_{\mathcal{X}} \delta_{\mathbf{x}}(W) p_\Xi(W)\delta W \tag{A.27}$$

which is the alternative definition of the first moment density and therefore the PHD. Also the higher order moment densities can be derived from the PGFL in a similar manner. For the general proof see [56, 38].

[7] Strictly speaking, the Leibniz integration rule was used to interchange the integral and derivative.

A.5.3 Examples of PHD

In case of having **exactly one object** to describe, i.e. $p(X) = 0$ for $|X| \neq 1$, the PHD is

$$D_\Xi(\mathbf{x}) = \int_\mathcal{X} p_\Xi(\{\mathbf{x}\} \cup W)\delta W = p_\Xi(\{\mathbf{x}\}) + \int_\mathcal{X} 0 \; d\mathbf{w}^1 + \dots = p_\Xi(\{\mathbf{x}\}) \tag{A.28}$$

In case of having **exactly two objects**, the PHD is

$$D_\Xi(\mathbf{x}) = 0 + \int_\mathcal{X} p_\Xi(\{\mathbf{x}, \mathbf{w}^1\}) d\mathbf{w}^1 + \frac{1}{2!}\int_{\mathcal{X}^2} 0 \; d\mathbf{w}^1 d\mathbf{w}^2 + \dots \tag{A.29a}$$

$$= \underbrace{p_N(2)}_{1} \cdot 2 \int_\mathcal{X} p_{\mathbf{X}|N}(\mathbf{x}, \mathbf{w}^1|2) d\mathbf{w}^1 = 2 \cdot \underbrace{p_{\mathbf{X}|N}(\mathbf{x}|2)}_{p(\mathbf{x})\triangleq} = 2 \cdot p(\mathbf{x}) \tag{A.29b}$$

If for example, $p_\Xi(\{x^1, x^2\}|2) = \frac{1}{2}\big(\mathcal{N}(x^1, 0, 1)\mathcal{N}(x^2, 1, 2) + \mathcal{N}(x^2, 0, 1)\mathcal{N}(x^1, 1, 2)\big)$ and zero otherwise is a PDF of some RFS on a real line, its PHD is

$$D_\Xi(x) = \overbrace{p_N(2)}^{1} \cdot 2 \int_\mathcal{X} \frac{1}{2}\big(\mathcal{N}(x, 0, 1)\mathcal{N}(w^1, 1, 2) + \mathcal{N}(w^1, 0, 1)\mathcal{N}(x, 1, 2)\big) dw^1 \tag{A.30a}$$

$$= \mathcal{N}(x, 0, 1)\int_\mathcal{X} \mathcal{N}(w^1, 1, 2) dw^1 + \mathcal{N}(x, 1, 2)\int_\mathcal{X} \mathcal{N}(w^1, 0, 1) dw^1 \tag{A.30b}$$

$$= \mathcal{N}(x, 0, 1) + \mathcal{N}(x, 1, 2) = \underbrace{2}_{\overline{N}} \cdot \underbrace{\left(\frac{1}{2}\mathcal{N}(x, 0, 1) + \frac{1}{2}\mathcal{N}(x, 1, 2)\right)}_{p(x)} \tag{A.30c}$$

In case of having **maximally two objects**, the PHD is

$$D_\Xi(\mathbf{x}) = p_\Xi(\{\mathbf{x}\}) + \int_\mathcal{X} p_\Xi(\{\mathbf{x}, \mathbf{w}^1\}) d\mathbf{w}^1 + \frac{1}{2!}\int_{\mathcal{X}^2} 0 \; d\mathbf{w}^1 d\mathbf{w}^2 + \dots \tag{A.31a}$$

$$= p_N(1) \cdot p_{\mathbf{X}|N}(\mathbf{x}|1) + p_N(2) \cdot 2 \int_\mathcal{X} p_{\mathbf{X}|N}(\mathbf{x}, \mathbf{w}^1|2) d\mathbf{w}^1 \tag{A.31b}$$

$$= p_N(1) \cdot p(\mathbf{x}) + p_N(2) \cdot 2p(\mathbf{x}) = \underbrace{(p_N(1) + 2 \cdot p_N(1))}_{\mathrm{E}_N[N] = \overline{N}} \cdot p(\mathbf{x}) \tag{A.31c}$$

If for example,

$$p_\Xi(X) = \begin{cases} 0.2, & \text{if } X = \emptyset \\ 0.5\,\mathcal{N}(x^1; 2, 3), & \text{if } X = \{x^1\} \\ 0.3\left(\mathcal{N}(x^1, 0, 1)\mathcal{N}(x^2, 1, 2) + \mathcal{N}(x^2, 0, 1)\mathcal{N}(x^1, 1, 2)\right), & \text{if } X = \{x^1, x^2\} \\ 0, & \text{otherwise} \end{cases} \tag{A.32}$$

is a PDF of some RFS on a real line, then, using the above results, its PHD is

$$D_\Xi(x) = \overbrace{p_N(1)}^{0.5} \cdot p_{\mathbf{X}|N}(x|1) + \overbrace{p_N(2)}^{0.3} \cdot 2 \int_{\mathcal{X}} p_{\mathbf{X}|N}(x, w^1|2)dw^1 \tag{A.33a}$$

$$= 0.5\,\mathcal{N}(x; 2, 3) + 0.3\left(\mathcal{N}(x, 0, 1) + \mathcal{N}(x, 1, 2)\right) \tag{A.33b}$$

$$= \underbrace{1.1}_{\overline{N}} \cdot \underbrace{\left(\frac{0.5}{1.1}\mathcal{N}(x; 2, 3) + \frac{0.3}{1.1}\mathcal{N}(x, 0, 1) + \frac{0.3}{1.1}\mathcal{N}(x, 1, 2)\right)}_{p(x)} \tag{A.33c}$$

APPENDIX B

The PHD Filter Derivation Using PGFLs

In section 5.2, a rather intuitive paraphrase of the original derivation of the intensity filter (or iFilter) from [23] was shown. In this appendix, the aim is to derive the classic PHD filter as it was originally done in [38], using probability generating functionals (PGFLs, see Section 2.2.5). The PGFLs find its use in derivation of filters involving moment-based approximations of multi-target densities, such as PHD, CPHD, or PMBM filters. If a reader considers a more detailed treatment of such filters, we recommend reading the following derivation, as it is a simple basis for the more advanced filters.

In the rest of this appendix, the goal is to derive the two formulas

$$D_{k|k-1}(\mathbf{x}_k) = D_B(\mathbf{x}_k) + \int_{\mathcal{X}} \big(P_S(\mathbf{x}_{k-1})\psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) + \\ + D_{Sp}(\mathbf{x}_k|\mathbf{x}_{k-1})\big) D_{k-1|k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1} \tag{B.1}$$

$$D_{k|k}(\mathbf{x}_k) = \left(1 - P_D(\mathbf{x}_k) + \sum_{\mathbf{z}\in Z_k} \frac{l(\mathbf{z}|\mathbf{x}_k)P_D(\mathbf{x}_k)}{\lambda c(\mathbf{z}) + D_{k|k-1}[l(\mathbf{z})P_D]}\right) D_{k|k-1}(\mathbf{x}_k) \tag{B.2}$$

which are the general case of the time-update and (approximate) Bayes-update of the PHD filter, respectively, where $D_B(\mathbf{x}_k) \triangleq N_B \cdot p_B(\mathbf{x}_k)$ is the birth process intensity, $D_{Sp}(\mathbf{x}_k|\mathbf{x}_{k-1})$ is the intensity of spawning, $P_S(\mathbf{x}_k)$ is the state-dependent probability of survival and the rest is as defined in Section 5.2.

