



**Università
degli Studi
di Ferrara**

DOCTORAL COURSE IN ENGINEERING SCIENCE

Cycle XXXVIII

Coordinator: Prof. Pier Ruggero Spina

**Multitarget tracking and classification
for sensing of things**

Scientific-Disciplinary Sector ING-INF/03

Candidate

Alessandro Vaccari

Supervisor

Prof. Andrea Conti

(signature)

(signature)

Year 2022-2025

Sommario

La conoscenza del contesto e della situazione in cui operare è un fattore cruciale per abilitare nuovi casi d'uso nelle reti wireless di prossima generazione (xG), inclusi quelli riguardanti sicurezza pubblica, veicoli autonomi, e ambienti intelligenti. Tracciamento e classificazione di obiettivi multipli (e.g., umani, robot, e veicoli) attraverso tecnologie basate su wireless sensing sono essenziali per estrarre conoscenza non solo da dispositivi connessi alla rete, ma anche da ciò che non sta comunicando informazioni. Infatti, l'ente 3GPP sta considerando, nel processo di standardizzazione di reti xG, casi d'uso che coinvolgono sia quegli obiettivi che non possono essere connessi alla rete (i.e., device-free), che quelli che non hanno interesse nel farsi rilevare (i.e., non-collaborative).

Raggiungere elevata affidabilità e accuratezza nel tracciamento e nella classificazione (e.g., identificazione e riconoscimento di attività) di obiettivi non connessi attraverso tecnologie di wireless sensing richiede lo sviluppo di nuovi approcci probabilistici e basati su intelligenza artificiale. Questa tesi mira a sviluppare nuove metodologie accurate e affidabili per estrarre congiuntamente informazioni sia di posizione che di classificazione da obiettivi non connessi, anche se inseriti in ambienti wireless complessi. In particolare, è stato proposto un approccio per stimare posizione e classe in modo integrato sfruttando approcci bayesiani e reti neurali per la classificazione di variazioni Doppler. Tale approccio è stato esteso con l'introduzione di un nuovo paradigma che prende il nome di joint tracking and classification (JTC). Introducendo la variabile di classificazione all'interno della ricorsione bayesiana, tipicamente utilizzata per il solo tracciamento, è possibile stimare congiuntamente posizione e classe di diversi obiettivi. Ulteriormente, viene proposta una derivazione per la stima congiunta attraverso JTC che abilita una forma chiusa di calcolo, la quale risulta necessaria per lo sviluppo di sistemi in grado di operare in tempo reale. Infine, è stata sviluppata un'architettura basata su microservizi, che consente di considerare fonti dati eterogenee e di distribuire il carico computazionale tra diversi servizi, sia locali che remoti. In tale architettura sono stati introdotti anche algoritmi di localizzazione e classificazione basati su immagini acquisite da telecamere, al fine di sfruttare sorgenti di dati eterogenee. L'accuratezza dei metodi proposti è stata quantificata attraverso sperimentazione in ambienti wireless complessi, tra cui quelli industriali. I risultati sperimentali dimostrano che gli approcci proposti garantiscono un errore di localizzazione di pochi centimetri e di superare il 90% di accuratezza nella classificazione operando con latenze inferiori a 100 ms.

Abstract

Situation awareness is crucial for enabling innovative use cases in next generation (xG) wireless networks, including public safety, autonomy, smart environments. Tracking and classification of humans, vehicles, and robots via wireless sensing guarantees information superiority by considering both connected and unconnected targets. Indeed, the 3GPP is considering use cases regarding unconnected (i.e., non-collaborative or device-free) targets – namely those either without interest in being identified or that are not possible nor convenient to connect to the network – in the standardization process of xG networks. In particular, innovative use cases for xG networks demand availability of several context information, which can be obtained by performing, simultaneously, tracking and classification (e.g., targets identification, activity recognition, and behavioral analysis) of non-collaborative targets via multimodal sensing. The goal of this thesis is to develop novel approaches for multimodal tracking and classification of non-collaborative targets.

The first contribution of this thesis is the development of an approach for accurate wireless sensing from samples of reflected radar signals. In particular, target positional information are employed to extract Doppler features of reflected signals used for recognizing target identity and activity. Since localization is essential for classification, both target identification and activity recognition also benefit positional inference. Then, an approach to estimate jointly position and class of multiple targets has been developed. Furthermore, a closed-form posterior distribution is derived to enable an amenable computation of the aforementioned estimate. Finally, localization and classification based on camera frames has been introduced to enhance sensing reliability and accuracy. The developed approaches have been exploited in a microservice-based architecture, which allows considering multiple and heterogeneous data sources and distribute the computation over different services either local or remote. The accuracy of the proposed approaches has been quantified via experimentation and field-trials in different indoor and outdoor environments considering a variety of people and mobile robots. An entire sensing system employing off-the-shelf radar and camera devices has been developed for operating in real-time (i.e., latency in the order of 100 ms) with high-accuracy (i.e., tracking precision in the order of decimeters and classification accuracy above 90 %) across different scenarios and environmental conditions.

Contents

List of tables	vii
List of figures	x
List of abbreviations	xi
1 Introduction	1
2 Preliminaries	7
2.1 Tracking of multiple targets	7
2.1.1 Bayesian filtering	8
2.1.2 Data association	11
2.2 Wireless sensing based on radar technologies	15
2.2.1 Different types of radar technologies	16
2.2.2 Signal processing for automotive radars	17
2.3 Wireless sensing based on camera technologies	19
2.3.1 Data transformation via homography	21
2.3.2 Calibration of the homography matrix	21
3 Integrated tracking and classification of multiple targets	25
3.1 System model	25
3.2 Detection technologies	26
3.3 Tracking technologies	28
3.4 Classification technologies	31
3.4.1 Target identification	32
3.4.2 Activity recognition	33
4 Joint tracking and classification of multiple targets	35
4.1 Joint tracking and classification filtering	35
4.1.1 Bayesian filtering	36
4.1.2 Data association	40
4.2 Joint tracking and classification via single MIMO radar at mmWaves .	43

4.2.1	Detection technologies	45
4.2.2	Tracking and classification technologies	46
5	Microservice-based architecture for multimodal sensing	49
5.1	Microservice-based architecture	49
5.1.1	Architecture components	51
5.1.2	Architecture data model	53
5.2	Multimodal wireless sensing	55
6	Field-trials and experimentation	59
6.1	Experimentation of integrated tracking and classification	59
6.1.1	System setting	59
6.1.2	Tracking algorithm	60
6.1.3	Identification algorithm	62
6.1.4	Activity recognition	63
6.2	Experimentation of joint tracking and classification	64
6.2.1	System setting	64
6.2.2	Experimental results	69
6.3	Field-trials and technology readiness	72
6.3.1	Tracking of people in an indoor environment	72
6.3.2	Tracking of people in an outdoor environment	73
6.3.3	Multitarget tracking and classification in an indoor environment	74
6.3.4	Multitarget tracking and classification in an industrial pilot line	75
6.3.5	Multitarget tracking and classification in an industrial environment	76
7	Conclusion	79
	List of contributions	81
	Bibliography	83

List of Tables

2.1	Variable definition	9
6.1	Experimentation parameters	65

List of Figures

1.1	Wireless sensing for situation awareness in xG wireless networks.	2
1.2	Situation awareness in industrial environments	3
3.1	Illustration of the ITC framework	26
3.2	Block diagram of integrated tracking and classification	27
3.3	RD map of before and after clutter mitigation	29
3.4	Doppler signatures of a human target and an industrial vehicle	33
3.5	Time-frequency images obtained by processing Doppler signatures via CWT	34
4.1	Illustration of the JTC framework	36
4.2	JTC classification tensor	42
4.3	Block diagram of the JTC framework	44
5.1	Block diagram of the proposed microservice-based architerture	50
5.2	Data model for the proposed microservice-based architecture	52
5.4	Block diagram of the proposed microservice-based architerture for wire- less sensing	55
5.5	GUI of the proposed microservice-based architecture	56
5.6	Digital twin interface consuming sensing analytics through the microservice- based architecture for an indoor office environment	57
5.7	Digital twin dashboard consuming sensing analytics through the microservice- based architecture for an industrial environment	57
6.1	Experimental environment for tracking and classification	60
6.2	ECDF of the tracking error	61
6.3	Representation of real-time multitarget classification	62
6.4	Experimentation environment for activity recognition	63
6.5	Confusion matrix for the activity recognition accuracy	64
6.6	ECDF of tracking error for human targets using the JTC approach	67

6.7	ECDF of tracking error for human targets and mobile robots using the JTC approach	68
6.8	Identification accuracy of humans and robots using the JTC approach .	70
6.9	Tracking of people in an indoor environment	72
6.10	Tracking of people in an outdoor environment	73
6.11	Multitarget tracking and classification in an indoor environment	74
6.12	Multitarget tracking and classification in an industrial pilot line	75
6.13	Multitarget tracking and classification in an industrial environment . .	76

List of abbreviations

3GPP	3rd Generation Partnership Project	GOSPA	generalized optimal subpattern assignment
AI	artificial intelligence	GUI	graphical user interface
AMQ	advanced message queuing	IF	intermediate frequency
BLER	block error rate	IPPP	inhomogeneous Poisson point process
BS	base station	ISAC	integrated sensing and communication
CNN	convolutional neural network	LPF	low-pass filter
CWT	continuous wavelet transform	JPDA	joint probabilistic data association
DFT	discrete Fourier transform	JSON	JavaScript object notation
DBSCAN	density-based spatial clustering of applications with noise	JTC	joint tracking and classification
DLT	direct linear transformation	MCS	modulation and coding scheme
ECDF	empirical cumulative distribution function	μ D	micro-Doppler
FMCW	frequency modulated continuous wave	mmWave	millimeter wave
FS	fast sampling	MTT	multitarget tracking
FPGA	field-programmable gate array	NDT	network digital twin
GNN	global nearest neighbor	NN	neural network

OFDM	orthogonal frequency division multiplexing	SRS	sounding reference signal
RA	range-angle	SS	slow sampling
RAN	radio access network	TRL	technology readiness level
RD	range-Doppler	TRP	transmission-reception point
RDA	range-Doppler-angle	UE	user equipment
ReLU	rectified linear unit	UML	unified modeling language
RHZU	right-handed Z up	xG	next generation
RSRP	reference signal received power	YOLO	You only look once

Chapter 1

Introduction

Sensing of things will be essential in next generation (xG) wireless networks for enabling innovative use cases including public safety [1–5], smart environment [6–10], and factory of the future [11–15]. The evolution toward beyond 5G (B5G) and 6G systems is driving a new paradigm that considers the integration of sensing and communication functionalities for a new generation of networks with the capability to perceive and interpret the environment [16–24]. The 3rd Generation Partnership Project (3GPP) is considering the integrated sensing and communication (ISAC) paradigm since release 19, proposing use cases and defining service-level requirements in terms of sensing accuracy, availability, and latency [25–35].

Situation awareness for xG wireless networks requires location awareness and context information, including activities and identity of people, objects, and vehicles (i.e., targets) in a monitored environment (see Fig. 1.1). Information superiority is guaranteed by considering both connected and unconnected targets simultaneously [36–47]. If connected targets can provide information regarding their position or identity, unconnected targets require the development of novel sensing approaches. In particular, unconnected targets are those which either can not be equipped with a dedicated device for communicating (i.e., device-free targets), or do not have any interest in being revealed (i.e., non-collaborative targets).

Multitarget tracking and classification (e.g., target identification and activity recognition) via wireless sensing is conventionally performed with approaches providing reliability and accuracy only under controlled conditions. For example, technologies based on cameras and microphones can operate only in line-of-sight (LOS) conditions. In addition, they pose privacy issues due to the need of recording images and sounds. Approaches exploiting reflected radiofrequency (RF) signals are promising for ubiquitous sensing of things, while respecting privacy constraints. However, wireless signals are affected by multipath propagation and clutter conditions, especially in complex wireless environments, such as indoor and industrial areas [48–51]. This calls for the

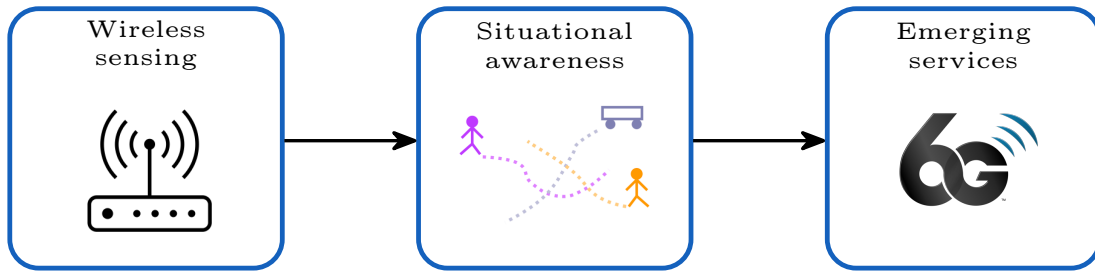


Figure 1.1: Wireless sensing for situation awareness in xG wireless networks.

development of resilient approaches to fulfill the stringent 3GPP requirements in terms of accuracy, availability, and latency.

Tracking and classification are independent tasks that can benefit each other. Localization over time of device-free targets is typically performed in a recursive manner, referred to as multitarget tracking (MTT) filtering [52–55]. Bayesian approaches are typically employed to perform MTT filtering due to an unknown number of targets which freely appear, disappear and move. At any time instant, MTT filtering consists of a positional prediction phase, based on a motion model, and an update positional phase based on the collected measurements [56–65]. The combinatorial problem to associate measurements to targets, referred to as data association, is challenging especially in complex wireless environments characterized by misdetections and false alarms due to the presence of obstacles and scatters [66–68]. Statistical processing and optimization techniques are needed to achieve a reliable data association [69–75]. Classification of device-free targets based on features of reflected RF signals calls for new methods relying on artificial intelligence (AI) due to the stochastic nature of the wireless channel [76–79]. In particular, neural networks (NNs) serves as powerful tool for signal features classification thanks to their capability of achieving high accuracy in mapping features to classes without requiring a mathematical problem description [80–85]. An accurate data association is also needed for considering reliable classification metrics. Wrong measurement-to-target associations lead to consider a coarse sequence of signal features (i.e., an unrepresentative classification metric) which produces uncertain classification. Thus, data association is a key task in MTT filtering, that defines positional measurements and signal features for tracking and classification, respectively.