B.1 Time-update

For the sake of simplicity, the time indices, and conditioning on measurement will be omitted for the derivation. Let Ξ denote the former posterior RFS, and Ψ the predicted RFS, then

$$D_\Xi(\mathbf{x}) \triangleq D_{k-1|k-1}(\mathbf{x}_{k-1}), \quad D_\Psi(\mathbf{y}) \triangleq D_{k|k-1}(\mathbf{x}_k), \quad \psi(\mathbf{y}|\mathbf{x}) \triangleq \psi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) \tag{B.3}$$

We are now to prove Eq. (B.1), which in notation defined in (B.3) simplifies to

$$D_\Psi(\mathbf{y}) = D_B(\mathbf{y}) + \int_{\mathcal{X}} \big(P_S(\mathbf{x})\psi(\mathbf{y}|\mathbf{x}) + D_{Sp}(\mathbf{y}|\mathbf{x})\big) D_\Xi(\mathbf{x}) d\mathbf{x} \tag{B.4}$$

B.1.1 PGFL of the Predicted RFS

Let $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$ denote a realization of Ξ, and let $Y = \{\mathbf{y}^1, \ldots, \mathbf{y}^m\}$ denote a realization of Ψ. In the simplified notation, the multi-target prediction given by the Chapman-Kolomogorov Eq. (2.25) can be written as

$$p_\Psi(Y) = \int_{\mathcal{X}} p_{\Psi|\Xi}(Y|X) p_\Xi(X) \delta X \tag{B.5}$$

The PGFL of Ψ is then given by

$$\begin{aligned}
G_\Psi[h] &= \int_{\mathcal{X}} h^Y p_\Psi(Y) \delta Y && \text{(B.6a)}\\
&= \int_{\mathcal{X}} h^Y \left(\int_{\mathcal{X}} p_{\Psi|\Xi}(Y|X) p_\Xi(X) \delta X \right) \delta Y && \text{(B.6b)}\\
&= \int_{\mathcal{X}} \underbrace{\left(\int_{\mathcal{X}} h^Y p_{\Psi|\Xi}(Y|X) \delta Y \right)}_{G_{\Psi|\Xi}[h|X] \triangleq} p_\Xi(X) \delta X && \text{(B.6c)}\\
&= \int_{\mathcal{X}} G_{\Psi|\Xi}[h|X] p_\Xi(X) \delta X && \text{(B.6d)}
\end{aligned}$$

where $G_{\Psi|\Xi}[h|X]$ is the conditional PGFL of the transition process. The desired PHD can be recovered from $G_\Psi[h]$ using Eq. (2.39) as, (see section 2.2.8)

$$D_\Psi(\mathbf{y}) = \left[\frac{\delta}{\delta \mathbf{y}} G_\Psi[h] \right]_{h=1} \tag{B.7}$$

In order to take the derivative, we first need to find the formula for $G_\Psi[h]$. For this purpose, we first describe the motion model and find its PGFL.

B.1.2 Motion Model

The transition process (an RFS), is assumed to contain independent random vectors, which are formed according to the model

$$\Psi|\Xi = \underbrace{\left(\bigcup_{\mathbf{x}\in\Xi} \Psi_S|\mathbf{x}\right)}_{\text{survived targets}} \cup \underbrace{\left(\bigcup_{\mathbf{x}\in\Xi} \Psi_{Sp}|\mathbf{x}\right)}_{\text{spawned targets}} \cup \underbrace{\Psi_B}_{\text{born targets}} \tag{B.8}$$

where $\Psi_S|\mathbf{x}$ is a set that is empty, or contains the transitioned target $\mathbf{x}$, $\Psi_{Sp}|\mathbf{x}$ is a set containing targets that were spawned from (generated due to) $\mathbf{x}$, and Ψ_B is a set of the newly born targets that are independent of Ξ. In particular, the following is assumed.

- The RFS $\Psi_S|\mathbf{x}$ is Bernoulli with the PDF, PHD and PGFL, respectively

$$p_{\Psi_S|\mathbf{x}}(Y|\mathbf{x}) = \begin{cases} 1 - P_S(\mathbf{x}), & \text{if } Y = \emptyset \\ P_S(\mathbf{x}) \cdot \psi(\mathbf{y}|\mathbf{x}), & \text{if } Y = \{\mathbf{y}\} \\ 0, & \text{otherwise} \end{cases} \tag{B.9}$$

$$D_S(\mathbf{y}) = P_S(\mathbf{x})\psi(\mathbf{y}|\mathbf{x}) \tag{B.10}$$

$$G_S[h](\mathbf{x}) = 1 - P_S(\mathbf{x}) + P_S(\mathbf{x}) \int_{\mathcal{X}} h(\mathbf{y})\psi(\mathbf{y}|\mathbf{x})d\mathbf{y} \tag{B.11}$$

  where $G_S[h](\mathbf{x})$ is a functional transformation, which for a given $\mathbf{x} \in \mathcal{X}$ is a functional.

- The RFS $\Psi_{Sp}|\mathbf{x}$ has a PHD $D_{Sp}(\mathbf{y}|\mathbf{x})$ and PFGL $G_{Sp}[h](\mathbf{x})$.

- The RFS Ψ_B has a PHD $D_B(\mathbf{y})$ and PFGL $G_B[h]$.

The PFGL of the transition process $\Psi|\Xi$ is then given by the product formula (2.34),

$$\begin{aligned} G_{\Psi|\Xi}[h|X] &= \left(\prod_{\mathbf{x}\in X} G_S[h](\mathbf{x})\right)\left(\prod_{\mathbf{x}\in X} G_{Sp}[h](\mathbf{x})\right) G_B[h] && \text{(B.12a)} \\ &= G_S[h]^X G_{Sp}[h]^X G_B[h] && \text{(B.12b)} \\ &= \left(G_S[h]G_{Sp}[h]\right)^X G_B[h] && \text{(B.12c)} \end{aligned}$$

where the exponential notation from (2.32) was used. It is now useful to define the functional transformation $\Phi[h](\mathbf{x}) \triangleq G_S[h](\mathbf{x})G_{Sp}[h](\mathbf{x})$.

B.1.3 Deriving the Time-updated PHD

Now, we can turn to the Eq. (B.7) and the derivative therein. Substitute the conditional PFGL of the transition process (B.12c) into the PGFL of the predicted process (B.6d),

$$
\begin{aligned}
G_{\Psi}[h] &= \int_{\mathcal{X}} G_{\Psi|\Xi}[h|X] p_{\Xi}(X)\delta X && \text{(B.13a)}\\
&= \int_{\mathcal{X}} \Phi[h]^{X} G_{B}[h] p_{\Xi}(X)\delta X && \text{(B.13b)}\\
&= G_{B}[h] \int_{\mathcal{X}} \Phi[h]^{X} p_{\Xi}(X)\delta X && \text{(B.13c)}\\
&= G_{B}[h] G_{\Xi}\big[\Phi[h]\big] && \text{(B.13d)}
\end{aligned}
$$

Now, the task is to take the functional derivative of (B.13d) and set $h = 1$. For the following, note that for any PFGL, $G[1] = \int_{\mathcal{X}} 1^{X} p(X)\delta X = 1$. Substituting (B.13d) into (B.7) results into

$$
\begin{aligned}
D_{\Psi}(\mathbf{y}) &= \left[\frac{\delta}{\delta \mathbf{y}} G_{\Psi}[h]\right]_{h=1} && \text{(B.14a)}\\
&= \left[\frac{\delta}{\delta \mathbf{y}} \big(G_{B}[h] G_{\Xi}\big[\Phi[h]\big]\big)\right]_{h=1} && \text{(B.14b)}\\
&= \left[\left(\frac{\delta}{\delta \mathbf{y}} G_{B}[h]\right) G_{\Xi}\big[\Phi[h]\big] + G_{B}[h]\left(\frac{\delta}{\delta \mathbf{y}} G_{\Xi}\big[\Phi[h]\big]\right)\right]_{h=1} && \text{(B.14c)}\\
&= \underbrace{\left[\frac{\delta}{\delta \mathbf{y}} G_{B}[h]\right]_{h=1}}_{D_{B}(\mathbf{y})} \underbrace{G_{\Xi}\big[\Phi[1]\big]}_{1} + \underbrace{G_{B}[1]}_{1} \left[\frac{\delta}{\delta \mathbf{y}} G_{\Xi}\big[\Phi[h]\big]\right]_{h=1} && \text{(B.14d)}\\
&= D_{B}(\mathbf{y}) + \left[\frac{\delta}{\delta \mathbf{y}} G_{\Xi}\big[\Phi[h]\big]\right]_{h=1} && \text{(B.14e)}
\end{aligned}
$$