Recently, several approaches to integrate classification of multiple targets and MTT have been proposed [86–95]. These approaches leverage positional information of a tracked target to collect specific signal features from the received waveforms. In particular, localization allows to determine which signal features employ as classification metric for a considered tracked target. Furthermore, location awareness provides information superiority to perform target classification by constraining the decision space, e.g., in a specific spatial position only a subset of classes is allowed. However, as

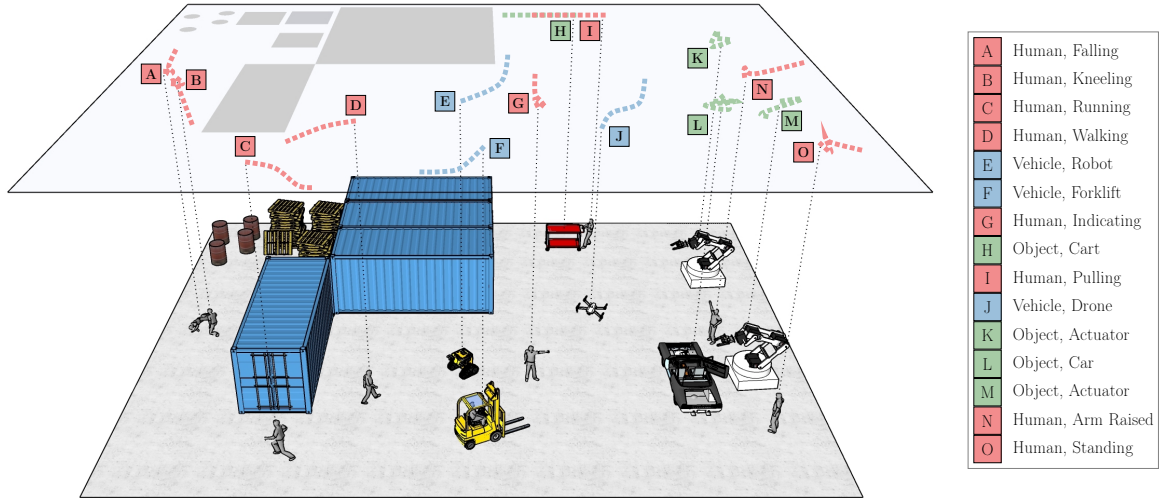


Figure 1.2: Situation awareness allows information superiority in complex wireless environments.

localization aids classification, target identification and activity recognition benefit tracking. Indeed, classification analytics enable the employment of personalized motion predictions for each tracked target based on specific kinematic models according to the target class and context information, including how targets interact each others. Data association is conventionally performed comparing statistical distances between target positional states and positional measurements. The availability of classification scores provides information superiority to complete data association.

A novel paradigm for tracking and classification is referred to as joint tracking and classification (JTC) and aims to estimate jointly position and class of device-free targets. Bayesian recursion is employed in JTC with the goal to estimate the classification analytics and positional information simultaneously, given a set of measurements. The prediction phase of the recursion considers also a classification model describing the class transition between subsequent measurements. Data association in JTC paradigm relies on both positional information and signal features. The posterior step aims to determine class and position of multiple targets by maximizing a joint distribution. The fundamental questions related to JTC are: (i) how to design an estimator which jointly estimate positions and classes of multiple targets; (ii) under which conditions is a closed-form estimate admitted; and (iii) which advantages can JTC provide by performing jointly tracking and classification? The answers to these questions allow the design and the development of JTC-based frameworks leading to more accurate and robust tracking and classification even in complex wireless environments. Fig. 1.2 shows a factory of the future in which several people, objects, and vehicles perform different activities and require different services enabled by simultaneous localization and classification.

A promising technology to perform wireless sensing of things based on reflected RF signals are radars operating at millimeter waves (mmWaves) [96–107]. In particular, orthogonal frequency division multiplexing (OFDM)-based radar and automotive radar exploiting frequency modulated continuous wave (FMCW) signals have emerged as two leading paradigms for high-resolution sensing. These technologies guarantee high resolution in both range and Doppler domains thanks to the large bandwidth available at mmWaves frequencies [108–115]. Moreover, short wavelengths enable compact antenna arrays which can be realized in multiple-input–multiple-output (MIMO) configurations increasing the angular resolution for target detection. Then, on one hand radar technologies allow sensing of device-free targets by processing reflected waveforms to reveal the target presence and obtain single-value estimates (SVEs) for positional inference, such as time-of-flight (TOF) and angle-of-arrival (AOA) [116–118]. On the other hand, features of a reflected waveform (e.g., micro-Doppler (μ D) shifts) characterize context information (e.g., target movements) and can be used as input for target classification [119–130]. A set of SVEs produces a positional measurement, while signal features are employed to obtain a classification metric.

Information superiority for tracking and classification can be achieved through multimodal sensing. Multimodal sensing approaches integrate heterogeneous sensing technologies, each offering distinct levels of reliability, spatial resolution, and accuracy. By combining complementary modalities, these systems enhance the robustness and completeness of environmental perception. In particular, data diversity allows to observe the monitored environment from multiple perspectives, allowing dynamic selection and fusion of the most informative modalities to optimize sensing performance under varying conditions [131–134]. However, the use of heterogeneous sensors, characterized by differing resolutions and data representations, necessitates the development of microservice-based architecture for effective data fusion and information extraction.

This thesis aims to introduce novel approaches for multitarget tracking and classification of unconnected targets through wireless and multimodal sensing, with the goal of enhancing situation awareness in xG networks. The reminder of this thesis is organized as follows. Chapter 2 provides the theoretical background on multitarget tracking, radar signal processing, and camera-based localization. Chapter 3 presents the proposed framework for integrated tracking and classification based on automotive radar. Chapter 4 introduces the JTC approach, enabling the simultaneous estimation of position and class. Chapter 5 extends the developed methods to multimodal sensing, combining radar and camera data within a microservice-based architecture for distributed and scalable processing. Chapter 6 reports the experimental validation and field-trials of the proposed approaches, demonstrating the effectiveness and robustness in realistic scenarios. Finally, conclusions summarize the main contributions of this dissertation.

The research activity described in this thesis has been conducted within national and international projects at the wireless communication and localization networks (WCLN) laboratory at the Department of Engineering, University of Ferrara. The results presented in this thesis have been published in premier journals and in proceedings of international conferences as indicated in the authors publication list at the end of the document. Part of the research activities have been conducted within the *ONR Global* projects nCOA (non-Collaborative Object Analytics), the *NextGenerationEU Italian NRRP RESTART Program* project TAP-IN (Technologies, Algorithms, and Protocols for use cases of Industrial Networks), and the *Programma di supporto alle tecnologie emergenti 5G* project CTE (Casa delle Tecnologie Emergenti).

Notation

The notation of this thesis is described in the following. Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by $\mathbf{x}$ and x ; a random vector and its realization are denoted by $\mathbf{x}$ and $\mathbf{x}$; a random matrix and its realization are denoted by $\mathbf{X}$ and $\mathbf{X}$, respectively. Sets and random sets are denoted by upright sans serif and calligraphic font, respectively. For example, a random set and its realization are denoted by $\mathbf{X}$ and $\mathcal{X}$, respectively. The m -by- n matrix of ones is denoted by $\mathbf{1}_{m \times n}$. The function $\mathbf{f}_{\mathbf{x}}(\mathbf{x})$ and, for brevity when possible, $\mathbf{f}(\mathbf{x})$ denote the probability density function (PDF) of a continuous random variable (RV) $\mathbf{x}$; $\mathbf{f}_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})$ and, for brevity when possible, $\mathbf{f}(\mathbf{x}|\mathbf{y})$ denote the PDF of $\mathbf{x}$ conditional on $\mathbf{y} = \mathbf{y}$. The function $\varphi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the probability distribution function (PDF) of a Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Integrals are over the entire space of the integration variable.

Chapter 2

Preliminaries

This chapter provides the theoretical and methodological foundations that underpin the approaches developed in the following chapters. First, the principles of MTT filtering and data association are introduced, which serve as the basis for estimating the trajectories of multiple targets in complex wireless environments. Then, radar sensing technologies based on OFDM and FMCW signals are discussed. Automotive radars, i.e., FMCW MIMO radars operating at mmWaves frequencies, are presented in detail highlighting their signal processing pipeline for range-Doppler-angle (RDA) estimation, clutter mitigation, and target detection. Finally, camera-based sensing is proposed as a complementary approach for localization and classification. This chapter establishes the background for the integrated and joint tracking and classification frameworks introduced in the rest of the dissertation. It also paves the way for the development of multimodal sensing frameworks based on the use of heterogeneous technologies, which is promising to enhance situational awareness and environmental perception in xG wireless networks.

2.1 Tracking of multiple targets

At the discrete time instant t_k , consider N_k tracked targets. The i -th target, for $i \in \{1, 2, \dots, N_k\}$, is characterized by its positional state $\mathbf{x}_k^{(i)} \in \mathbb{R}^{n_x}$ and its class $\mathbf{c}_k^{(i)} \in \mathcal{C}_t$, where $\mathcal{C}_t$ is the set containing all the possible classes which can be assigned to a tracked target. Specifically, the positional state $\mathbf{x}_k^{(i)}$ is a random column vector of dimension n_x , which includes target position and mobility parameters, such as instant velocity and acceleration. The class $\mathbf{c}_k^{(i)}$ is the target class, e.g., the target type and/or the activity performed. Since multiple targets are tracked and classified simultaneously, consider the multitarget positional state $\mathbf{x}_k = [\mathbf{x}_k^{(1)\top}, \mathbf{x}_k^{(2)\top}, \dots, \mathbf{x}_k^{(N_k)\top}]^\top$ and the multitarget class $\mathbf{c}_k = [\mathbf{c}_k^{(1)}, \mathbf{c}_k^{(2)}, \dots, \mathbf{c}_k^{(N_k)}]^\top$ containing all the estimated positional

states and classes, respectively.

At the discrete time instant k , consider M_k measurements (i.e., detections of sensor networks) obtained by processing samples of reflected waveforms. The j -th measurement, for $j \in \{1, 2, \dots, M_k\}$, is a random vector of dimension m_z defined as $\mathbf{z}_k^{(j)} \in \mathbb{R}^{m_z}$. It contains positional information and other signal features of a detection originated from either a target or clutter. Positional information allow tracking, while signal features determine the class $\mathbf{c}_k^{(i)} \in \mathcal{C}$, with $i \in \{1, 2, \dots, N_k\}$ and $\mathcal{C} = \mathcal{C}_t \cup \mathcal{C}_\zeta$, where $\mathcal{C}_\zeta$ is the set of classes that is not of interest to track. For example, in a certain scenario, it may be of interest to track children and not adults. Therefore, it results $\mathcal{C}_t = \{\text{child}\}$ and $\mathcal{C}_\zeta = \{\text{adult, clutter}\}$. The multi-measurement vector $\mathbf{z}_k = [\mathbf{z}_k^{(1)\text{T}}, \mathbf{z}_k^{(2)\text{T}}, \dots, \mathbf{z}_k^{(M_k)\text{T}}]^\text{T}$ contains all the revealed detections at the current time instant. In particular, the multi-measurement vector $\mathbf{z}_k$ consists of $M_{o,k}$ target-related observations (i.e., subset of measurements due to targets of interest) and $M_{\zeta,k}$ false-alarm measurements (i.e., subset of measurements due to clutter), such that $M_k = M_{o,k} + M_{\zeta,k}$. Thus, at the current time k , we can define $\mathbf{z}_k = [\mathbf{o}_k^\text{T}, \boldsymbol{\zeta}_k^\text{T}]^\text{T}$, where $\mathbf{o}_k = [\mathbf{o}_k^{(1)\text{T}}, \mathbf{o}_k^{(2)\text{T}}, \dots, \mathbf{o}_k^{(M_{o,k})\text{T}}]^\text{T}$, with $\mathbf{o}_k^{(j)} \in \mathbb{R}^{m_z}$ and $j \in \{1, 2, \dots, M_{o,k}\}$, is the vector of observations and $\boldsymbol{\zeta}_k = [\boldsymbol{\zeta}_k^{(1)\text{T}}, \boldsymbol{\zeta}_k^{(2)\text{T}}, \dots, \boldsymbol{\zeta}_k^{(M_{\zeta,k})\text{T}}]^\text{T}$, with $\boldsymbol{\zeta}_k^{(j)} \in \mathbb{R}^{m_z}$ and $j \in \{1, 2, \dots, M_{\zeta,k}\}$, is the vector of false alarms. For the regards of this work consider that: (i) an observation can be generated by at most one target and a target can generate at most one observation,¹ i.e., $M_{o,k} \leq N_k$; and (ii) all the measurements are independent of each other. Tab. 2.1 summarizes the variables introduced so far.

In the remaining of this section, we address MTT filtering and, therefore, only the positional states are considered.

2.1.1 Bayesian filtering

Location awareness of multiple targets simultaneously can be achieved by leveraging MTT filtering. To perform MTT filtering consists in (i) completing data association, namely the problem to associate measurements collected from sensing devices (e.g., radar sensor networks) to the tracked targets and (ii) estimating the new positional state of each target given data association. In particular, at each time instant, a target positional state can be estimated as follows. First, the prediction phase evaluates the positional prediction density relying on a kinematic model that describes the target motion. Then, the update phase evaluates the posterior density by considering both the positional prediction density and the positional measurement associated to the target.

¹According to the availability of high-resolution sensors, several detectors output a set of observations for each target, which can be reduced to a single estimate by performing clustering and considering the cluster centroids.

Table 2.1: Variable definition.

Notation	Definition at any time instant k
N_k	Number of targets
M_k	Number of measurements
$M_{o,k}$	Number of target-related observations
$M_{\zeta,k}$	Number of false-alarms measurements
$\mathbf{x}_k^{(i)}$	Positional state of the i -th target
$\mathbf{x}_k$	Multitarget positional state
$\mathbf{z}_k^{(j)}$	Positional measurement j -th
$\mathbf{z}_k$	Multitarget measurement
$\mathbf{o}_k^{(j)}$	Positional target-related observation j -th
$\mathbf{o}_k$	Vector of target-related observations
$\boldsymbol{\zeta}_k^{(j)}$	False-alarm measurement j -th
$\boldsymbol{\zeta}_k$	Vector of false alarm measurements
$\theta_k^{(i)}$	Data association for the i -th target
$\boldsymbol{\theta}_k$	Multitarget data association vector
Θ_k	Set of all the valid data association vectors
$c_k^{(i)}$	Class for the i -th target
$\mathcal{C}$	Set of possible classes
$\mathbf{c}_k$	Multitarget class vector
P_k	Set of all the valid class vectors

Consider the multitarget positional state $\mathbf{x}_k$, which evolves as a first-order Markov process. In particular, the positional Markov transition density $f(\mathbf{x}_k|\mathbf{x}_{k-1})$ describes the probability that $\mathbf{x}_{k-1}$ evolves to $\mathbf{x}_k$ according to the transition equation given by

$$\mathbf{x}_k = \boldsymbol{\gamma}(\mathbf{x}_{k-1}, \mathbf{v}_k) \quad (2.1)$$

where $\boldsymbol{\gamma}(\cdot, \cdot)$ is a nonlinear transformation and $\mathbf{v}_k$ is the transition noise. By the Chapman-Kolmogorov equation the prediction density can be defined as

$$f(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int f(\mathbf{x}_k|\mathbf{x}_{k-1})f(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1} \quad (2.2)$$

where $f(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ is the posterior density at the previous time instant and $\mathbf{z}_{1:k-1} = [\mathbf{z}_1^T, \mathbf{z}_2^T, \dots, \mathbf{z}_{k-1}^T]^T$ is the measurement history from the first time instant to $k-1$.

The positional measurement likelihood $f(\mathbf{z}_k|\mathbf{x}_k)$ defines the probability to have the multi-measurement vector $\mathbf{z}_k$ given the multitarget positional state $\mathbf{x}_k$ according to the observation equation given by

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{w}_k) \quad (2.3)$$

where $\mathbf{h}(\cdot, \cdot)$ is a nonlinear transformation and $\mathbf{w}_k$ is the observation noise. The multitarget positional density, i.e., Bayes posterior density, can be defined as

$$f(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{f(\mathbf{z}_k|\mathbf{x}_k) f(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{f(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \quad (2.4)$$

where $f(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ is the normalization term given by

$$f(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int f(\mathbf{z}_k|\mathbf{x}_k) f(\mathbf{x}_k|\mathbf{z}_{1:k-1}) d\mathbf{x}_k. \quad (2.5)$$

A closed-form solution for the MTT filtering is given by the Kalman filter (KF) if linear and Gaussian models are considered.² Specifically, Gaussian transition equation and Gaussian observation equation are, respectively, given by

$$\gamma(\mathbf{x}_{k-1}, \mathbf{v}_k) = \varphi(\mathbf{x}_k; \mathbf{\Gamma}_k \mathbf{x}_{k-1}, \mathbf{V}_k) \quad (2.6a)$$

$$\mathbf{h}(\mathbf{x}_k, \mathbf{w}_k) = \varphi(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k, \mathbf{W}_k). \quad (2.7a)$$

The transition matrix $\mathbf{\Gamma}_k$ defines the linear transform from $\mathbf{x}_{k-1}$ to $\mathbf{x}_k$, while the observation matrix $\mathbf{H}_k$ provides a linear relation between states and observations. Transition noise $\mathbf{v}_k$ and observation noise $\mathbf{w}_k$ are considered independent zero-mean Gaussian vectors, with covariance matrix $\mathbf{V}_k$ and $\mathbf{W}_k$, respectively. Considering Gaussian posterior density at the first time instant, all the subsequent densities are Gaussian and a closed-form solution for the Bayes recursion is given by the KF. Then, the Gaussian prediction density can be written as

$$f(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \varphi(\mathbf{x}_k; \tilde{\mathbf{x}}_k, \tilde{\mathbf{P}}_k) \quad (2.8)$$

where

$$\begin{aligned} \tilde{\mathbf{x}}_k &= \mathbf{\Gamma}_k \hat{\mathbf{x}}_{k-1} \\ \tilde{\mathbf{P}}_k &= \mathbf{V}_{k-1} + \mathbf{\Gamma}_k \hat{\mathbf{P}}_{k-1} \mathbf{\Gamma}_k^T \end{aligned} \quad (2.9)$$

²For nonlinear non-Gaussian system models, computationally feasible approximate algorithms include variants of the KF, such as the extended KF and the unscented KF. Both the extended KF and the unscented KF are based on the Kalman equations but perform different approximations of the system models.

and $\mathbf{I}_k^T$ is the transpose of $\mathbf{I}_k$. The Gaussian Bayes posterior density can be written as

$$f(\mathbf{x}_k | \mathbf{z}_{1:k}) = \varphi(\mathbf{x}_k; \hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k) \quad (2.10)$$

where

$$\mathbf{K}_k = \tilde{\mathbf{P}}_k \mathbf{H}_k^T \Sigma_k^{-1} \quad (2.11a)$$

$$\Sigma_k = \mathbf{W}_k + \mathbf{H}_k \tilde{\mathbf{P}}_k \mathbf{H}_k^T \quad (2.11b)$$

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \tilde{\mathbf{x}}_k) \quad (2.11c)$$

$$\hat{\mathbf{P}}_k = [\mathbf{I} - \mathbf{K}_k \mathbf{H}_k] \tilde{\mathbf{P}}_k. \quad (2.11d)$$

Specifically, $\mathbf{K}_k$ is referred to as Kalman gain and Σ_k innovation covariance matrix, while Σ_k^{-1} denotes the inverse of Σ_k .

2.1.2 Data association

In MTT, distinguishing which target generates which measurement is a challenging task especially in harsh propagation environments due to the high probability of mis-detections and false-alarms. Let $\theta_k^{(i)}$ denote the data association variable representing the association between the i -th target and the j -th measurement, at the time instant k , as follows

$$\theta_k^{(i)} = \begin{cases} j & \text{if target } i \text{ is associated to a measurement } j \\ 0 & \text{if target } i \text{ is associated to no measurement.} \end{cases} \quad (2.12)$$

In the presence of multiple targets, the multitarget data association vector $\boldsymbol{\theta}_k = [\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N_k)}]^T$ denotes the measurement-to-target associations for all the tracked targets. In particular, the multitarget data association vector is such that $\boldsymbol{\theta}_k \in \Theta_k$, where Θ_k is the set of all the valid measurement-to-target association vectors. A valid association requires that (i) each target is either associated to a measurement or misdetected and (ii) any pair of detected targets is not associated to the same measurement.

The multitarget data association vector $\boldsymbol{\theta}_k$ determines the false alarms vector $\boldsymbol{\zeta}_k = [\zeta_k^{(1)T}, \zeta_k^{(2)T}, \dots, \zeta_k^{(M_{\zeta,k})T}]^T$, which is typically described as an inhomogeneous Poisson point process (IPPP) parameterized by the clutter PDF defined as

$$f(\boldsymbol{\zeta}_k) = \frac{\lambda_k(\boldsymbol{\zeta}_k)}{\lambda_k} \quad (2.13)$$

where $\lambda_k(\zeta_k)$ is the clutter intensity function of the considered environment and

$$\bar{\lambda}_k = \int \lambda_k(\zeta_k) d\zeta_k \quad (2.14)$$

is the expected number of clutter detections, i.e., the normalization term. Tab. 2.1 summarizes the variables introduced so far.

A measurement is an observation, or a misdetection, generated by the i -th target with probability

$$\begin{cases} P_d(\mathbf{x}_k^{(i)}) f(\mathbf{z}_k^{(\theta_k^{(i)})} | \mathbf{x}_k^{(i)}) & \text{if } \mathbf{o}_k = [\mathbf{z}_k^{(\theta_k^{(i)})}] \\ 1 - P_d(\mathbf{x}_k^{(i)}) & \text{if } \mathbf{o}_k = \emptyset \end{cases} \quad (2.15)$$

where $P_d(\cdot)$ is the detection probability and $f(\mathbf{z}_k^{(\theta_k^{(i)})} | \mathbf{x}_k^{(i)})$ is the probability to observe $\mathbf{z}_k^{(\theta_k^{(i)})}$ given the positional state $\mathbf{x}_k^{(i)}$.

Consider the Bayes posterior density from (2.4). Employing the law of the total probability and by assuming independent targets, it can be written

$$f(\mathbf{x}_k | \mathbf{z}_{1:k}) \propto f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{z}_{1:k-1}) \quad (2.16a)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} f(\mathbf{z}_k, \boldsymbol{\theta}_k, \mathbf{x}_k | \mathbf{z}_{1:k-1}) \quad (2.16b)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} \prod_{i=1}^{N_k} f(\mathbf{z}_k^{(\theta_k^{(i)})}, \theta_k^{(i)}, \mathbf{x}_k^{(i)} | \mathbf{z}_{1:k-1}^{(\theta_{1:k-1}^{(i)})}). \quad (2.16c)$$

For a given data association vector $\boldsymbol{\theta}_k$ (i.e., a data association hypothesis), $\theta_k^{(i)}$ and $\mathbf{z}_k^{(\theta_k^{(i)})}$ can be seen as parameters of a function $g(\cdot)$ defined as

$$g(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) = f(\mathbf{z}_k^{(\theta_k^{(i)})}, \theta_k^{(i)}, \mathbf{x}_k^{(i)} | \mathbf{z}_{1:k-1}^{(\theta_{1:k-1}^{(i)})}). \quad (2.17a)$$

The normalized version of $g(\cdot)$, denoted as $\bar{g}(\cdot)$, is given by

$$\bar{g}(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) = \frac{g(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)})}{w(\mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)})} \quad (2.18)$$

where

$$w(\mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) = \int g(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) d\mathbf{x}_k^{(i)}. \quad (2.19)$$

Combining (2.16c) and (2.18) results

$$f(\mathbf{x}_k | \mathbf{z}_{1:k}) \propto \sum_{\boldsymbol{\theta}_k \in \Theta_k} \prod_{i=1}^{N_k} f\left(\mathbf{z}_k^{(\theta_k^{(i)})}, \theta_k^{(i)}, \mathbf{x}_k^{(i)} | \mathbf{z}_{1:k-1}^{(\theta_{1:k-1}^{(i)})}\right) \quad (2.20a)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} \prod_{i=1}^{N_k} w\left(\mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \times \prod_{i=1}^{N_k} \bar{g}\left(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \quad (2.20b)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \times \prod_{i=1}^{N_k} \bar{g}\left(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \quad (2.20c)$$

where $w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)$ is referred to as data association weight and it can be maximized in order to identify the most likely data association vector $\boldsymbol{\theta}_k^*$. The Bayes posterior density is given by

$$f(\mathbf{x}_k | \mathbf{z}_{1:k}) = \sum_{\boldsymbol{\theta}_k \in \Theta_k} \bar{w}(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \times \prod_{i=1}^{N_k} \bar{g}\left(\mathbf{x}_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \quad (2.21)$$

where $\bar{w}(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)$ is such that

$$\bar{w}(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) = \frac{w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)}{\sum_{\boldsymbol{\theta}_k \in \Theta_k} w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)} \quad (2.22)$$

and $\bar{w}(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)$ is referred to as data association density.

Introducing a cost matrix $\mathbf{L}$ and an assignment matrix $\mathbf{A}$ data association can be formulated as an optimization problem. In a general optimization problem, the cost matrix element $l^{(i,j)}$ of $\mathbf{L}$ represents the cost to associate the row element i with the column element j , while the assignment matrix element $a^{(i,j)} \in \mathbf{A}$ is such that

$$a^{(i,j)} = \begin{cases} 1 & \text{if column } j \text{ is assigned to row } i \\ 0 & \text{otherwise.} \end{cases} \quad (2.23)$$

Thus, the total cost of multiple assignments can be obtained as

$$c = \sum_i \sum_j a^{(i,j)} l^{(i,j)} = \text{tr}(\mathbf{A}^T \mathbf{L}). \quad (2.24)$$

The optimal assignment problem consists in finding the solution $\mathbf{A}^*$ given $\mathbf{L}$ such that minimizes $\text{tr}(\mathbf{A}^T \mathbf{L})$ under the following conditions

$$a^{(i,j)} \in \{0, 1\}, \forall i, j \quad (2.25a)$$

$$\sum_j a^{(i,j)} = 1, \forall i \quad (2.25b)$$

$$\sum_i a^{(i,j)} = 1, \forall j. \quad (2.25c)$$

Both $\mathbf{A}$ and $\mathbf{L}$ are defined as an $N_k \times (M_k + N_k)$ matrix, in which each row represents a target and each column either a measurement or a misdetection. Specifically, $\mathbf{A}$ and $\mathbf{L}$ are composed as follows:

- for $i \in \{1, 2, \dots, N_k\}$ and $j \in \{1, 2, \dots, M_k\}$, the assignment variable $a^{(i,j)}$ states the association between the j -th measurement and the i -th target with cost $l^{(i,j)}$;
- for $i \in \{1, 2, \dots, N_k\}$ and $j = M_k + i'$ with $i = i'$, the assignment variable $a^{(i,j)}$ states whether the i -th target is misdetected according to the cost $l^{(i,j)}$; and
- for $i \in \{1, 2, \dots, N_k\}$ and $j = M_k + i'$ with $i \neq i'$, the assignment variable $a^{(i,j)}$ is forced to zero according to $l^{(i,j)} = \infty$.

The optimization problem aims to find the most likely data association vector $\boldsymbol{\theta}_k^*$, namely the one providing the maximum data association weight $w^*(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)$. A different formulation of the problem aims to minimize the logarithm of the inverse of $w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)$, as follows

$$\boldsymbol{\theta}_k^* = \arg \max_{\boldsymbol{\theta}_k \in \Theta_k} w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \quad (2.26a)$$

$$= \arg \max_{\boldsymbol{\theta}_k \in \Theta_k} \log \left(\prod_{i=1}^{N_k} w(\mathbf{z}_{1:k}^{(\boldsymbol{\theta}_k^{(i)})}, \boldsymbol{\theta}_k^{(i)}) \right) \quad (2.26b)$$

$$= \arg \min_{\boldsymbol{\theta}_k \in \Theta_k} \sum_{i=1}^{N_k} -\log \left(w(\mathbf{z}_{1:k}^{(\boldsymbol{\theta}_k^{(i)})}, \boldsymbol{\theta}_k^{(i)}) \right). \quad (2.26c)$$

Hence, the cost matrix elements can be defined as

$$l^{(i,j)} = -\log \left(w(\mathbf{z}_{1:k}^{(\boldsymbol{\theta}_k^{(i)})}, \boldsymbol{\theta}_k^{(i)}) \right) \quad (2.27a)$$

$$= -\log \left(\int f(\mathbf{z}_k^{(\boldsymbol{\theta}_k^{(i)})}, \boldsymbol{\theta}_k^{(i)}, \mathbf{x}_k^{(i)} | \mathbf{z}_{1:k-1}^{(\boldsymbol{\theta}_k^{(i)})}) d\mathbf{x}_k^{(i)} \right) \quad (2.27b)$$

where $i \in \{1, 2, \dots, N_k\}$, $j \in \{1, 2, \dots, M_k, M_k + i\}$, and $\boldsymbol{\theta}_k^{(i)} \in \{0, 1, 2, \dots, M_k\}$.

A closed form solution to compute data association costs is defined under the following conditions:

- constant detection probability, i.e., $P_d(\mathbf{x}_k^{(i)}) = P_d, \forall i \in \{1, 2, \dots, N_k\}$;
- uniform distributed clutter in the considered environment, i.e., $\lambda_k(\zeta_k) = \lambda_k, \forall \zeta_k$;
- Gaussian prediction density with independent targets, i.e.,

$$f(\mathbf{x}_k | \mathbf{z}_{1:k-1}) = \prod_{i=1}^{N_k} \varphi(\mathbf{x}_k^{(i)}; \tilde{\mathbf{x}}_k^{(i)}, \tilde{\mathbf{P}}_k^{(i)}); \text{ and} \quad (2.28)$$

- Gaussian measurement likelihood with independent targets, i.e.,

$$f(\mathbf{z}_k | \mathbf{x}_k) = \prod_{i=1}^{N_k} \varphi(\mathbf{z}_k^{(i)}; \mathbf{H}_k \mathbf{x}_k^{(i)}, \mathbf{W}_k). \quad (2.29)$$

The closed-form solution for computing the data association cost is given by

$$\begin{aligned} l^{(i,j)} = & -\log\left(\frac{P_d}{\lambda_k}\right) + \frac{1}{2} \log\left(\det(2\pi \Sigma_k)\right) \\ & + \frac{1}{2} \left(\mathbf{z}_k^{(j)} - \mathbf{H}_k \tilde{\mathbf{x}}_k^{(i)}\right)^T \Sigma_k^{-1} \left(\mathbf{z}_k^{(j)} - \mathbf{H}_k \tilde{\mathbf{x}}_k^{(i)}\right) \end{aligned} \quad (2.30)$$

if measurement j is assigned to target i , or by

$$l^{(i,j)} = -\log(1 - P_d) \quad (2.31)$$

if target i is misdetected [54].

Finally, different algorithms, such as, global nearest neighbor (GNN) and joint probabilistic data association (JPDA), can be employed to complete tracking of multiple targets.

2.2 Wireless sensing based on radar technologies

Radar sensing is emerging as a key enabler for the evolution of xG wireless networks, where communication, localization, and environmental perception are expected to operate in a unified framework. In 5G systems, particularly those operating at mmWaves frequencies, the high bandwidth and spatial resolution of multi-antenna transceivers can be leveraged not only for data transmission but also for high-precision sensing of the propagation environment. This convergence has led to the concept of ISAC, where radar-like capabilities are embedded directly into the communication infrastructure. In this context, the same transmitted waveforms, antenna arrays, and spectral resources used for data exchange can be reused to estimate range, velocity, and angular

information of surrounding objects, allowing network nodes to act as distributed radar sensors.