Noting that a PGFL can be rewritten as an expectation, the second term in (B.14e) can be rewritten using "linearity" (Leibniz integration rule) of the expectation as

$$D_\Psi(\mathbf{y}) = \ldots = D_B(\mathbf{y}) + \left[\frac{\delta}{\delta \mathbf{y}} \mathrm{E}_\Xi \left[\Phi[h]^\Xi\right]\right]_{h=1} \tag{B.15a}$$

$$= D_B(\mathbf{y}) + \mathrm{E}_\Xi \left[\left[\frac{\delta}{\delta \mathbf{y}} \Phi[h]^\Xi\right]_{h=1} \right] \tag{B.15b}$$

$$= D_B(\mathbf{y}) + \int_{\mathcal{X}} \left[\frac{\delta}{\delta \mathbf{y}} \prod_{\mathbf{x}\in X} \Phi[h](\mathbf{x})\right]_{h=1} p_\Xi(X)\delta X \tag{B.15c}$$

using the product rule for derivatives we get

$$D_\Psi(\mathbf{y}) = \ldots = D_B(\mathbf{y}) + \int_{\mathcal{X}} \left[\sum_{\mathbf{x}\in X} \left(\frac{\delta}{\delta \mathbf{y}} \Phi[h](\mathbf{x})\right) \prod_{\substack{\mathbf{w}\in X\\ \mathbf{w}\neq \mathbf{x}}} \Phi[h](\mathbf{w})\right]_{h=1} p_\Xi(X)\delta X \tag{B.16a}$$

$$= D_B(\mathbf{y}) + \int_{\mathcal{X}} \sum_{\mathbf{x}\in X} \left[\frac{\delta}{\delta \mathbf{y}} \Phi[h](\mathbf{x})\right]_{h=1} p_\Xi(X)\delta X \tag{B.16b}$$

For every fixed $\mathbf{x}$ and for any valid h, the term $\Phi[h](\mathbf{x})$ is a PGFL of the union of survived and spawned targets. Therefore, by the recovery of PHDs using PGFLs Eq. (2.39), we get

$$D_{S\cup Sp}(\mathbf{y}|\mathbf{x}) \triangleq \left[\frac{\delta}{\delta \mathbf{y}} \Phi[h](\mathbf{x})\right]_{h=1} \tag{B.17a}$$

$$= \left[\frac{\delta}{\delta \mathbf{y}} \left(G_S[h](\mathbf{x}) G_{Sp}[h](\mathbf{x})\right)\right]_{h=1} \tag{B.17b}$$

$$= \left[\left(\frac{\delta}{\delta \mathbf{y}} G_S[h](\mathbf{x})\right) \left(G_{Sp}[h](\mathbf{x})\right) + \left(G_S[h](\mathbf{x})\right) \left(\frac{\delta}{\delta \mathbf{y}} G_{Sp}[h](\mathbf{x})\right)\right]_{h=1} \tag{B.17c}$$

$$= \underbrace{\left[\frac{\delta}{\delta \mathbf{y}} G_S[h](\mathbf{x})\right]_{h=1}}_{D_S(\mathbf{y}|\mathbf{x})} \underbrace{G_{Sp}[1](\mathbf{x})}_{1} + \underbrace{G_S[1](\mathbf{x})}_{1} \underbrace{\left[\frac{\delta}{\delta \mathbf{y}} G_{Sp}[h](\mathbf{x})\right]_{h=1}}_{D_{Sp}(\mathbf{y}|\mathbf{x})} \tag{B.17d}$$

$$= P_S(\mathbf{x})\psi(\mathbf{y}|\mathbf{x}) + D_{Sp}(\mathbf{y}|\mathbf{x}) \tag{B.17e}$$

The second term in Eq. (B.16b) could be viewed as an expectation of the co-called *random sum*[1] that simplify dramatically. Assume the PHD $D_\Xi(\mathbf{x})$ of Ξ is known, then

$$D_\Psi(\mathbf{y}) = \cdots = D_B(\mathbf{y}) + \int_{\mathcal{X}} \sum_{\mathbf{x}\in X} D_{S\cup Sp}(\mathbf{y}|\mathbf{x}) p_\Xi(X)\delta X \tag{B.18a}$$

$$= D_B(\mathbf{y}) + \sum_{n=1}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} \big(D_{S\cup Sp}(\mathbf{y}|\mathbf{x}^1) + \cdots + D_{S\cup Sp}(\mathbf{y}|\mathbf{x}^n)\big) p_\Xi(\{\mathbf{x}^1,\ldots,\mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n \tag{B.18b}$$

$$= D_B(\mathbf{y}) + \sum_{n=1}^{+\infty} \frac{1}{n!} \sum_{i=1}^{n} \int_{\mathcal{X}^n} D_{S\cup Sp}(\mathbf{y}|\mathbf{x}^i) p_\Xi(\{\mathbf{x}^1,\ldots,\mathbf{x}^i,\ldots,\mathbf{x}^n\}) d\mathbf{x}^1 \cdots d\mathbf{x}^i \cdots d\mathbf{x}^n \tag{B.18c}$$

$$= D_B(\mathbf{y}) + \sum_{n=1}^{+\infty} \frac{\cancel{n}}{\cancel{n}(n-1)!} \int_{\mathcal{X}^n} D_{S\cup Sp}(\mathbf{y}|\mathbf{w}) p_\Xi(\{\mathbf{x}^1,\ldots,\mathbf{x}^{n-1},\mathbf{w}\}) d\mathbf{x}^1 \cdots d\mathbf{x}^{n-1} d\mathbf{w} \tag{B.18d}$$

$$= D_B(\mathbf{y}) + \int_{\mathcal{X}} D_{S\cup Sp}(\mathbf{y}|\mathbf{w}) \left(\sum_{n=1}^{+\infty} \frac{1}{(n-1)!} \int_{\mathcal{X}^{n-1}} p_\Xi(\{\mathbf{x}^1,\ldots,\mathbf{x}^{n-1},\mathbf{w}\}) d\mathbf{x}^1 \cdots d\mathbf{x}^{n-1} \right) d\mathbf{w} \tag{B.18e}$$

$$= D_B(\mathbf{y}) + \int_{\mathcal{X}} D_{S\cup Sp}(\mathbf{y}|\mathbf{w}) \underbrace{\left(\sum_{n=0}^{+\infty} \frac{1}{n!} \int_{\mathcal{X}^n} p_\Xi(\{\mathbf{x}^1,\ldots,\mathbf{x}^n\} \cup \{\mathbf{w}\}) d\mathbf{x}^1 \cdots d\mathbf{x}^n \right)}_{D_\Xi(\mathbf{w})} d\mathbf{w} \tag{B.18f}$$

$$= D_B(\mathbf{y}) + \int_{\mathcal{X}} D_{S\cup Sp}(\mathbf{y}|\mathbf{w}) D_\Xi(\mathbf{w}) d\mathbf{w} \tag{B.18g}$$

Finally, substituting (B.17e) into (B.18g) yields the final expression

$$D_\Psi(\mathbf{y}) = \ldots = D_B(\mathbf{y}) + \int_{\mathcal{X}} \big(P_S(\mathbf{x})\psi(\mathbf{y}|\mathbf{x}) + D_{Sp}(\mathbf{y}|\mathbf{x})\big) D_\Xi(\mathbf{w}) d\mathbf{w} \tag{B.19}$$

Rewriting this expression with appropriate symbols defined in (B.3) yield the desired time-update equation (B.1), which finishes the proof. Note, that no approximations were needed during the derivation, an thus the time-update of PHD hold generally, for the given motion model, for any posterior distribution. In case of no spawning and constant P_S, we can observe that the result is equivalent to that of the intensity filter (Section 5.2), where it was explicitly assumed, that the former posterior is a PPP on the augmented space $\mathcal{X}^A$.