2.2.1 Different types of radar technologies

In 6G networks, radar functionality is expected to play an even more central role, providing real-time situational awareness and context information to support autonomous systems, intelligent transportation, smart manufacturing, and human-machine interaction. By exploiting massive MIMO architectures, ultra-wideband (UWB) waveforms, and sub-THz carrier frequencies, 6G radar sensing can achieve centimeter-level localization accuracy and micro-Doppler resolution for motion characterization. The fusion of radar-derived information with communication metrics (e.g., channel state information, beam patterns, or signal power maps) enables a new level of network intelligence, where the wireless system dynamically adapts its operation based on environmental perception. Furthermore, the ability to perform joint sensing and communication reduces hardware redundancy, spectral congestion, and power consumption, facilitating energy-efficient and spectrum-cooperative network design. Consequently, radar sensing is envisioned as a foundational technology for 5G-Advanced and 6G systems, transforming wireless networks from connectivity providers into cognitive, perception-driven infrastructures capable of understanding and interacting with their environment.

Modern radar systems exploit diverse waveform designs to achieve accurate estimation of target range, velocity, and angular position. The choice of waveform directly impacts resolution in both range and Doppler domains, as well as interference robustness and the feasibility of integration with communication systems. Among the most prominent solutions, OFDM-based radar and automotive radar exploiting FMCW signals have emerged as two leading paradigms for high-resolution sensing, each characterized by distinct processing architectures and implementation trade-offs.

A distinctive advantage of OFDM radar is its compatibility with communication waveforms used in broadband wireless systems such as IEEE 802.11, LTE, and 5G NR. This compatibility allows the same signal to serve both sensing and communication functions, forming the foundation for ISAC systems. Radars based on OFDM signals leverage the transmission of multicarrier waveforms composed of a large number of orthogonal subcarriers. Each subcarrier can be independently modulated in amplitude and phase, allowing the encoding of both sensing and communication information. The orthogonality of subcarriers ensures that, in the absence of synchronization errors, there is no mutual interference among them, which simplifies the spectral design and enables efficient use of bandwidth. In addition, OFDM signals provide robustness to multipath propagation and frequency-selective fading. In radar operation, the environment is illuminated with the OFDM signal, and the reflections from objects are processed to

estimate range, velocity, and angle information. Range estimation in OFDM radar is achieved through analysis of the phase or frequency response of individual subcarriers, exploiting the inherent frequency diversity of the multicarrier structure. Doppler estimation can be performed by observing phase changes across consecutive OFDM symbols. Estimation of range and Doppler information generate a range-Doppler (RD) maps, which is promising for high-resolution target detection. However, OFDM radar also introduces certain challenges, including the need for precise synchronization between transmitter and receiver, high peak-to-average power ratio leading to non-linear distortion in power amplifiers, and increased computational complexity due to multi-dimensional Fourier-based processing.

Automotive radars based on FMCW signals employ a different modulation strategy, transmitting a continuous wave whose instantaneous frequency varies linearly with time, commonly referred to as a chirp. This linear frequency modulation allows the extraction of range and Doppler information from the frequency shifts between transmitted and received signals, after processing them using a mixer and a low-pass filter. In particular, the frequency slope of a chirp determines the range resolution, while the repetition of chirps in a single measurement enables Doppler estimation through phase variation across successive frames. Compared to pulsed or multicarrier systems, FMCW radar provides continuous illumination of the scene, high signal-to-noise ratio (SNR), and efficient use of available bandwidth, making it highly suitable for short and medium range sensing in real-time applications. In MIMO configurations, FMCW radar can synthesize a virtual antenna array, significantly increasing angular resolution without requiring a large physical aperture. The combination of frequency modulation and spatial diversity enables three-dimensional localization of targets in range, Doppler, and angle domains. Furthermore, FMCW processing architectures are well established and can be implemented efficiently on digital signal processors or field-programmable gate arrays (FPGAs), ensuring real-time operation. Compared to OFDM radar, FMCW systems exhibit lower implementation complexity, higher robustness to hardware imperfections, and reduced sensitivity to phase noise. This thesis considers an off-the-shelf automotive radar for experimentation and field trials. Signal processing techniques for automotive radar employing FMCW signals is described in the following.

2.2.2 Signal processing for automotive radars

A single measurement collected via a FMCW radar consists of multiple frequency modulated signals, referred to as chirp signals. Consider an FMCW MIMO radar transmitting linear chirps, i.e., sinusoidal signals with frequencies increasing linearly from f_m to $f_M = f_m + B$ over time, where B is the signal bandwidth. The frequency

of linear chirp signals increases in a time T_c , referred to as chirp duration, thus with a slope $S = B/T_c$. A measurement via FMCW MIMO radar consists of $N_c > 1$ chirp signals, each transmitted every T_c seconds. The chirp signal echoes after multipath propagation and target reflections can be collected at the receiver and then processed to obtain dynamics of the reflecting targets. The n -th received chirp at each antenna can be written as

$$s_R^{(n)}(t) = A_R^{(n)} \cos \left(2\pi f_m(t_n - t_d) + \pi S(t_n - t_d)^2 \right) \quad (2.32)$$

where $A_R^{(n)}$ is the n -th signal amplitude, $t_n = t - nT_c$, and t_d is a time delay representing the TOF. Specifically, the time delay is $t_d = 2(r + vt)/c$, where r and v are the range and velocity of the scatterer that reflected the signal and c is the propagation speed of electromagnetic waves.

At the receiver of the FMCW radar, the transmitted and received waveforms are multiplied using a mixer and their product is then filtered using a low-pass filter (LPF). The resulting signal is referred to as intermediate frequency (IF) signal and represents the difference of instantaneous frequency between the transmitted and received waveforms. The IF signal obtained by the transmission of the n -th chirp is given by

$$s_{IF}^{(n)}(t) = \frac{A_R^{(n)}}{2} \cos \left(2\pi(f_p t_n + f_D T_c) \right) \quad (2.33)$$

where $f_p = 2B r/cT_c$ and $f_D = 2f_m v/c$. Each IF signal can be sampled with period T_s in N_s samples, while the entire IF signal sequence with period T_c in N_c samples, i.e., one for each transmitted chirp. The sampling processes with periods T_s and T_c are referred to as fast sampling (FS) and slow sampling (SS), respectively. In MIMO radars, multiple antennas receive the same reflected signal with a time delay that depends on the AOA of the waveform. Therefore, samples of the IF signal sequence obtained via FS and SS can be arranged in a 3-D tensor, referred to as *radar data cube*, in which FS and SS represent the FS and SS dimensions, respectively, while the third dimension is due to multiple antennas. By processing the radar data cube via discrete Fourier transform (DFT) we obtain the RDA map, which consists of energy bins for each frequency shift considered over FS, SS, and angular dimension. Let $\tilde{f}_t$, $\tilde{f}_s$, and $\tilde{f}_a$ denote the frequency shifts over FS, SS, and angular dimension, respectively. The range is determined as $\hat{r} = c T_c \tilde{f}_t / (2B)$, velocity as $\hat{v} = c \tilde{f}_s / (2f_m)$, and AOA $\hat{\theta}$ such that $\sin(\hat{\theta}) = c \tilde{f}_a / (2\pi \delta f_m)$, where δ is the shift between antennas. The marginalization of the RDA map with respect to the angular dimension provides a range-velocity representation of the environment, referred to as RD map. The RD map is promising to distinguish between energy bins (i.e., frequency shifts) generated by the target presence from those produced by clutter and multipath propagation.

The reflected chirp signals contain context information, including the time-domain representation of micro-Doppler shifts generated by the target movements. A time-domain sequence encapsulating the micro-Doppler signal features generated by the target activity is promising to perform target classification. In fact, movements and activities generate specific sequence of micro-Doppler shifts, which are referred to as Doppler signatures. The time-domain representation of a Doppler signature contains information regarding the target dynamics preserving time correlation. Conventional methods for classification of tracked targets are based on either the motion velocity or the radar cross section (RCS). However, such methods do not consider the situation in which a vehicle can move slower than a pedestrian and that the RCS of a target varies due to its orientation. Compared to conventional classification features, such as RCS and target velocity, the time-domain analysis of Doppler signatures allows to characterize even the micro-movements performed by a specific target and their correlations over time. In particular, frequency-based metrics can provide information about the target identity and the activity performed since even the smallest movements produce Doppler shifts in the reflected signal.

Consider a measurement collected using a FMCW MIMO radar and arranged as a data cube. By applying the DFT to FS and angular dimensions, we obtain a sequence of N_c samples for each positional coordinate of the monitored area. Such a sequence of N_c samples is the time domain representation of the Doppler shifts and contains information regarding movements at a specific spatial position. The employment of FMCW radars operating at mmWaves is promising to collect high resolution Doppler shifts. Consider N_c chirps transmitted in a time frame ΔT_c , then the velocity resolution of an FMCW radar is defined as

$$v_{\text{res}} = \frac{\lambda_c}{2 \Delta T_c} \quad (2.34)$$

where $\lambda_c = 1/f_c$ and f_c is the chirp signal frequency. The high resolution available at the Doppler domain enables identification of multiple targets by performing classification of their type of moving, as well as to recognize activities employing the sequence of movements.

2.3 Wireless sensing based on camera technologies

The employment of NN-based approaches is promising to detect multiple targets from camera frames. In particular, convolutional neural networks (CNNs) allow to reveal, classify, and localize inside the image targets of interest, including humans, vehicles, and robots. Then, the homography transformation allows mapping the image space into real-world coordinates enabling location awareness. Unlike radar-based approaches,

which require clutter mitigation and Doppler processing for classification, camera-based approaches directly exploit visual information to obtain reliable sensing analytics.

At the discrete time instant k , we acquire a video frame from the camera. A CNN detector, namely You only look once (YOLO), is employed for processing the frame in order to detect multiple targets simultaneously. The YOLO model divides the input image into a grid of cells and, for each cell, predicts the presence of bounding boxes, confidence scores, and class probabilities. The inference step produces a set of detections

$$\mathcal{B}_k = \left\{ \mathbf{b}_k^{(j)} \right\}_{j=1}^{M_k} \quad (2.35)$$

where M_k is the number of detections at time k and each bounding box is defined as

$$\mathbf{b}_k^{(j)} = \left[u_{\min, k}^{(j)}, v_{\min, k}^{(j)}, u_{\max, k}^{(j)}, v_{\max, k}^{(j)} \right]^T \quad (2.36)$$

with $\mathbf{u} = [u_{\min, k}^{(j)}, u_{\max, k}^{(j)}]^T$ and $\mathbf{v} = [v_{\min, k}^{(j)}, v_{\max, k}^{(j)}]^T$ denoting the horizontal and vertical pixel coordinates, respectively. Each detection $\mathbf{b}_k^{(j)}$ is also associated with a class label $c_k^{(j)} \in \mathcal{C}_t$, where $\mathcal{C}_t$ is the set of semantic categories of interest (e.g., human, vehicle, robot) and a confidence score $\rho_k^{(j)} \in [0, 1]$, quantifying the likelihood that the bounding box corresponds to a real target of class $c_k^{(j)}$. Therefore, the output of YOLO at each instant k , is the the detection vector $\tilde{\mathbf{z}}_k^{(j)}$, which is given by

$$\tilde{\mathbf{z}}_k^{(j)} = [\mathbf{b}_k^{(j)}, c_k^{(j)}, \rho_k^{(j)}]^T \quad (2.37)$$

with $j = 1, 2, \dots, M_k$. The confidence threshold ρ_{thr} is applied to filter out unreliable detections. Only detections satisfying $\rho_k^{(j)} \geq \rho_{\text{thr}}$ are retained for subsequent localization.

While bounding boxes characterize the extent of the target in the image plane, they must be reduced to a single representative point to enable geometric mapping into the world coordinate system. Consider the bottomcenter point of each bounding box as the representative point, since it approximates the contact point of the target with the ground plane. Then, the estimated coordinate in the pixel domain is $\tilde{\mathbf{p}}_k^{(j)} = [u_k^{(j)}, v_k^{(j)}, z_k^{(j)}]^T$, where the position information are given by

$$u_k^{(j)} = \frac{1}{2}(u_{\min, k}^{(j)} + u_{\max, k}^{(j)}) \quad (2.38a)$$

$$v_k^{(j)} = v_{\max, k}^{(j)}, \quad (2.38b)$$

and the normalization term is $z_k^{(j)} = 1$.

2.3.1 Data transformation via homography

To transform the representative point from pixel coordinates into real-world coordinates, we employ a homography mapping. A homography matrix $\mathbf{H} \in \mathbb{R}^{3 \times 3}$ is a projective transformation that relates two planes, i.e., the image plane and the ground plane of the monitored area. Given the representative point $\tilde{\mathbf{p}}_k^{(j)}$ in homogeneous pixel coordinates, the corresponding world coordinate $\mathbf{p}_k^{(j)} = [x_k^{(j)}, y_k^{(j)}, z_k^{(j)}]^T$ is obtained as

$$\mathbf{p}_k^{(j)} = \mathbf{H} \tilde{\mathbf{p}}_k^{(j)}. \quad (2.39)$$

Since $\mathbf{p}_k^{(j)}$ is expressed in homogeneous coordinates, the real-world positions $x_k^{(j)}$ and $y_k^{(j)}$ are given by

$$x_k^{(j)} = \frac{1}{z_k^{(j)}} u_k^{(j)} \quad (2.40a)$$

$$y_k^{(j)} = \frac{1}{z_k^{(j)}} v_k^{(j)} \quad (2.40b)$$

where $z_k^{(j)}$ is the normalization term. By combining the detection step with the homography-based localization, each detection j at time k is characterized by the tuple

$$\hat{\mathbf{z}}_k^{(j)} = [\mathbf{p}_k^{(j)}, c_k^{(j)}, \rho_k^{(j)}] \quad (2.41)$$

where $\mathbf{p}_k^{(j)} \in \mathbb{R}^2$ is the world position, $c_k^{(j)}$ is the semantic class, and $\rho_k^{(j)}$ is the associated confidence.

This representation enables simultaneous localization and classification of multiple targets in real time. The output of the camera-based system can then be directly employed in higher-level modules, e.g., trajectory prediction or multi-sensor fusion with radar.

Remark – The camera-based approach presents complementary characteristics with respect to radar-based sensing. On the one hand, YOLO provides fine-grained semantic classification that is difficult to extract from radar reflections. On the other hand, camera-based localization is sensitive to illumination changes, occlusions, and calibration errors. Therefore, the integration of both sensing modalities enhances the robustness and reliability of multitarget perception.

2.3.2 Calibration of the homography matrix

The computation of the homography matrix $\mathbf{H}$ requires an offline calibration step that establishes the geometric relationship between the camera image plane and the

ground plane of the monitored area. The quality of the homography estimation depends critically on the spatial distribution of calibration points. Such points should cover the region of interest uniformly and avoid collinear configurations to reduce numerical instability. Let $\{(x_\ell, y_\ell)\}_{\ell=1}^{N_H}$ denote a set of reference points with known real-world coordinates expressed in a local ground plane reference system, and let $\{(u_\ell, v_\ell)\}_{\ell=1}^{N_H}$ denote the corresponding pixel coordinates observed in the camera image.

The mapping from the camera image to the real-world coordinate is given by the projective relation

$$\lambda_\ell \begin{bmatrix} u_\ell \\ v_\ell \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_\ell \\ y_\ell \\ 1 \end{bmatrix} \quad (2.42)$$

where $\ell \in \{1, 2, \dots, N_H\}$, $\lambda_\ell \in \mathbb{R}$ is a non-zero scale factor and $\mathbf{H} \in \mathbb{R}^{3 \times 3}$ is the homography matrix.

From the (2.42) we obtain two independent linear constraints per correspondence $((x_\ell, y_\ell), (u_\ell, v_\ell))$, with $\ell \in \{1, 2, \dots, N_H\}$. Therefore, a minimum of $N_H \geq 4$ non-collinear point correspondences is required to uniquely estimate $\mathbf{H}$ up to a scale factor. In practice, more correspondences are collected to improve robustness, and the estimation is solved as a least-squares problem. By following the direct linear transformation (DLT) method each correspondence $((x_\ell, y_\ell), (u_\ell, v_\ell))$ contributes to the linear system

$$\begin{bmatrix} 0 & 0 & 0 & -x_\ell & -y_\ell & -1 & v_\ell x_\ell & v_\ell y_\ell & v_\ell \\ x_\ell & y_\ell & 1 & 0 & 0 & 0 & -u_\ell x_\ell & -u_\ell y_\ell & -u_\ell \end{bmatrix} \tilde{\mathbf{h}} = 0 \quad (2.43)$$

where $\tilde{\mathbf{h}}$ is the flattened version of $\mathbf{H}$, i.e., $\tilde{\mathbf{h}} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]$. Stacking all $2N_H$ equations leads to a homogeneous system $A\mathbf{h} = 0$, with $\mathbf{h} \in \mathbb{R}^9$ the vectorized homography. The solution is obtained as the right singular vector of A corresponding to the smallest singular value.

Since the DLT algorithm is sensitive to scale differences in the data, normalization of both world and image coordinates is performed before solving (2.43). This consists of translating and scaling the coordinates such that they are centered at the origin and have an average distance of $\sqrt{2}$ from the origin. After solving for $\mathbf{H}$ in normalized coordinates, the transformation is denormalized to obtain the final homography matrix.

Finally, refinement of $\mathbf{H}$ can be performed by minimizing the geometric reprojection error through nonlinear optimization:

$$\mathbf{H} = \min_{\tilde{\mathbf{H}}} \sum_{\ell=1}^{N_H} \left\| \begin{bmatrix} u_\ell \\ v_\ell \end{bmatrix} - \pi \left(\tilde{\mathbf{H}} \begin{bmatrix} x_\ell \\ y_\ell \\ 1 \end{bmatrix} \right) \right\|^2 \quad (2.44)$$

where the function $\pi(\cdot)$ denotes the projection from homogeneous to Cartesian coordinates. This refinement step ensures improved accuracy in the presence of noise or measurement uncertainties.

Chapter remark

Reliable multitarget tracking relies on Bayesian filtering and accurate data association. Probabilistic frameworks based on recursive estimation of target position allow precise localization even in complex wireless environment affected by multipath and clutter conditions. Understanding radar and camera sensing principles reveals how heterogeneous data sources complement each other: radar provides robust positional information, while camera data enhances semantic understanding. These foundations are crucial for developing novel approaches for wireless sensing.

Chapter 3

Integrated tracking and classification of multiple targets

Context awareness via wireless sensing is guaranteed by the estimation of positional information and classification analytics (e.g., target identity and activity). However, tracking and classification of people, objects, and vehicles in complex wireless environments require the development of novel approaches for fulfilling the 3GPP requirements. This chapter introduces a framework for integrated tracking and classification of multiple targets using an FMCW MIMO radar operating at mmWaves. The proposed framework aims to design a complete processing chain for detecting, tracking, and classifying multiple non-collaborative targets in complex wireless environments. In the following, the employed signal processing techniques, including clutter mitigation, target detection, and the estimation of kinematic parameters from RDA maps, are described. Multitarget tracking filtering based on Bayesian recursion and data association, as well as classification methods employing CNNs for target identification and activity recognition are reported. The proposed framework for tracking and classification allows the deployment of accurate and reliable wireless sensing in complex wireless environments, such as industrial areas. Fig. 3.1 shows an illustration of the ITC framework.

3.1 System model

At each discrete time instant k , consider a random number $\mathbf{N}_k$ of targets, which can appear, disappear, and move freely. The k -th target is described by its positional state $\mathbf{x}_k^{(i)} \in \mathbb{R}^{n_x}$, with $i \in \{1, 2, \dots, N_k\}$ for a realization N_k of $\mathbf{N}_k$. The target positional state $\mathbf{x}_k^{(i)}$ consists of parameters such as position, velocity, and acceleration. The target class is denoted by $\mathbf{c}_k^{(i)} \in \mathcal{C}_t$, where $\mathcal{C}_t$ is the set of all possible identity classes. Positional states and classes of tracked targets are denoted by the multitarget positional

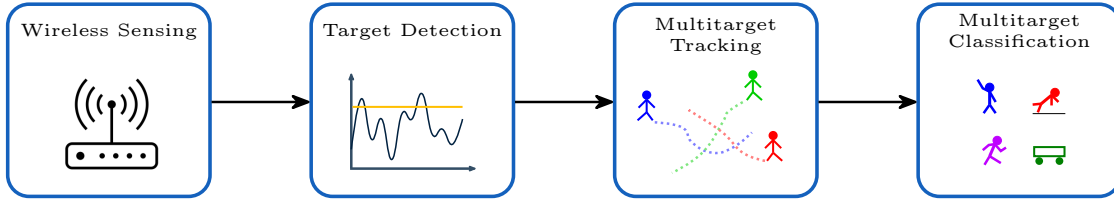


Figure 3.1: Illustration of the ITC framework.

state $\mathbf{x}_k = [\mathbf{x}_k^{(1)\top}, \mathbf{x}_k^{(2)\top}, \dots, \mathbf{x}_k^{(N_k)\top}]^\top$ and multitarget class $\mathbf{c}_k = [\mathbf{c}_k^{(1)}, \mathbf{c}_k^{(2)}, \dots, \mathbf{c}_k^{(N_k)}]^\top$, respectively.

Consider M_k measurements obtained from sensors at the current time k . Each measurement $\mathbf{z}_k^{(j)} \in \mathbb{R}^{m_z}$, with $j \in \{1, 2, \dots, M_k\}$ for a realization M_k of M_k , contains positional information to be employed as input for the MTT filtering. The measurement vector $\mathbf{z}_k = [\mathbf{z}_k^{(1)\top}, \mathbf{z}_k^{(2)\top}, \dots, \mathbf{z}_k^{(M_k)\top}]^\top$ contains all the detected measurements, including false alarms. In this work, we consider that: (i) all the measurements are independent of each other; (ii) a measurement is generated by at most one target; and (iii) a target can generate at most one measurement. In the remaining of this chapter, it has been described the signal processing techniques, including clutter mitigation and target detection methods, for collecting a set of measurements when employing a FMCW MIMO radar. Fig. 3.2 shows the block block diagram of the presented framework for integrated tracking and classification of multiple targets.

3.2 Detection technologies

In cluttered environments, target detection is challenging due to RD maps corrupted by signal distortion and noise. The proposed technique for clutter mitigation consists of an offline and an online phase. In the offline phase, the environmental clutter intensity distribution is estimated by the collection of measurements with no target in the area of interest. For each point with coordinate $\mathbf{p}$ in the RD or RDA map, a specific threshold is determined based on the clutter distribution previously estimated.¹ In the online phase, RD map elements are compared with the designed thresholds. If the magnitude of the RD map element under test is below the corresponding threshold, its value is forced to zero, otherwise is kept.

The threshold for a point at the coordinate $\mathbf{p}$ is defined by considering as Gaussian distributed its noise power $n_{\mathbf{p}}$, i.e., $n_{\mathbf{p}} \sim \varphi(n_{\mathbf{p}}; \bar{n}_{\mathbf{p}}, \sigma_{\mathbf{p}}^2)$. The distribution parameters can be estimated during the offline learning by the collection of multiple noisy measurements without any target in the monitored area. The threshold $\xi_{\mathbf{p}}$ for a point at

¹For notational convenience, we denote $\mathbf{p} \in \mathbb{R}^2$ a RD map coordinate and a $\mathbf{p} \in \mathbb{R}^3$ RDA map coordinate.

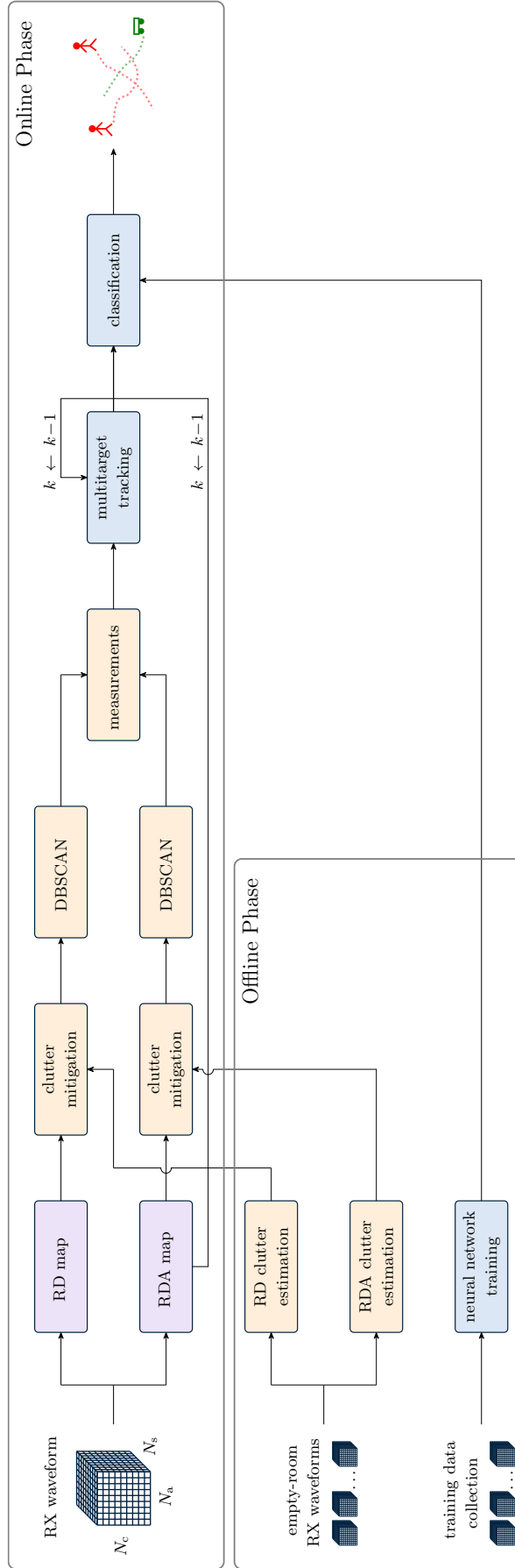


Figure 3.2: Block diagram of the presented framework for tracking and classification via single automotive radar operating at mmWaves. In particular, lilac blocks represent the signal processing phase, orange blocks the clutter mitigation and detection phase, and blue blocks the tracking and classification phase.

the coordinate $\mathbf{p}$ is given by

$$\xi_{\mathbf{p}} = \bar{n}_{\mathbf{p}} + \alpha_{\mathbf{p}} \sigma_{\mathbf{p}} \quad (3.1)$$

where $\alpha_{\mathbf{p}} \in \mathbb{R}^+$ tunes the false alarm and misdetection probabilities. A higher $\alpha_{\mathbf{p}}$ reduces the false alarm probability, while increases the number of misdetections, and vice-versa.

The high range and velocity resolutions of mmWave FMCW radars and the noise introduced by multipath propagation generate a representation of each detected target as an extended object, i.e., characterized by a set of range-velocity coordinates corresponding to an over threshold magnitudes. The DBSCAN algorithm allows to group such coordinates in clusters by leveraging their density in the RD map [135]. A weighted mean of the cluster elements provides a single-value estimate, referred to as centroid, which consists of a range-velocity coordinate. In particular, by denoting as Ω_j the j -th cluster of coordinates, with $j \in \{1, 2, \dots, M\}$ and M representing the number of clusters obtained by computing the DBSCAN algorithm, the centroid $\boldsymbol{\mu}_j$ is given by

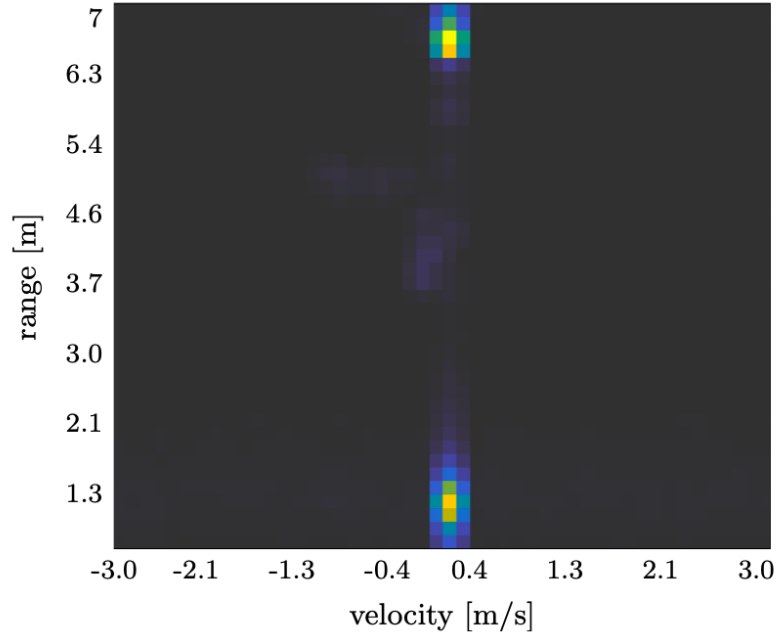
$$\boldsymbol{\mu}_j = \frac{\sum_{\mathbf{p} \in \Omega_j} P_{\mathbf{p}} \mathbf{p}}{\sum_{\mathbf{p} \in \Omega_j} P_{\mathbf{p}}} \quad (3.2)$$

where $\mathbf{p}$ denotes a range-velocity coordinate and $P_{\mathbf{p}}$ is the magnitude of the corresponding energy bin in the RD map. The centroid $\boldsymbol{\mu}_j$ is the estimated range-velocity coordinate for representing the cluster j . At a range-velocity coordinate $\boldsymbol{\mu}_j$, the reflection AOA is obtained by considering the angle with highest magnitude in the RDA map.

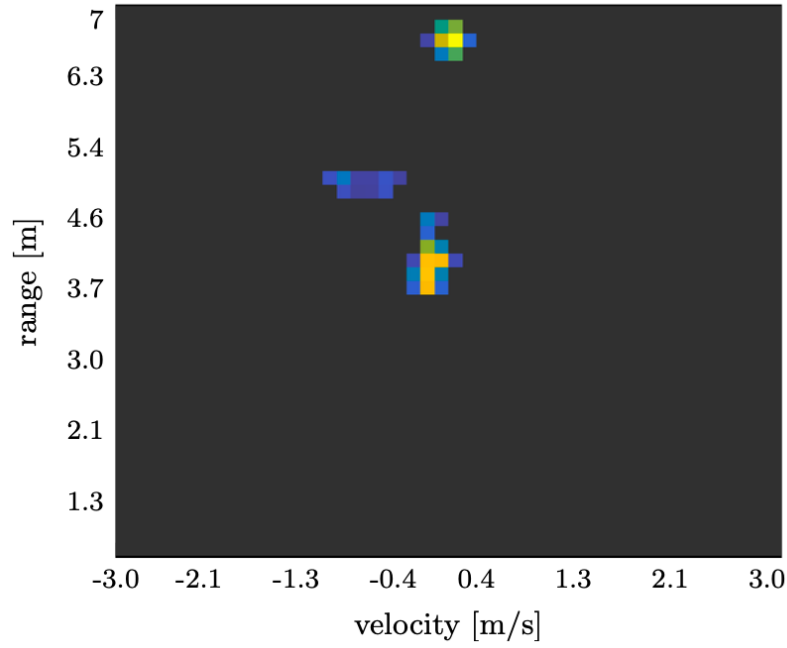
The knowledge of the target positional states at the previous time instant is exploited to further reduce the number of misdetections and false alarms. From the RDA map, a sub-tensor for each previously estimated positional state can be extracted and, then, processed via the DBSCAN algorithm to enlarge the set of detections. The $\mathbf{M}_k$ measurements are composed by the union of centroids collected evaluating both the RD map and the RDA map sub-tensors. Therefore, the set of measurements consist of range-velocity-AOA coordinates, which can be used as input of the MTT filtering.