[1] Note, that in the following set integral, X must be at least singleton (and so $n > 0$), since $\sum_{\mathbf{x}\in X} D_{S\cup Sp}(\mathbf{y}|\mathbf{x})$ for $X = \emptyset$ is undefined.

B.2 Bayes-update

Similarly to the time-update, time indices will be omitted for the derivation, introducing Ψ denoting the predicted RFS, Ξ denoting the posterior RFS, and Σ denoting the newly-arriving measurement RFS thus

$$D_\Xi(\mathbf{x}) := D_{k|k}(\mathbf{x}_k), \quad D_\Psi(\mathbf{x}) := D_{k|k-1}(\mathbf{x}_k), \quad Z = Z_k, \quad l(\mathbf{z}|\mathbf{x}) = l(\mathbf{z}_k|\mathbf{x}_k) \tag{B.20}$$

Explicit conditioning on the whole measurement sequence will be omitted for simplicity. We are now to prove Eq. (B.2), which in notation defined in (B.20) simplifies to

$$D_\Xi(\mathbf{x}) = \left(1 - P_D(\mathbf{x}) + \sum_{\mathbf{z}\in Z} \frac{l(\mathbf{z}|\mathbf{x})P_D(\mathbf{x})}{\lambda c(\mathbf{z}) + D_\Psi[l(\mathbf{z})P_D]}\right) D_\Psi(\mathbf{x}) \tag{B.21}$$

with $l(\mathbf{z}) \triangleq l(\mathbf{z}|\cdot)$ suppressing the entry vector $\mathbf{x}$ to denote the entire function.

B.2.1 PFGL of the Posterior RFS

In the simplified notation, the multi-target measurement-update given by the Bayes rule (2.26), can be written as

$$p_\Xi(X) = \frac{l(Z|X)p_\Psi(X)}{\int_\mathcal{X} l(Z|X)p_\Psi(X)\delta X} = \frac{l(Z|X)p_\Psi(X)}{p_\Sigma(Z)} \tag{B.22}$$

The PGFL of Ξ is then given by

$$G_\Xi[h] = \int_\mathcal{X} h^X p_\Xi(X)\delta X = \frac{\int_\mathcal{X} h^X l(Z|X)p_\Psi(X)\delta X}{p_\Sigma(Z)} \tag{B.23}$$

Similarly to the time-update, the desired PHD can be recovered from $G_\Xi[h]$ as

$$D_\Xi(\mathbf{x}) = \left[\frac{\delta}{\delta\mathbf{x}} G_\Xi[h]\right]_{h=1} = \frac{1}{p_\Sigma(Z)} \left[\frac{\delta}{\delta\mathbf{x}} \int_\mathcal{X} h^X l(Z|X)p_\Psi(X)\delta X\right]_{h=1} \tag{B.24}$$

The expression is naturally different than the one of the time-update. The Bayes-update is harder to derive. We have to determine an expression for both the numerator, and the denominator. To continue, we have to determine the measurement model.

B.2.2 Measurement Model

The measurement process (an RFS), is assumed to contain independent random vectors, which are formed according to the model

$$\Sigma|\Xi = \underbrace{\left(\bigcup_{\mathbf{x}\in\Xi} O|\mathbf{x}\right)}_{\text{object-generated}} \cup \underbrace{C}_{\text{clutter}} \tag{B.25}$$

where $O|\mathbf{x}$ is a set that is either empty, or it contains a measurement generated by the target $\mathbf{x}$, and C is a set containing the clutter. In particular, the following is assumed.

- The RFS $O|\mathbf{x}$ is Bernoulli with the PDF, PHD, and PGFL, respectively

$$p_{O|\mathbf{x}}(Z|\mathbf{x}) = \begin{cases} 1 - P_D(\mathbf{x}), & \text{if } X = \emptyset \\ P_D(\mathbf{x}) \cdot l(\mathbf{z}|\mathbf{x}), & \text{if } X = \{\mathbf{x}\} \\ 0, & \text{otherwise} \end{cases} \tag{B.26}$$

$$D_O(\mathbf{y}) = P_D(\mathbf{x})l(\mathbf{z}|\mathbf{x}) \tag{B.27}$$

$$G_O[g](\mathbf{x}) = 1 - P_D(\mathbf{x}) + P_D(\mathbf{x}) \int_{\mathcal{X}} g(\mathbf{z})l(\mathbf{z}|\mathbf{x})d\mathbf{z} = 1 - P_D(\mathbf{x}) + P_D(\mathbf{x})l[g](\mathbf{x}) \tag{B.28}$$

  where $l[g](\mathbf{x}) \triangleq \int_{\mathcal{X}} g(\mathbf{z})l(\mathbf{z}|\mathbf{x})d\mathbf{z}$ is a functional transformation.

- The RFS C is Poisson with the PDF, PHD, and PFGL, respectively

$$p_C(Z) = e^{-\lambda} \prod_{\mathbf{z}\in Z} \lambda c(\mathbf{z}) \tag{B.29}$$

$$D_C(\mathbf{z}) = \lambda c(\mathbf{z}) \tag{B.30}$$

$$G_C[g] = \exp\left(\int_{\mathcal{X}} g(\mathbf{z})\lambda c(\mathbf{z})d\mathbf{z} - \lambda\right) = e^{\lambda c[g]-\lambda} \tag{B.31}$$

  where $c[g] \triangleq \int_{\mathcal{X}} g(\mathbf{z})c(\mathbf{z})d\mathbf{z}$ is a functional transformation.

The PFGL of the measurement process $\Psi|\Xi$ is then given by the product formula (2.34),

$$G_{\Sigma|\Xi}[g|X] = \left(\prod_{\mathbf{x}\in X} G_O[g](\mathbf{x})\right) G_C[g] = G_O[g]^X G_C[g] \tag{B.32a}$$

B.2.3 Deriving the Bayes-updated PHD

Now we can turn back to (B.24). For tractability, define the two-variable PGFL as

$$F[g,h] \triangleq \int_{\mathcal{X}} \int_{\mathcal{X}} h^X g^Z l(Z|X) p_\Psi(X) \delta X \delta Z \tag{B.33}$$

Both the numerator and denominator of (B.24) can be expressed in terms of this PGFL. Since the constant $p_\Sigma(Z)$ is a PDF of the multi-target measurement process (evaluated at the arrived measurement-set Z), it can be recovered from its PGFL $G_\Sigma(Z)$ using (2.38) as

$$p_\Sigma(Z) = \left[\frac{\delta}{\delta Z} G_\Sigma[g] \right]_{g=0} \tag{B.34a}$$

$$= \left[\frac{\delta}{\delta Z} \int_{\mathcal{X}} g^Z p_\Sigma(Z) \delta Z \right]_{g=0} \tag{B.34b}$$

$$= \left[\frac{\delta}{\delta Z} \int_{\mathcal{X}} g^Z \int_{\mathcal{X}} l(Z|X) p_\Psi(X) \delta X \delta Z \right]_{g=0} \tag{B.34c}$$

$$= \left[\frac{\delta}{\delta Z} \int_{\mathcal{X}} \int_{\mathcal{X}} 1^X g^Z l(Z|X) p_\Psi(X) \delta X \delta Z \right]_{g=0} \tag{B.34d}$$

$$= \left[\frac{\delta^m}{\delta \mathbf{z}^m \ldots \delta \mathbf{z}^1} F[g,1] \right]_{g=0} \tag{B.34e}$$