3.3 Tracking technologies

Tracking of multiple device-free targets requires data association, which is the combinatorial problem of determining (i) which target generates a given measurement, (ii) if the measurement is due to false alarm, and (iii) if a target is misdetected. Let $\boldsymbol{\theta}_k^{(i)}$



(a) RD map before clutter mitigation.



(b) RD map before after mitigation.

Figure 3.3: Heatmap of a RD map collected with 3 targets in the environment. Representation before and after filtering using the proposed approach for clutter mitigation.

be the data association variable denoting the assignment of the measurement j , with $j \in \{1, 2, \dots, M_k\}$, to the target i , with $i \in \{1, 2, \dots, N_k\}$, at the time instant k , as

follows

$$\theta_k^{(i)} = \begin{cases} j & \text{if target } i \text{ is associated to a measurement } j \\ 0 & \text{if target } i \text{ is associated to no measurement.} \end{cases} \quad (3.3)$$

In the presence of multiple targets, we define the multitarget data association vector $\boldsymbol{\theta}_k = [\theta_k^{(1)}, \theta_k^{(2)}, \dots, \theta_k^{(N_k)}]^T$ providing the measurement-to-target associations for all the tracked targets. In particular, the multitarget data association vector is such that $\boldsymbol{\theta}_k \in \Theta_k$, where Θ_k is the set containing all possible vectors of valid multitarget data associations. A valid multitarget data association requires that (i) each target is either associated to a measurement or misdetections and (ii) any pair of detected targets is not associated to the same measurement. To compute data association and MTT filtering analytically, we exploit the KF solution considering linear and Gaussian system models.²

Data association can be seen as an optimization problem, where a data association probability $P_a(\theta_k^{(i)})$, for linear and Gaussian modeling, is given by

$$P_a(\theta_k^{(i)}) = \begin{cases} P_d(\mathbf{x}_k^{(i)}) \varphi(\mathbf{z}_k^{(\theta_k^{(i)})}; \mathbf{H}_k \tilde{\mathbf{x}}_k^{(i)}, \boldsymbol{\Sigma}_k) & \theta_k^{(i)} \neq 0 \\ 1 - P_d(\mathbf{x}_k^{(i)}) & \theta_k^{(i)} = 0 \end{cases} \quad (3.4)$$

where $P_d(\cdot)$ is the detection probability and $\mathbf{H}_k$ is the linear measurement model, while $\tilde{\mathbf{x}}_k^{(i)}$ and $\boldsymbol{\Sigma}_k$ are the positional prediction and the innovation covariance of the KF, respectively. The data association costs

$$l_{i,j} = -P_a(\theta_k^{(i)} = j) \quad (3.5)$$

can be arranged in a matrix $\mathbf{L}$, where rows represent targets and columns represent both measurements and misdetections. By processing $\mathbf{L}$ using the Murty's algorithm [137], we define $\hat{\Theta}_k \subseteq \Theta_k$ representing the set of the Q -best data association hypotheses obtained evaluating the data association costs.

The JPDA algorithm leverages a soft-decision approach for data association by accounting for the uncertainty in matching measurements with tracked targets [138]. In particular, JPDA assigns an association score $\beta_k^{(i,j)}$ to each possible measurement-to-target association, based on how the measurement aligns with the predicted positional

²For nonlinear non-Gaussian system models, computationally feasible approximate algorithms include variants of the KF such as the extended KF and the unscented KF [136]. Both the extended KF and the unscented KF are based on the Kalman equations but perform different approximations of the system models.

state of the target. The association score can be calculated as

$$\beta_k^{(i,j)} = \sum_{\theta_k \in \{\theta_k \in \hat{\Theta}_k, \theta_k^{(i)} = j\}} P_a(\theta_k^{(i)}) \quad (3.6)$$

where $j \in \{0, 1, \dots, M_k\}$. Based on the association score, the distribution of the multitarget positional state for independent targets is given by

$$f(\mathbf{x}_k | \mathbf{z}_{1:k}) = \prod_{i=1}^{N_k} \sum_{j=0}^{M_k} \beta_k^{(i,j)} f(\mathbf{x}_k^{(i)} | \mathbf{z}_k^{(j)}) . \quad (3.7)$$

where $f(\mathbf{x}_k^{(i)} | \mathbf{z}_k^{(j)})$ can be calculated employing the KF for linear and Gaussian models.

Appearing and disappearing targets require the initialization and the deletion of new and existing tracks, respectively. Track initialization is performed from a sequence of measurements collected over time as follows. At the first time instant, all the measurements are considered as potential tracks, i.e., each generates a track hypothesis. At the next time instants consider the greediest data association hypothesis. The measurements that are not assigned to any existing track are either associated to a track hypothesis or considered as new potential tracks. Finally, if a track hypothesis achieves a length of K measurements is validated as real track and processed via the JPDA algorithm. Given that the positional states are represented using a Gaussian density, the covariance matrix serves as a crucial indicator of the plausibility of the track hypothesis. In particular, until the trace of the covariance matrix is below a deleter threshold, the hypothesis is valid. By adjusting the deleter threshold, the same approach can be employed to delete existing tracks processed via the JPDA algorithm. The choice of the threshold value is a key aspect for achieving accuracy in track removal tasks. Typically, the best threshold is estimated empirically through several trials and exploiting the knowledge of the operative conditions, including the environment, the sensing device, and the types of targets in the monitored area.

3.4 Classification technologies

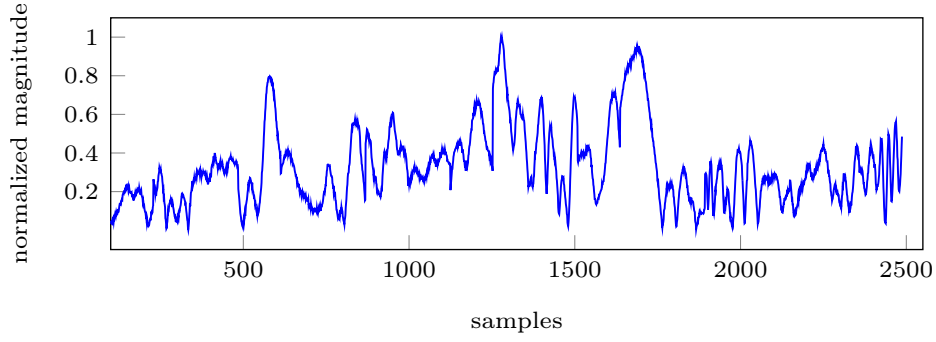
The analytical characterization of complex wireless environments is challenging due to the stochastic nature of the channel, especially in the presence of multiple targets and clutter. Specifically, the presence of obstacles and scatter objects make difficult the mathematical characterization of a mapping function which inputs the Doppler signature and outputs the activity performed or the target identity. Machine learning and, in particular, deep learning, is promising for the classification of Doppler signatures extracted from reflected signals. Approaches based on NNs allow to determine the most

significant signal features to perform an accurate classification without requiring a mathematical formulation of the waveform propagation. During the offline phase of the proposed approach (i.e., training phase) by the processing of training data, a NN infers a function which, given a Doppler signature as input, outputs the probability of belonging to each class among all the considered. To achieve reliable classification performance in several environmental conditions, an important aspect to consider is the training generalization. Indeed, it is important that a trained classification model is capable of operating in scenarios not explored during the training phase, e.g., different people in different spatial positions of the environment. The issue of forecasting capability limited to the training data is referred to as overfitting. By employing CNNs and exploiting data augmentation and bagging techniques, high classification accuracy can be achieved even in environments not explored during training. Next, we describe the developed methods for target identification and activity recognition.

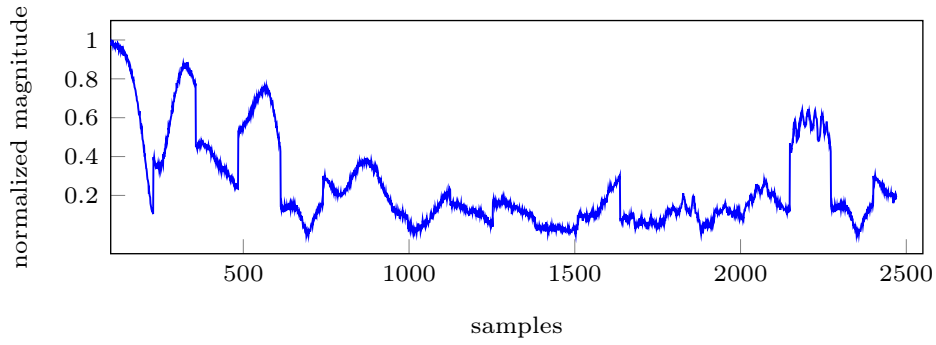
3.4.1 Target identification

Identification requires the extraction of specific signal features for each target tracked. In particular, MTT filtering provides range-angle coordinates, which can be exploited to determine specific classification analytics for each target from reflected chirp signals. Consider a collected radar data cube, by applying the DFT to the FS and angular dimensions, we obtain range-angle coordinates, each characterized by the N_c samples of the SS dimension. Such samples provide a time-domain representation of the micro-Doppler shifts generated by movements at a specific range-angle coordinate. Fig. 3.4 shows the time-domain representation of the micro-Doppler signatures generated by a human target and an industrial vehicle.

We consider a classification problem for target identification taking as input a sequence of WN_c time-domain samples collected by exploiting target positions estimated via MTT. In particular, the parameter $W \in \mathbb{N}$ defines the duration of the observation window in which the classification analytics are collected. The entire sequence of WN_c samples is organized as a linear array to preserve the correlations over time. Increasing W enables performing a more accurate characterization of the target motion. Therefore, by collecting multiple Doppler shift signals over a time frame, we obtain a Doppler signature for each target containing the micro-movement characterization that can be processed employing a NN for a classification task. In particular, 1D-CNNs are an effective tool for the classification of time series. Compared to conventional approaches based on recurrent NNs, CNNs guarantee higher accuracy thanks to the capability to determine temporal correlations spread over a longer time period.



(a) Doppler signature generated by the movements of a human target.



Doppler signature generated by the movements of an industrial vehicle.

Figure 3.4: Representation of Doppler signatures used as input for classification generated by human target and industrial vehicle, respectively.

3.4.2 Activity recognition

Activity recognition is performed considering the location awareness provided by a localization algorithm. As mentioned above, a sequence of Doppler signals can be collected over time from the RDA map relying on the target positional state to obtain the Doppler signature of a completed activity. The classification of activities requires knowledge about the time correlation among the obtained Doppler shifts at different frequency-variation scales. The CWT applied to a time domain signal produces a time-frequency image which allows to extract local frequency spectrum and temporal information simultaneously in compliance with the Heisenberg uncertainty theorem. Specifically, the time-frequency resolution of the transformation can be adapted to capture small variations in time when high frequency are considered and vice-versa. Then, the output of the CWT is a time-frequency image containing the sequence of frequency variations over time extracted using a suitable time-frequency resolution. The CWT is defined as

$$\mathcal{W}_{\mathbf{x}}(u, s) = \sum_{t=0}^{N_c} \mathbf{x}[t] \psi_{u,s}(t) dt \quad (3.8)$$

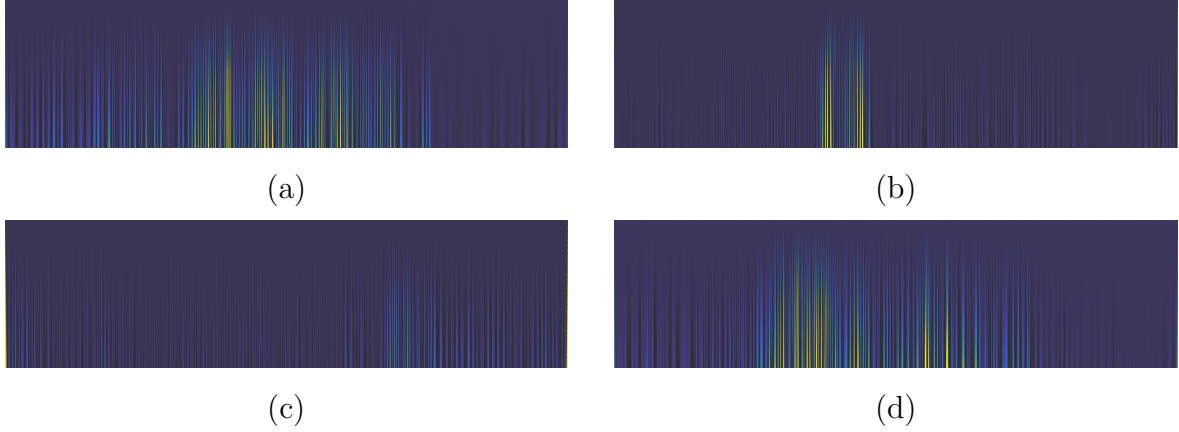


Figure 3.5: Time-frequency images obtained by processing Doppler signatures via CWT. The x-axis represents time domain, y-axis the frequency domain, and the color range (from blue to yellow) the normalized magnitude of the CWT. In particular, time-frequency images consider: (a) human presence, (b) human falling, (c) no human, and (d) arm movement.

where $\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$ and $\psi \in \mathbf{L}^2(\mathbb{R})$ is referred to as the mother wavelet. In particular, u and s represent the time translation and the scale parameter of the employed mother wavelet. Changing the scale s allow to analyze different signal frequencies. By using the CWT, we extract the temporal correlation characteristics of the Doppler signature to be employed as input for the NN. Fig. 3.5 shows time-frequency images obtained by processing via CWT the Doppler signatures originated from different activities.

Chapter remark

Integrated tracking and classification demonstrate that positional and classification tasks are mutually beneficial. Radar-based signal processing, data association, and neural-based classification form a promising framework for reliable context awareness. Exploiting Bayesian recursion for tracking and NNs for Doppler features classification provides accurate motion and identity estimation. Derivation of closed-form solution for the Bayesian recursion allows real-time device-free sensing in complex wireless environments.

Chapter 4

Joint tracking and classification of multiple targets

Information superiority in complex wireless environment is guaranteed by the introduction of the JTC paradigm. Sensing of things baed on JTC approaches enable the simultaneous estimation of position and class of multiple unconnected targets. Unlike the integrated framework, JTC approaches rely on a joint probabilistic formulation where tracking and classification are performed within a unified Bayesian recursion. The chapter presents the derivation of the JTC filtering equations, discusses the conditions under which a closed-form solution can be obtained, and details the joint data association that exploit both positional and classification information. An implementation of the JTC framework using an automotive radar operating at mmWaves is proposed to demonstrate the advantages of joint estimation in terms of accuracy and reliability. Fig. 4.1 shows an illustration of the JTC framework.