The numerator of (B.24) can be expanded, noting that the multi-target PDF of the measurement process (measurement likelihood) $l(Z|X)$ can be recovered from its PGFL $G_{\Sigma|\Xi}[g|X]$. The expression for the desired PHD then becomes

$$D_\Xi(\mathbf{x}) = \ldots = \frac{1}{p_\Sigma(Z)} \left[\frac{\delta}{\delta \mathbf{x}} \int_{\mathcal{X}} h^X l(Z|X) p_\Psi(X) \delta X \right]_{h=1} \tag{B.35a}$$

$$= \frac{1}{p_\Sigma(Z)} \left[\frac{\delta}{\delta \mathbf{x}} \int_{\mathcal{X}} h^X \left[\frac{\delta}{\delta Z} G_{\Sigma|\Xi}[g|X] \right]_{g=0} p_\Psi(X) \delta X \right]_{h=1} \tag{B.35b}$$

$$= \frac{1}{p_\Sigma(Z)} \left[\frac{\delta}{\delta Z} \left[\frac{\delta}{\delta \mathbf{x}} \int_{\mathcal{X}} h^X G_{\Sigma|\Xi}[g|X] p_\Psi(X) \delta X \right]_{h=1} \right]_{g=0} \tag{B.35c}$$

$$= \frac{1}{p_\Sigma(Z)} \left[\frac{\delta}{\delta Z} \left[\frac{\delta}{\delta \mathbf{x}} \int_{\mathcal{X}} h^X \left(\int_{\mathcal{X}} g^Z l(Z|X) \delta Z \right) p_\Psi(X) \delta X \right]_{h=1} \right]_{g=0} \tag{B.35d}$$

$$= \frac{1}{p_\Sigma(Z)} \left[\frac{\delta^m}{\delta \mathbf{z}^m \ldots \delta \mathbf{z}^1} \left[\frac{\delta}{\delta \mathbf{x}} F[g,h] \right]_{h=1} \right]_{g=0} \tag{B.35e}$$

In the literature, this relation can be found in a simplified form, expressed as

$$D_{\Xi}(\mathbf{x}) = \cdots = \left(\left[\frac{\delta}{\delta Z}F[g,1]\right]_{g=0}\right)^{-1}\left[\frac{\delta}{\delta Z}\left[\frac{\delta}{\delta \mathbf{x}}F[g,h]\right]_{h=1}\right]_{g=0} = \frac{\frac{\delta^{m+1}F}{\delta \mathbf{z}^m \ldots \delta \mathbf{z}^1 \delta \mathbf{x}}[0,1]}{\frac{\delta^m F}{\delta \mathbf{z}^m \ldots \delta \mathbf{z}^1}[0,1]} \tag{B.36}$$

To continue, expand the two-variable PGFL,

$$F[g,h] = \int_{\mathcal{X}}\int_{\mathcal{X}} h^X g^Z l(Z|X) p_{\Psi}(X) \delta X \delta Z \tag{B.37a}$$

$$= \int_{\mathcal{X}} h^X \underbrace{\left(\int_{\mathcal{X}} g^Z l(Z|X)\delta Z\right)}_{G_{\Sigma|\Xi}[g|X]} p_{\Psi}(X)\delta X \tag{B.37b}$$

$$= \int_{\mathcal{X}} h^X \big(G_O[g]^X G_C[g]\big) p_{\Psi}(X)\delta X \tag{B.37c}$$

$$= G_C[g] \underbrace{\int_{\mathcal{X}} (h \cdot G_O[g])^X \, p_{\Psi}(X)\delta X}_{G_{\Xi}\left[h\cdot G_O[g]\right]} \tag{B.37d}$$

$$= G_C[g] G_{\Psi}\big[h \cdot G_O[g]\big] \tag{B.37e}$$

In order to derive a closed-form expression, assume that Ψ is Poisson RFS. Since we know its PHD, we can form its PDF, and especially PGFL as

$$p_{\Psi}(X) \approx e^{-\overline{N}_{\Psi}} \prod_{\mathbf{x}\in X} D_{\Psi}(\mathbf{x}), \qquad \text{with} \qquad \overline{N}_{\Psi} = \int_{\mathcal{X}} D_{\Psi}(\mathbf{x}) d\mathbf{x} \tag{B.38}$$

$$G_{\Psi}[h] \approx e^{D_{\Psi}[h]-\overline{N}_{\Psi}}, \qquad \text{with} \qquad D_{\Psi}[h] = \int_{\mathcal{X}} h(\mathbf{x}) D_{\Psi}(\mathbf{x}) d\mathbf{x} \tag{B.39}$$

Then $F[g,h]$ simplifies into

$$F[g,h] = \ldots = G_C[g] G_{\Psi}\big[h \cdot G_O[g]\big] \tag{B.40a}$$

$$= e^{\lambda c[g]-\lambda} e^{D_{\Psi}\left[h\cdot G_O[g]\right]-\overline{N}_{\Psi}} \tag{B.40b}$$

$$= \exp\big(\lambda c[g] - \lambda + D_{\Psi}\big[h \cdot \big(1 - P_D + P_D l[g]\big)\big] - \overline{N}_{\Psi}\big) \tag{B.40c}$$

Now, we can reveal the expression for the PHD (B.36) step by step.

The Denominator

First, set $h = 1$,

$$F[g,1] = \exp\left(\lambda c[g] - \lambda + D_\Psi\left[1 - P_D + P_D l[g]\right] - \overline{N}_\Psi\right) \tag{B.41a}$$

$$= \exp\left(\lambda c[g] - \lambda + \cancel{D_\Psi[1]} - D_\Psi[P_D] + D_\Psi\left[P_D l[g]\right] - \cancel{\overline{N}_\Psi}\right) \tag{B.41b}$$

$$= \exp\left(\lambda c[g] - \lambda - D_\Psi[P_D] + D_\Psi\left[P_D l[g]\right]\right) \tag{B.41c}$$

Second, take the functional derivative

$$\frac{\delta^m}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^1} F[g,1] = \frac{\delta^{m-1}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^2}\left(\frac{\delta}{\delta \mathbf{z}^1} F[g,1]\right) \tag{B.42a}$$

$$= \frac{\delta^{m-1}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^2}\left(\frac{\delta}{\delta \mathbf{z}^1} \exp\left(\lambda c[g] - \lambda - D_\Psi[P_D] + D_\Psi\left[P_D l[g]\right]\right)\right) \tag{B.42b}$$

$$\left|\begin{array}{c}\text{derivative of}\\\text{exponential}\\\text{functional}\end{array}\right| \quad = \frac{\delta^{m-1}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^2}\left(F[g,1] \cdot \frac{\delta}{\delta \mathbf{z}^1}\left(\lambda c[g] - \lambda - D_\Psi[P_D] + D_\Psi\left[P_D l[g]\right]\right)\right) \tag{B.42c}$$

$$= \frac{\delta^{m-1}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^2}\left(F[g,1] \cdot \left(\lambda c(\mathbf{z}^1) + D_\Psi\left[P_D l(\mathbf{z}^1)\right]\right)\right) \tag{B.42d}$$

$$= \frac{\delta^{m-2}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^3}\left(\frac{\delta}{\delta \mathbf{z}^2}\left[F[g,1] \cdot \left(\lambda c(\mathbf{z}^1) + D_\Psi\left[P_D l(\mathbf{z}^1)\right]\right)\right]\right) \tag{B.42e}$$

$$= \frac{\delta^{m-2}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^3}\Bigg(\overbrace{\left[\frac{\delta}{\delta \mathbf{z}^2} F[g,1]\right]}^{\text{analogous as for } \mathbf{z}^1}\left(\lambda c(\mathbf{z}^1) + D_\Psi\left[P_D l(\mathbf{z}^1)\right]\right) + \\ + F[g,1]\underbrace{\left[\frac{\delta}{\delta \mathbf{z}^2}\left(\lambda c(\mathbf{z}^1) + D_\Psi\left[P_D l(\mathbf{z}^1)\right]\right)\right]}_{0}\Bigg) \tag{B.42f}$$

$$= \frac{\delta^{m-2}}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^3}\left(F[g,1] \prod_{\mathbf{z} \in \{\mathbf{z}^1, \mathbf{z}^2\}}\left(\lambda c(\mathbf{z}) + D_\Psi\left[P_D l(\mathbf{z})\right]\right)\right) \tag{B.42g}$$

$$\left|\begin{array}{c}\text{by induction}\\= \dots =\end{array}\right| \quad = F[g,1] \prod_{\mathbf{z} \in Z}\left(\lambda c(\mathbf{z}) + D_\Psi\left[P_D l(\mathbf{z})\right]\right) \tag{B.42h}$$

Finally, set $g = 0$, (compare this result with the PDF (5.14c) from Section 5.2)

$$p_\Sigma(Z) = \left[\frac{\delta^m}{\delta \mathbf{z}^m \dots \delta \mathbf{z}^1} F[g,1]\right]_{g=0} = F[0,1] \prod_{\mathbf{z} \in Z}\left(\lambda c(\mathbf{z}) + D_\Psi\left[P_D l(\mathbf{z})\right]\right) \tag{B.43a}$$

$$= e^{-\lambda - D_\Psi[P_D]} \prod_{\mathbf{z} \in Z}\left(\lambda c(\mathbf{z}) + D_\Psi\left[P_D l(\mathbf{z})\right]\right) \tag{B.43b}$$

The Numerator

First, take the functional derivative of $F[g,h]$ assuming g is constant,

$$\frac{\delta}{\delta \mathbf{x}} F[g,h] = F[g,h] \frac{\delta}{\delta \mathbf{x}} \left(\lambda c[g] - \lambda + D_\Psi \left[h \cdot \left(1 - P_D + P_D l[g] \right) \right] - \overline{N}_\Psi \right) \tag{B.44a}$$

$$= F[g,h] \left(0 - 0 + \frac{\delta}{\delta \mathbf{x}} \int_{\mathcal{X}} h(\mathbf{x}) \big(1 - P_D(\mathbf{x}) + P_D(\mathbf{x}) l[g](\mathbf{x}) \big) D_\Psi(\mathbf{x}) d\mathbf{x} - 0 \right) \tag{B.44b}$$

$$= F[g,h] \underbrace{\big(1 - P_D(\mathbf{x}) + P_D(\mathbf{x}) l[g](\mathbf{x}) \big) D_\Psi(\mathbf{x})}_{f[g](\mathbf{x}) \triangleq} \tag{B.44c}$$

$$= F[g,h] f[g](\mathbf{x}) \tag{B.44d}$$

where $f[g](\mathbf{x})$ is a functional transformation. Second, set $h = 1$ and take the derivative with respect to Z. Using the product rule (for functional derivatives [4, pp. 389]), we get

$$\frac{\delta}{\delta Z} \left[\frac{\delta}{\delta \mathbf{x}} F[g,h] \right]_{h=1} = \frac{\delta}{\delta Z} \Big(f[g](\mathbf{x}) \cdot F[g,1] \Big) = \sum_{W \subseteq Z} \frac{\delta f[g](\mathbf{x})}{\delta W} \cdot \frac{\delta F[g,1]}{\delta (Z \setminus W)} \tag{B.45}$$

Take a look at the expression $\frac{\delta f[g](\mathbf{x})}{\delta W}$. One can deduce that

a) if $\quad W = \emptyset \qquad \Rightarrow \qquad \dfrac{\delta f[g](\mathbf{x})}{\delta \emptyset} = f[g](\mathbf{x}) \qquad$ (B.46)

b) if $\quad W = \{\mathbf{z}^i\},\ i \in (1, \ldots, m) \qquad \Rightarrow \qquad \dfrac{\delta f[g](\mathbf{x})}{\delta \mathbf{z}^i} = 0 + P_D(\mathbf{x}) \left(\dfrac{\delta}{\delta \mathbf{z}^i} l[g](\mathbf{x}) \right) D_\Psi(\mathbf{x})$

$$= P_D(\mathbf{x}) l(\mathbf{z}^i | \mathbf{x}) D_\Psi(\mathbf{x}) \tag{B.47}$$

c) if $\quad W = \{\mathbf{z}^i, \mathbf{z}^j\},\ i \neq j \in (1, \ldots, m) \qquad \Rightarrow \qquad \dfrac{\delta f[g](\mathbf{x})}{\delta \mathbf{z}^j \delta \mathbf{z}^i} = \dfrac{\delta}{\delta \mathbf{z}^j} \left(\dfrac{\delta f[g](\mathbf{x})}{\delta \mathbf{z}^i} \right) = 0 \qquad$ (B.48)

and so if $\quad |W| \geq 2 \qquad \Rightarrow \qquad \dfrac{\delta f[g](\mathbf{x})}{\delta W} = 0 \qquad$ (B.49)

The summation in (B.45) then simplifies into

$$\frac{\delta}{\delta Z} \left[\frac{\delta}{\delta \mathbf{x}} F[g,h] \right]_{h=1} = \ldots = \frac{\delta f[g](\mathbf{x})}{\delta \emptyset} \cdot \frac{\delta F[g,1]}{\delta Z} + \sum_{\{\mathbf{z}\} \subseteq Z} \frac{\delta f[g](\mathbf{x})}{\delta \{\mathbf{z}\}} \cdot \frac{\delta F[g,1]}{\delta (Z \setminus \{\mathbf{z}\})} + 0 \tag{B.50}$$

$$= f[g](\mathbf{x}) \cdot \underbrace{\frac{\delta F[g,1]}{\delta Z}}_{\text{derived before}} + \sum_{\mathbf{z} \in Z} \frac{\delta f[g](\mathbf{x})}{\delta \mathbf{z}} \cdot \frac{\delta F[g,1]}{\delta (Z \setminus \{\mathbf{z}\})} \tag{B.51}$$

The term $\frac{\delta F[g,1]}{\delta(Z\setminus\{\mathbf{z}\})}$ can be determined from the already derived $\frac{\delta F[g,1]}{\delta Z}$, Eq. (B.42h). One can deduce it can be expressed as

$$\frac{\delta F[g,1]}{\delta(Z\setminus\{\mathbf{z}\})} = \frac{F[g,1]\prod_{\mathbf{w}\in Z}\big(\lambda c(\mathbf{w}) + D_\Psi\big[P_D l(\mathbf{w})\big]\big)}{\lambda c(\mathbf{z}) + D_\Psi\big[P_D l(\mathbf{z})\big]} \tag{B.52}$$

Therefore, Eq. (B.51) becomes

$$\begin{aligned}\frac{\delta}{\delta Z}\left[\frac{\delta}{\delta \mathbf{x}}F[g,h]\right]_{h=1} &= \overbrace{\big(1 - P_D(\mathbf{x}) + P_D(\mathbf{x})l[g](\mathbf{x})\big)D_\Psi(\mathbf{x})}^{f[g](\mathbf{x})}\overbrace{F[g,1]\prod_{\mathbf{w}\in Z}\big(\lambda c(\mathbf{w}) + D_\Psi\big[P_D l(\mathbf{w})\big]\big)}^{\frac{\delta F[g,1]}{\delta Z}} + \\ &\quad + \sum_{\mathbf{z}\in Z}\underbrace{P_D(\mathbf{x})l(\mathbf{z}|\mathbf{x})D_\Psi(\mathbf{x})}_{\frac{\delta f[g](\mathbf{x})}{\delta \mathbf{z}}}\underbrace{\frac{F[g,1]\prod_{\mathbf{w}\in Z}\big(\lambda c(\mathbf{w}) + D_\Psi\big[P_D l(\mathbf{w})\big]\big)}{\lambda c(\mathbf{z}) + D_\Psi\big[P_D l(\mathbf{z})\big]}}_{\frac{\delta F[g,1]}{\delta(Z\setminus\{\mathbf{z}\})}} && \text{(B.53)}\\ &= F[g,1]\prod_{\mathbf{w}\in Z}\big(\lambda c(\mathbf{w}) + D_\Psi\big[P_D l(\mathbf{w})\big]\big)\, D_\Psi(\mathbf{x})\times && \text{(B.54)}\\ &\quad\times\left(1 - P_D(\mathbf{x}) + P_D(\mathbf{x})l[g](\mathbf{x}) + \sum_{\mathbf{z}\in Z}\frac{P_D(\mathbf{x})l(\mathbf{z}|\mathbf{x})}{\lambda c(\mathbf{z}) + D_\Psi[P_D l(z)]}\right) && \text{(B.55)}\end{aligned}$$