4.1 Joint tracking and classification filtering

In several scenarios multiple kinds of targets performing different types of activities need to be tracked and classified at the same time. For this purpose, it can be considered the multitarget class $\mathbf{c}_k = [\mathbf{c}_k^{(1)}, \mathbf{c}_k^{(2)}, \dots, \mathbf{c}_k^{(N_k)}]^T \in \mathcal{P}_k$, where $\mathbf{c}_k^{(i)} \in \mathcal{C}_t$, with $i \in \{1, 2, \dots, N_k\}$, is the class of the i -th target and $\mathcal{P}_k$ is the set of all the valid classification vectors. Specifically, a valid classification vector requires that each tracked target is associated to one, and only one, class from $\mathcal{C}_t$. Therefore, the set $\mathcal{P}_k$ contains all the permutation with repetition of N_k elements from the class set $\mathcal{C}_t$. See Tab. 2.1 for the variable definition summary.

To perform tracking and classification simultaneously, consider a measurement $\mathbf{z}_k^{(j)} \in \mathbb{R}^{m_z}$, with $j \in \{1, 2, \dots, M_k\}$, which represents both positional information and the classification features of a detection. In particular, we define a measurement

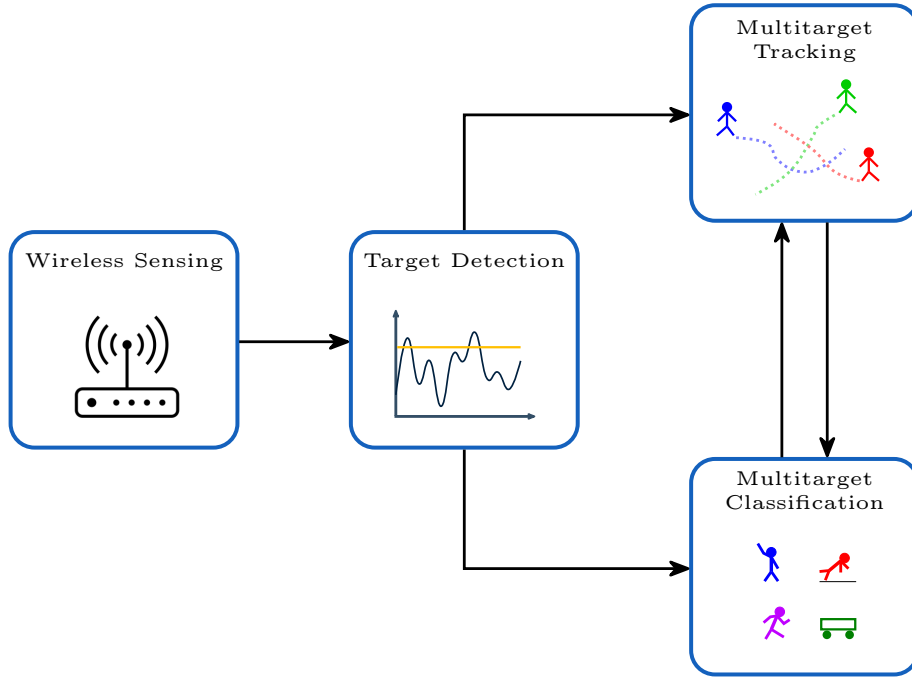


Figure 4.1: Illustration of the JTC framework.

$\mathbf{z}_k^{(j)}$ as a vector of dimension m_z , containing the target mobility parameters (e.g., position and velocity) and signal features which can be employed as classification metric (e.g., μ D shifts).

In the remaining of this section, we, first, propose the JTC filtering and a closed-form solution for an amenable implementation. Then, we introduce the JTC data association to perform measurement-to-target assignments considering the information superiority given by the classification.

4.1.1 Bayesian filtering

Simultaneous location awareness and classification require a joint estimator for $\mathbf{x}_k$ and $\mathbf{c}_k$ given the measurements $\mathbf{z}_{1:k}$. Integrating the multitarget class $\mathbf{c}_k$ into the MTT filtering leads to JTC filtering. In JTC filtering, prior and posterior densities estimate jointly target position and class of multiple targets. In particular, the JTC posterior density results¹

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) = \frac{f(\mathbf{z}_k | \mathbf{x}_k, \mathbf{c}_k) f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k-1})}{f(\mathbf{z}_k | \mathbf{z}_{1:k-1})} \quad (4.1)$$

where

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k-1}) = \quad (4.2)$$

¹Note that all the assumptions made in Section 2.1 are still valid.

$$\sum_{\mathbf{c}_{k-1} \in \mathcal{P}_{k-1}} \int f(\mathbf{x}_k, \mathbf{c}_k, \mathbf{x}_{k-1}, \mathbf{c}_{k-1} | \mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1}$$

is the JTC prediction density and

$$f(\mathbf{z}_k | \mathbf{z}_{1:k-1}) = \sum_{\mathbf{c}_k \in \mathcal{P}_k} \int f(\mathbf{z}_k | \mathbf{x}_k, \mathbf{c}_k) \times f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k-1}) d\mathbf{x}_k. \quad (4.3)$$

is the normalization term. The JTC measurement likelihood $f(\mathbf{z}_k | \mathbf{x}_k, \mathbf{c}_k)$ defines the probability that the positional measurement vector $\mathbf{z}_k$ is generated from the multitarget positional state $\mathbf{x}_k$ classified as $\mathbf{c}_k$. Thus, the JTC observation equation is

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{c}_k, \mathbf{w}_k) \quad (4.4)$$

where $\mathbf{h}(\cdot, \cdot, \cdot)$ is a nonlinear transformation and $\mathbf{w}_k$ is the observation noise.

Consider the multitarget positional state $\mathbf{x}_k$ and the multitarget class $\mathbf{c}_k$ as a joint state, which evolves as a first-order Markov process. In particular, the JTC Markov transition density $f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{x}_{k-1}, \mathbf{c}_{k-1})$ describes the probability that $\mathbf{x}_{k-1}$ and $\mathbf{c}_{k-1}$ evolve to $\mathbf{x}_k$ and $\mathbf{c}_k$, respectively, according to the transition equation given by

$$\mathbf{x}_k, \mathbf{c}_k = \gamma(\mathbf{x}_{k-1}, \mathbf{c}_{k-1}, \mathbf{v}_k) \quad (4.5)$$

where $\gamma(\cdot, \cdot, \cdot)$ is a nonlinear transformation and $\mathbf{v}_k$ is the transition noise. By the Chapman-Kolmogorov equation the JTC prediction density can be defined as

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k-1}) = \sum_{\mathbf{c}_{k-1} \in \mathcal{P}_{k-1}} \int f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{x}_{k-1}, \mathbf{c}_{k-1}) f(\mathbf{x}_{k-1}, \mathbf{c}_{k-1} | \mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1} \quad (4.6)$$

where $f(\mathbf{x}_{k-1}, \mathbf{c}_{k-1} | \mathbf{z}_{1:k-1})$ is the joint posterior at the previous time instant.

To provide a closed-form solution for the JTC recursion let consider (i) independency between the classifier and the position estimator² and (ii) linear and Gaussian models to exploit the KF solution described for the MTT recursion. In particular, the JTC Markov transition density for independent estimators results

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{x}_{k-1}, \mathbf{c}_{k-1}) = f(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{c}_{k-1}) f(\mathbf{c}_k | \mathbf{x}_{k-1}, \mathbf{c}_{k-1}) \quad (4.7)$$

and represents the joint class and positional state transition over time as a product of independent transitions. Hence, the positional state $\mathbf{x}_{k-1}$, given the class $\mathbf{c}_{k-1}$, evolves

²Note that classification still depends on the target position, and vice-versa, due to the observation generated by the target.

to $\mathbf{x}_k$ in accordance with a positional transition model. The multitarget class $\mathbf{c}_{k-1}$, given the class $\mathbf{x}_{k-1}$, evolves to $\mathbf{c}_k$ in accordance with a class transition model. We introduce the *constant class* transition model, i.e., a liner model to predict the multitarget class $\mathbf{c}_k$ given $\mathbf{c}_{k-1}$. Specifically, the constant class transition model assumes that the multitarget class is not varying over two subsequent time instants, at any time k . Therefore, the class transition density is given by

$$f(\mathbf{c}_k|\mathbf{x}_{k-1}, \mathbf{c}_{k-1}) = f(\mathbf{c}_k|\mathbf{c}_{k-1}) \quad (4.8)$$

such that

$$f(\mathbf{c}_k|\mathbf{c}_{k-1}) = \begin{cases} 1 & \text{if } \mathbf{c}_k = \mathbf{c}_{k-1} \\ 0 & \text{otherwise.} \end{cases} \quad (4.9)$$

and the JTC prediction can be written as

$$\begin{aligned} f(\mathbf{x}_k, \mathbf{c}_k|\mathbf{z}_{1:k-1}) &= f(\mathbf{c}_k|\mathbf{c}_{k-1}) \\ &\times \sum_{\mathbf{c}_{k-1} \in \mathcal{P}_{k-1}} \int f(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{c}_{k-1}) f(\mathbf{x}_{k-1}, \mathbf{c}_{k-1}|\mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1}. \end{aligned} \quad (4.10)$$

The Gaussian JTC positional transition density can be defined as

$$f(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{c}_{k-1}) = \varphi\left(\mathbf{x}_k; \mathbf{\Gamma}_k^{(\mathbf{c}_{k-1})} \mathbf{x}_{k-1}, \mathbf{V}_{p,k}\right) \quad (4.11)$$

where $\mathbf{\Gamma}_k^{(\mathbf{c}_{k-1})}$ is the linear motion model which best characterizes the target kinematics according to the target classification and $\mathbf{V}_{p,k}$ is the positional noise covariance. Considering a Gaussian JTC posterior density at previous time instant, the JTC prediction results

$$f(\mathbf{x}_k, \mathbf{c}_k|\mathbf{z}_{1:k-1}) = \varphi\left(\mathbf{x}_k; \tilde{\mathbf{x}}_k, \tilde{\mathbf{P}}_k\right) \quad (4.12)$$

where

$$\tilde{\mathbf{c}}_k = \hat{\mathbf{c}}_{k-1} \quad (4.13a)$$

$$\tilde{\mathbf{x}}_k = \mathbf{\Gamma}_k^{(\tilde{\mathbf{c}}_k)} \hat{\mathbf{x}}_{k-1} \quad (4.13b)$$

$$\tilde{\mathbf{P}}_k = \mathbf{V}_{k-1} + \mathbf{\Gamma}_k^{(\tilde{\mathbf{c}}_k)} \hat{\mathbf{P}}_{k-1} \left(\mathbf{\Gamma}_k^{(\tilde{\mathbf{c}}_k)}\right)^T. \quad (4.13c)$$

Leveraging the conditional probability properties, the joint posterior density, defined in (4.1), can be defined as

$$f(\mathbf{x}_k, \mathbf{c}_k|\mathbf{z}_{1:k}) = \frac{f(\mathbf{c}_k|\mathbf{z}_{1:k}, \mathbf{x}_k) f(\mathbf{z}_k|\mathbf{x}_k) f(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{f(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \quad (4.14a)$$

$$= \frac{\frac{f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k})}{f(\mathbf{x}_k | \mathbf{z}_{1:k})} f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{z}_{1:k-1})}{f(\mathbf{z}_k | \mathbf{z}_{1:k-1})}. \quad (4.14b)$$

Then, the normalization term is given by

$$f(\mathbf{z}_k | \mathbf{z}_{1:k-1}) = \sum_{\mathbf{c}_k \in \mathcal{P}_k} \int \frac{f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k})}{f(\mathbf{x}_k | \mathbf{z}_{1:k})} f(\mathbf{z}_k | \mathbf{x}_k) \times f(\mathbf{x}_k | \mathbf{z}_{1:k-1}) d\mathbf{x}_k. \quad (4.15)$$

where $f(\mathbf{z}_k | \mathbf{x}_k)$ is the positional measurement likelihood and $f(\mathbf{x}_k | \mathbf{z}_{1:k-1})$ the positional prediction. Since it has been assuming independency between the classifier and the position estimator, the (4.14b) results

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) = \frac{f(\mathbf{c}_k | \mathbf{z}_{1:k}) f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{z}_{1:k-1})}{f(\mathbf{z}_k | \mathbf{z}_{1:k-1})} \quad (4.16a)$$

$$= \frac{f(\mathbf{c}_k | \mathbf{z}_{1:k})}{\sum_{\mathcal{P}_k} f(\mathbf{c}_k | \mathbf{z}_{1:k})} \times \frac{f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{z}_{1:k-1})}{\int f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{z}_{1:k-1}) d\mathbf{x}_k}. \quad (4.16b)$$

Considering linear and Gaussian models, from (2.4) the JTC posterior can be written as

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) = \frac{f(\mathbf{c}_k | \mathbf{z}_{1:k})}{\sum_{\mathcal{P}_k} f(\mathbf{c}_k | \mathbf{z}_{1:k})} \varphi(\mathbf{x}_k; \hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k) = \bar{f}(\mathbf{c}_k | \mathbf{z}_{1:k}) \varphi(\mathbf{x}_k; \hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k) \quad (4.17a)$$

where

$$\mathbf{K}_k = \tilde{\mathbf{P}}_k \left(\mathbf{H}_k^{(\hat{\mathbf{c}}_k)} \right)^T \Sigma_k^{-1} \quad (4.18a)$$

$$\Sigma_k = \mathbf{W}_k + \mathbf{H}_k^{(\hat{\mathbf{c}}_k)} \tilde{\mathbf{P}}_k \left(\mathbf{H}_k^{(\hat{\mathbf{c}}_k)} \right)^T \quad (4.18b)$$

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + \mathbf{K}_k \left(\mathbf{z}_k - \mathbf{H}_k^{(\hat{\mathbf{c}}_k)} \tilde{\mathbf{x}}_k \right) \quad (4.18c)$$

$$\hat{\mathbf{P}}_k = [\mathbf{I} - \mathbf{K}_k \mathbf{H}_k] \tilde{\mathbf{P}}_k. \quad (4.18d)$$

Specifically, $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{P}}_k$ are given by the Chapman-Kolmogorov prediction, while $\mathbf{H}_k^{(\hat{\mathbf{c}}_k)}$ is the linear measurement model based on the target classification. The multitarget class $\hat{\mathbf{c}}_k$ is estimated by using a generic classifier, which maps signal features in target classes. For independent targets the joint Bayes posterior density can be written as

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) = \prod_{i=1}^{N_k} \bar{f} \left(c_k^{(i)} | \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})} \right) \varphi \left(\mathbf{x}_k^{(i)}; \hat{\mathbf{x}}_k^{(i)}, \hat{\mathbf{P}}_k^{(i)} \right) \quad (4.19a)$$

where the weight $\bar{f}(c_k^{(i)} | \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})})$ is the normalized classification probability given $\mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}$, i.e., the history of measurement-to-target assignments for the i -th target. The mean $\hat{\mathbf{x}}_k^{(i)}$ and covariance $\hat{\mathbf{P}}_k^{(i)}$ characterize the target positional state and can be obtained leveraging the KF solution.

In the JTC posterior, the positional density is weighted by the classification scores, which are obtained evaluating signal features associated to each target by data association. In the JTC filtering, data association must consider both positional and class information to figure out the most likely measurement-to-target assignments providing the best estimate of $\mathbf{x}_k$ and $\mathbf{c}_k$, i.e., the multitarget joint state.