Finally, set $g = 0$. Noting that $p_\Sigma(Z) = F[0,1]\prod_{\mathbf{z}\in Z}\big(\lambda c(\mathbf{z}) + D_\Psi\big[P_D l(\mathbf{z})\big]\big)$, we can simplify the expression as

$$\left[\frac{\delta}{\delta Z}\left[\frac{\delta}{\delta \mathbf{x}}F[g,h]\right]_{h=1}\right]_{g=0} = p_\Sigma(Z)D_\Psi(\mathbf{x})\left(1 - P_D(\mathbf{x}) + 0 + \sum_{\mathbf{z}\in Z}\frac{P_D(\mathbf{x})l(\mathbf{z}|\mathbf{x})}{\lambda c(\mathbf{z}) + D_\Psi[P_D l(z)]}\right) \tag{B.56a}$$

The Final Bayes-update

Now, substitute the denominator (B.43b) and numerator (B.56a) into (B.36) to get

$$\begin{aligned}D_\Xi(\mathbf{x}) = \ldots &= \left(\left[\frac{\delta}{\delta Z}F[g,1]\right]_{g=0}\right)^{-1}\left[\frac{\delta}{\delta Z}\left[\frac{\delta}{\delta \mathbf{x}}F[g,h]\right]_{h=1}\right]_{g=0} && \text{(B.57)}\\ &= \left(1 - P_D(\mathbf{x}) + \sum_{\mathbf{z}\in Z}\frac{P_D(\mathbf{x})l(\mathbf{z}|\mathbf{x})}{\lambda c(\mathbf{z}) + D_\Psi[P_D l(z)]}\right)D_\Psi(\mathbf{x}) && \text{(B.58)}\end{aligned}$$

Rewriting this expression with appropriate symbols defined in (B.20) yields the desired Bayes-update equation (B.2), which finishes the proof. Note, that Ψ was assumed to be Poisson, but Ξ was not. Its true distribution is like (5.17d), and stays for the time-update.

Bibliography

[1] Y. Bar-Shalom, P. K. Willet, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms.* YBS Publishing, 2011.

[2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems.* Artech House, 1999.

[3] S. Challa, M. R. Morelande, D. Mušicki, and R. J. Evans, *Fundamentals of Object Tracking.* Cambridge University Press, 2011.

[4] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion.* Artech House, 2007.

[5] B.-N. Vo, M. Mallick, Y. Bar-Sshalom, S. Coraluppi, R. Osborne III, R. Mahler, and B.-T. Vo, *Multi-target Tracking*, pp. 1–15. Wiley Encyclopedia of Electrical and Electronics Engineering, 2015.

[6] S. Särkkä, *Bayesian Filtering and Smoothing.* Institute of Mathematical Statistics Textbooks, Cambridge University Press, 2013.

[7] R. E. Kalman, "A new approach fo linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering*, vol. 82, pp. 35–45, March 1960.

[8] A. H. Jazwinski, "Stochastic processes and filtering theory," *Dover, New York*, 126-139.

[9] H. Tanizaki, *Nonlinear Filters: Estimation and Applications (2nd ed.).* Springer-Verlag, 1996.

[10] H. Sorenson and D. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465 – 479, 1971.

[11] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, pp. 477–482, March 2000.

[12] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, 2000.

[13] J. Matoušek, J. Duník, and O. Straka, "Point-mass filter: Density specific grid design and implementation," *Preprints of the 15th European Workshop on Advanced Control and Diagnosis, Bologna, Italy*, 2019.

[14] J. Duník, O. Straka, and M. Šimandl, "Stochastic integration filter," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1561–1566, 2013.

[15] Z. Chen *et al.*, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[16] Z. Feng, X. Wen-fang, and L. Xi, "Overview of nonlinear Bayesian filtering algorithm," *Procedia Engineering*, vol. 15, pp. 489 – 495, 2011. CEIS 2011.

[17] C. Morefield, "Application of 0-1 integer programming to multitarget tracking problems," *IEEE Transactions on Automatic Control*, vol. 22, pp. 302–312, June 1977.

[18] D. B. Reid, "An algorithm for tracking multiple targets," in *1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, pp. 1202–1211, 1978.

[19] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, pp. 843–854, December 1979.

[20] C.-Y. Chong, S. Mori, and D. B. Reid, "Forty years of multiple hypothesis tracking - a review of key developments," *2018 21st International Conference on Information Fusion (FUSION)*, pp. 452–459, 2018.

[21] I. Goodman and R. Mahler, *Mathematics of Data Fusion.* Kluwer Academic Publishers, 1997.

[22] R. Mahler, "A brief survey of advances in random-set fusion," in *2015 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 62–67, Oct 2015.

[23] R. L. Streit, *Poisson Point Processes: Imaging, Tracking, and Sensing.* Springer, 2010.

[24] R. L. Streit, "Multisensor multitarget intensity filter," in *2008 11th International Conference on Information Fusion*, pp. 1–8, June 2008.

[25] R. L. Streit and L. D. Stone, "Bayes derivation of multitarget intensity filters," in *2008 11th International Conference on Information Fusion*, pp. 1–8, June 2008.

[26] R. Mahler, "Statistics 103" for multitarget tracking," *Sensors*, no. 19(1):202, 2019.

[27] R. Mahler, "On point processes in multitarget tracking," *arXiv:1603.02373 [stat.OT]*, 2016.

[28] L. Svensson, K. Granström, and Y. Xia, "Multi-object tracking for automotive systems [online course, lecture videos]." edX, accessible at: `https://www.edx.org/course/multi-object-tracking-for-automotive-systems`, 2019.

[29] K. Granström, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview, and applications," *Journal of Advances in Information Fusion*, vol. 12, pp. 139–174, December 2017.

[30] B.-N. Vo, B.-T. Vo, N. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *IEEE Transactions on Signal Processing*, vol. 58, pp. 5129–5141, Oct 2010.

[31] C.-Y. Chong, K.-C. Chang, and S. Mori, "A review of forty years of distributed estimation," in *21st International Conference on Information Fusion (FUSION)*, pp. 70–77, 2018.

[32] B. Anderson and J. Moore, *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hall, 1979.

[33] F. L. Lewis, L. Xie, and D. Popa, *Optimal and Robust Estimation With an Introduction to Stochastic Control Theory, second edition.* Boca Raton: CRC Press, Taylor&Francis Group, 2008.

[34] R. Mahler, "Exact closed-form multitarget Bayes filters," *Sensors*, no. 19(12):2818, 2019.

[35] R. Mahler, "Measurement-to-track association and finite-set statistics," *arXiv:1701.07078 [stat.ME]*, 2017.

[36] P. Billingsley, *Probability and Measure, Anniversary Edition*. Wiley Series in Probability and Statistics, Wiley, 2012.

[37] B.-N. Vo and S. Singh, "On the Bayes filtering equations of finite set statistics," in *2004 5th Asian Control Conference (IEEE Cat. No.04EX904)*, vol. 2, pp. 1264–1269 Vol.2, July 2004.

[38] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, pp. 1152–1178, Oct 2003.

[39] C. Degen, R. Streit, and W. Koch, "On the functional derivative with respect to the Dirac Delta," in *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pp. 1–8, Oct 2015.

[40] D. Clark and R. Mahler, "Generalized PHD filters via a general chain rule," in *2012 15th International Conference on Information Fusion*, pp. 157–164, July 2012.

[41] A. Gad, F. Majdi, and M. Farooq, "A comparison of data association techniques for target tracking in clutter," in *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, vol. 2, pp. 1126–1133 vol.2, July 2002.

[42] X. Rong Li and Y. Bar-Shalom, "Tracking in clutter with nearest neighbor filters: analysis and performance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 995–1010, July 1996.

[43] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter," *Proceedings of the IEEE*, vol. 92, pp. 536–557, March 2004.

[44] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filter," *IEEE Control Systems Magazine*, vol. 29, pp. 82–100, Dec 2009.

[45] B. Ristic, B.-N. Vo, B.-T. Vo, and A. Farina, "A tutorial on Bernoulli filters: Theory, implementation and applications," *IEEE Transactions on Signal Processing*, vol. 61, pp. 3406–3430, July 2013.

[46] J. Houssineau, E. D. Delande, and D. E. Clark, "Notes of the summer school on finite set statistics," *arXiv:1308.2586 [math.PR]*, 2013.

[47] R. Streit, C. Degen, and W. Koch, "The pointillist family of multitarget tracking filters," *arXiv:1505.08000 [stat.AP]*, 2015.

[48] B.-N. Vo and B.-T. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, 2013.

[49] B.-N. Vo, B.-T. 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, 2014.

[50] J. L. Williams, "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, 2015.

[51] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: Direct derivation and implementation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1883–1901, 2018.

[52] Á. F. García-Fernández, L. Svensson, and M. R. Morelande, "Multiple target tracking based on sets of trajectories," *Accepted for the IEEE Transactions on Aerospace and Electronic Systems*, 2019.

[53] E. Brekke and M. Chitre, "Relationship between finite set statistics and the multiple hypothesis tracker," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1902–1917, 2018.

[54] R. Mahler, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, 2007.

[55] B.-N. Vo, B.-T. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3553–3567, 2007.

[56] R. P. S. Mahler, "A theoretical foundation for the Stein-Winter "Probability Hypothesis Density (PHD)" multitarget tracking approach," *Proceedings of the 2000 MSS National Symposium on Sensor and Data Fusion, Vol. I (unclassified), San Antonio TX, 99-117, June 20-22*, 2000.

[57] Á. F. García-Fernández and B. 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, 2015.

[58] R. L. Streit, "PHD intensity filtering is one step of a MAP estimation algorithm for positron emission tomography," in *2009 12th International Conference on Information Fusion*, pp. 308–315, 2009.

[59] Shozo Mori and Chee-Yee Chong, "An alternative form of cardinalized PHD filter or i.i.d.-approximation filter," in *2007 10th International Conference on Information Fusion*, pp. 1–8, 2007.

[60] E. Resmerita, H. W. Engl, and A. N. Iusem, "The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis," *Inverse Problems*, vol. 23, pp. 2575–2588, nov 2007.

[61] C. Pouchol and O. Verdier, "The ML–EM algorithm in continuum: sparse measure solutions," *Inverse Problems*, vol. 36, p. 035013, feb 2020.

[62] 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)*, vol. 39, no. 1, pp. 1–38, 1977.

[63] C. Byrne and P. P. B. Eggermont, *EM Algorithms*, pp. 305–388. New York, NY: Springer New York, 2015.

[64] A. Frigyik, S. Srivastava, and M. R. Gupta, "An introduction to functional derivatives," tech. rep., UWEETR-2008-0001, Dept of Electrical Engineering, University of Washington, 2008.

[65] 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, 2006.

[66] Yang Wang, Zhongliang Jing, and Shiqiang Hu, “Data association for PHD filter based on MHT,” in *2008 11th International Conference on Information Fusion*, pp. 1–8, 2008.

[67] K. Panta, D. E. 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, 2009.

[68] K. Panta, B.-N. Vo, S. Singh, and A. Doucet, “Probability hypothesis density filter versus multiple hypothesis tracking,” in *Signal Processing, Sensor Fusion, and Target Recognition XIII* (I. Kadar, ed.), vol. 5429, pp. 284 – 295, International Society for Optics and Photonics, SPIE, 2004.

[69] Á. F. García-Fernández and L. Svensson, “Trajectory probability hypothesis density filter,” in *2018 21st International Conference on Information Fusion (FUSION)*, pp. 1430–1437, 2018.

[70] Á. F. García-Fernández and L. Svensson, “Trajectory PHD and CPHD filters,” *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5702–5714, 2019.

[71] B.-N. Vo, B.-T. Vo, and A. Cantoni, “Bayesian filtering with random finite set observations,” *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1313–1326, 2008.

[72] S. R. Balaji and S. Karthikeyan, “A survey on moving object tracking using image processing,” in *2017 11th International Conference on Intelligent Systems and Control (ISCO)*, pp. 469–474, 2017.

[73] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, “Deep learning in video multi-object tracking: A survey,” *Neurocomputing*, vol. 381, pp. 61 – 88, 2020.

[74] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, “Deep learning for generic object detection: A survey,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, 2020.

[75] Z. Zou, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: A survey,” *arXiv:1905.05055 [cs.CV]*, 2019.

[76] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” *arXiv:1804.02767 [cs.CV]*, 2018.

[77] J. Redmon, “Darknet: Open source neural networks in c.” `http://pjreddie.com/darknet/`, 2013–2016.

[78] AlexeyAB, “Yolo-v3 and yolo-v2 for windows and linux.” `https://github.com/AlexeyAB/darknet?files=1`, accessed March 9th 2020.

[79] S. Dubuisson and C. Gonzales, “A survey of datasets for visual tracking,” *Machine Vision and Applications*, vol. 27, no. 1, pp. 23–52, 2016.

[80] D. Schuhmacher, B.-N. Vo, and B.-T. Vo, “A consistent metric for performance evaluation of multi-object filters,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.

[81] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, “Generalized optimal sub-pattern assignment metric,” in *2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–8, 2017.

[82] X. Rong Li and V. P. Jilkov, "Survey of maneuvering target tracking. part I. dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.

[83] J. Krejčí, "Google drive folder with attachments to this thesis." https://drive.google.com/drive/folders/1wMj2Nf2DoRDpdEtGO0lOZIxIxdDjgcxV?usp=sharing, uploaded April 11th 2020.

[84] P. Jodoin, L. Maddalena, A. Petrosino, and Y. Wang, "Extensive benchmark and survey of modeling methods for scene background initialization," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5244–5256, 2017.

[85] "EC Funded CAVIAR project/IST 2001 37540, found at URL:." http://homepages.inf.ed.ac.uk/rbf/CAVIAR/, accessed April 4th 2020.

[86] Y. Wu, Haibin Ling, Jingyi Yu, Feng Li, Xue Mei, and Erkang Cheng, "Blurred target tracking by blur-driven tracker," in *2011 International Conference on Computer Vision*, pp. 1100–1107, 2011.

[87] Y. Wu, J. Lim, and M. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.

[88] "Visual tracker benchmark dataset." http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html, accessed April 4th 2020.

[89] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," in *2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 1–6, 2009.

[90] W. Koch, "On anti-symmetry in multiple target tracking," in *2018 21st International Conference on Information Fusion (FUSION)*, pp. 957–964, 2018.

[91] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure, 2nd edition.* Springer, 2008.

[92] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods: 1st edition.* Springer, 2003.

[93] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications, 2nd edition.* Wiley, September 1995.

[94] B.-N. Vo, S. Singh, and A. Doucet, "Sequential monte carlo implementation of the PHD filter for multi-target tracking," in *Sixth International Conference of Information Fusion, 2003. Proceedings of the*, vol. 2, pp. 792–799, July 2003.