4.1.2 Data association

Joint data association is required in JTC filtering to estimate the multitarget joint state by considering positional and class information. In particular, the posterior density in the JTC filtering leverages the data association results to update the target position and class given the collected measurements. The unnormalized joint posterior density can be written as

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) \propto f(\mathbf{c}_k | \mathbf{z}_{1:k}) f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{z}_{1:k-1}) \quad (4.20a)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} f(\mathbf{x}_k, \mathbf{c}_k, \mathbf{z}_k, \boldsymbol{\theta}_k | \mathbf{z}_{1:k-1}) \quad (4.20b)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} f(\mathbf{c}_k | \mathbf{x}_k, \mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \times f(\mathbf{x}_k, \mathbf{z}_k, \boldsymbol{\theta}_k | \mathbf{z}_{1:k-1}) \quad (4.20c)$$

and given the independency between $\mathbf{c}_k$ and $\mathbf{x}_k$ results

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) \propto \sum_{\boldsymbol{\theta}_k \in \Theta_k} f(\mathbf{c}_k | \mathbf{z}_{1:k}, \boldsymbol{\theta}_k) f(\mathbf{x}_k, \mathbf{z}_k, \boldsymbol{\theta}_k | \mathbf{z}_{1:k-1}). \quad (4.21)$$

For independent targets

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) \propto \sum_{\boldsymbol{\theta}_k \in \Theta_k} \prod_{i=1}^{N_k} f(c_k^{(i)} | \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) \times f(\mathbf{x}_k^{(i)}, \mathbf{z}_k^{(\theta_k^{(i)})}, \theta_k^{(i)} | \mathbf{z}_{1:k-1}^{(\theta_{1:k-1}^{(i)})}) \quad (4.22)$$

and given a data association hypothesis $\boldsymbol{\theta}_k$, then, $\theta_k^{(i)}$ and $\mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}$, can be seen as parameters of a function $g(\cdot)$ defined as

$$g(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) = f(c_k^{(i)} | \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}) f(\mathbf{x}_k^{(i)}, \mathbf{z}_k^{(\theta_k^{(i)})}, \theta_k^{(i)} | \mathbf{z}_{1:k-1}^{(\theta_{1:k-1}^{(i)})}). \quad (4.23)$$

The normalized version of $g(\cdot)$, denoted as $\bar{g}(\cdot)$, is given by

$$\bar{g}\left(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) = \frac{g\left(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right)}{w\left(\mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right)} \quad (4.24)$$

where

$$w\left(\mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) = \sum_{c_k^{(i)} \in \mathcal{C}_t} \int g\left(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) d\mathbf{x}_k^{(i)} \quad (4.25)$$

is referred to as joint data association weight given the measurements $\mathbf{z}_{1:k}^{(\theta_k^{(i)})}$ and the assignment $\theta_k^{(i)}$ for the i -th target. Thus, from (4.24), the joint posterior can be written as

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) \propto \sum_{\boldsymbol{\theta}_k \in \Theta_k} \prod_{i=1}^{N_k} w\left(\mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \times \bar{g}\left(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \quad (4.26a)$$

$$= \sum_{\boldsymbol{\theta}_k \in \Theta_k} w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \times \prod_{i=1}^{N_k} \bar{g}\left(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \quad (4.26b)$$

and employing the normalized data association weight $\bar{w}(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k)$, which is defined in (2.22) and referred to as joint data association density, then

$$f(\mathbf{x}_k, \mathbf{c}_k | \mathbf{z}_{1:k}) = \sum_{\boldsymbol{\theta}_k \in \Theta_k} \bar{w}(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \times \prod_{i=1}^{N_k} \bar{g}\left(\mathbf{x}_k^{(i)}, c_k^{(i)}; \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right) \quad (4.27a)$$

The most likely data association vector can be found out solving the optimization problem described in Section 2.1.2 for the conventional MTT. In particular, the optimization problem aims to find $\boldsymbol{\theta}_k^*$, such that

$$\boldsymbol{\theta}_k^* = \arg \max_{\boldsymbol{\theta}_k \in \Theta_k} w(\mathbf{z}_{1:k}, \boldsymbol{\theta}_k) \quad (4.28a)$$

$$= \arg \min_{\boldsymbol{\theta}_k \in \Theta_k} \sum_{i=1}^{N_k} -\log\left(w\left(\mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right)\right). \quad (4.28b)$$

Therefore, the cost matrix elements $l^{(i,j)} \in \mathbf{L}$ can be specialized for the JTC data

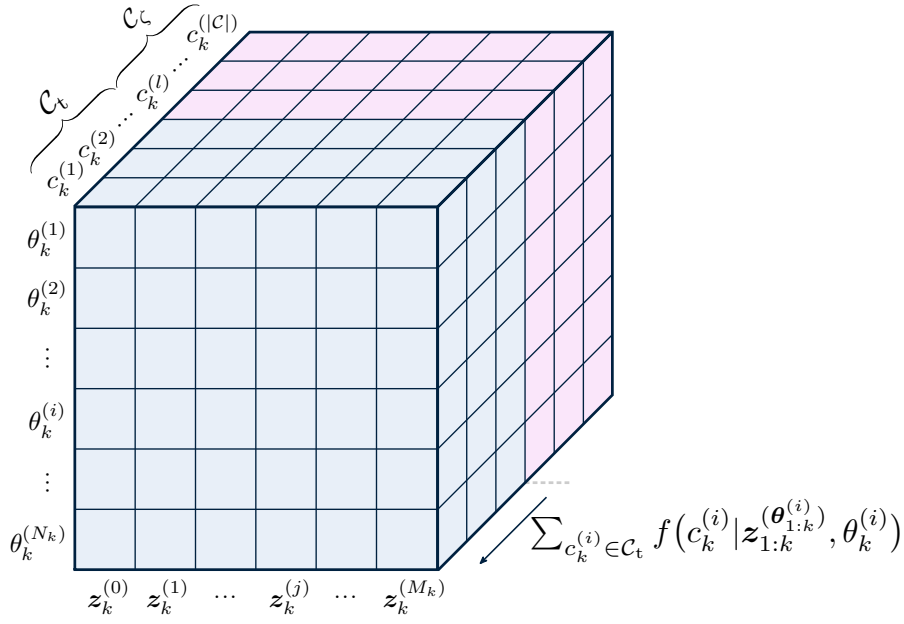


Figure 4.2: Graphical representation of equation (4.33). The probability of data association between the i -th target and the j -th measurement (i. e., $\theta_k^{(i)} = j$) is given by the summation over the classes in which a tracked target can be classified.

association problem as

$$l^{(i,j)} = -\log\left(w\left(z_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right)\right) \quad (4.29a)$$

$$= -\log\left(\sum_{c_k^{(i)} \in C_t} f\left(c_k^{(i)} | z_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right)\right) \\ - \log\left(\int f\left(\mathbf{x}_k^{(i)}, z_k^{(\theta_k^{(i)})}, \theta_k^{(i)} | z_{1:k-1}^{(\theta_k^{(i)})}\right) d\mathbf{x}_k^{(i)}\right) \quad (4.29b)$$

$$= l_C^{(i,j)} + l_T^{(i,j)} \quad (4.29c)$$

where

$$l_T^{(i,j)} = -\log\left(\int f\left(\mathbf{x}_k^{(i)}, z_k^{(\theta_k^{(i)})}, \theta_k^{(i)} | z_{1:k-1}^{(\theta_k^{(i)})}\right) d\mathbf{x}_k^{(i)}\right) \quad (4.30)$$

and

$$l_C^{(i,j)} = -\log\left(\sum_{c_k^{(i)} \in C_t} f\left(c_k^{(i)} | z_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)}\right)\right). \quad (4.31)$$

Specifically, $l_T^{(i,j)}$ is the data association cost conventionally employed in MTT, which can be evaluated using the closed-form solution provided in (2.30) and (2.31). On the other hand, $l_C^{(i,j)}$ is an additional data association cost, which is provided by

the classification scores. In particular, let define

$$f\left(c_k^{(i)}, \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}\right) \quad (4.32)$$

where $c_k^{(i)} \in \mathcal{C}_t$, as a probability mass function (PMF). Therefore,

$$\sum_{c_k^{(i)} \in \mathcal{C}_t} f\left(c_k^{(i)} | \mathbf{z}_{1:k}^{(\theta_{1:k}^{(i)})}, \theta_k^{(i)}\right) \quad (4.33)$$

is the marginalization, given a data association hypothesis, of the (4.32) with respect to the classification variables $c_k^{(i)} \in \mathcal{C}_t$, i.e., the classes which can be assigned to a tracked target. Fig. 4.2 gives a graphical representation of the (4.33) considering data association hypotheses and classification scores.

In equation (4.29b), it has been defined an enhanced data association cost, referred to as JTC data association cost, which considers both the statistical distance between two positional states and a classification score based on the performed classification. In particular, the classification score $l_C^{(i,j)}$ defines how likely a data association hypothesis is by considering only the performed classification. Then, $l_C^{(i,j)}$ can either confirm or change the result obtained employing the conventional measurement-to-target association approaches benefiting both soft and hard-decision based data association.

The algorithms developed for conventional MTT can be completely integrated in the JTC approach as positional state estimators. In JTC filtering, classification allows (i) the choice of a specific motion model for each target to evaluate the prediction density, (ii) weights the JTC posterior density, and (iii) contributes in data association. In particular, the JTC data association guarantees information superiority to associate measurements to targets enabling association costs tuned by the classification scores.

4.2 Joint tracking and classification via single MIMO radar at mmWaves

This section aims to propose an entire framework to perform JTC in complex wireless environments employing a single FMCW MIMO radar operating at mmWaves. In particular, FMCW radars allow to infer range and velocity of multiple targets, relying on a sequence of reflected chirps, i.e., signals with a varying frequency over time. In addition, MIMO technologies guarantee the availability of multiple antennas, which enable to estimate the AOA of a reflected waveform and, then, to perform localization using a single device.

Fig. 4.3 outlines the proposed framework to perform JTC via a single FMCW

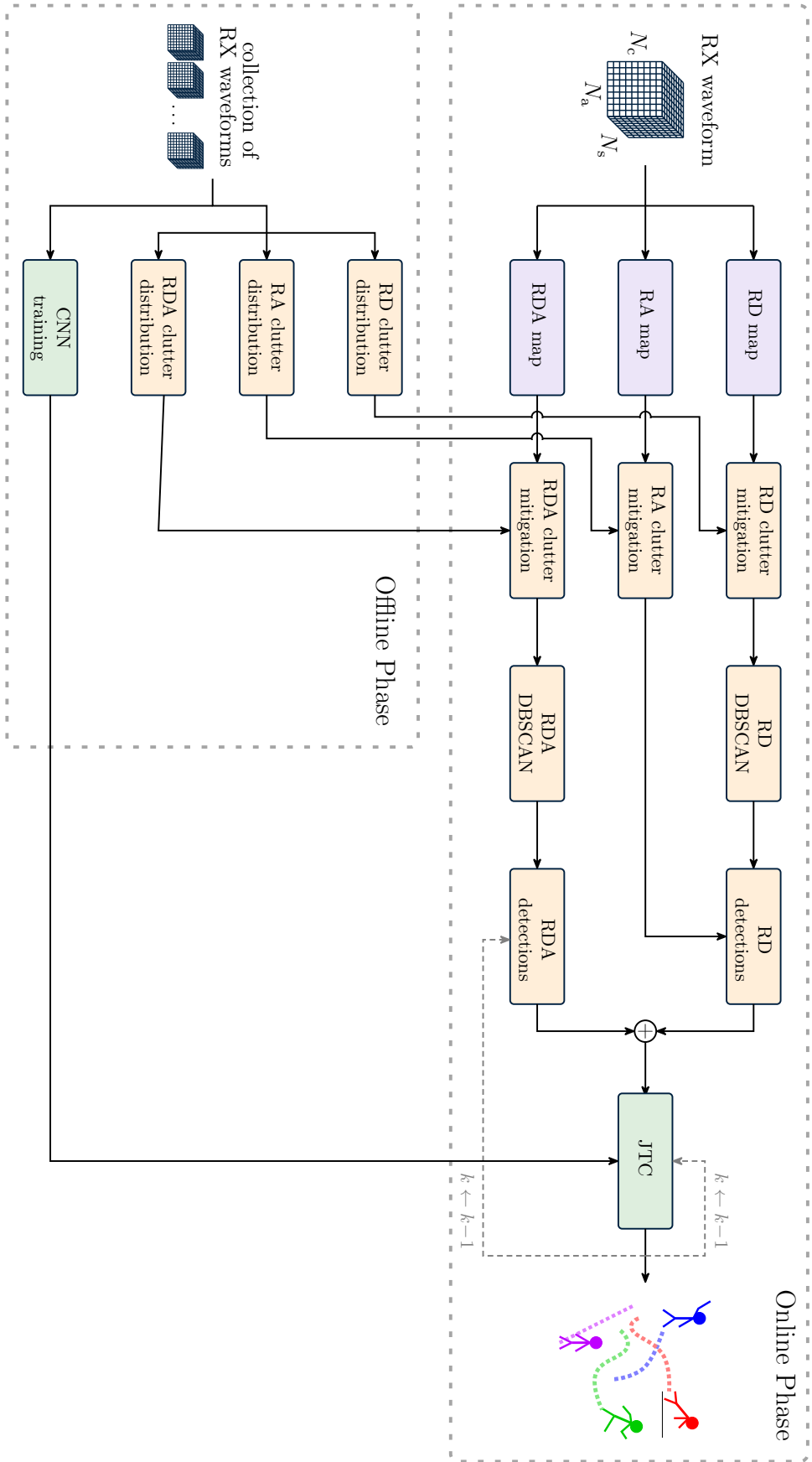


Figure 4.3: Block diagram of the proposed JTC framework to perform tracking and classification via single FMCW MIMO radar operating at mmWaves. In particular, lilac blocks represent the signal processing phase, orange blocks represent the clutter mitigation and detection phase, and green blocks the tracking and classification phase.

MIMO radar. First, signal processing and clutter mitigation techniques need to be applied to the raw samples of reflected waveforms. Then, detection is required to reveal the presence of targets from the processed samples. Finally, for each detected target a positional measurement and a classification metric can be obtained. The JTC filtering processes the collected measurements to jointly infer the multitarget state $\mathbf{x}_k$ and the multitarget class $\mathbf{c}_k$, at any time instant k .

4.2.1 Detection technologies

In complex wireless environment the employment of clutter mitigation techniques is needed to achieve high accuracy localization and classification. In particular, an effective approach to remove static clutter originated from static scatters is to remove the background noise from the reflected waveform. It requires an offline phase in which the static clutter distribution can be inferred collecting a large amount of measurements without any target in the environment.

Consider the range-angle (RA) map which discretizes the monitored space in a finite set of points. By collecting a large amount of measurements, it can be inferred the clutter distribution for each considered spatial location. The same approach can be employed for estimating the noise distribution of the elements in the RD and RDA maps.

For each point with coordinate $\mathbf{p}$ in the RD, or RDA, map a specific threshold can be defined based on its own estimated noise power distribution.³ In particular, the noise power $\mathbf{n}_{\mathbf{p}}$, of a point at the coordinate $\mathbf{p}$, can be approximated as Gaussian distributed and given by

$$\mathbf{n}_{\mathbf{p}} \sim \varphi\left(\mathbf{n}_{\mathbf{p}}; \bar{\mathbf{n}}_{\mathbf{p}}, \sigma_{\mathbf{p}}^2\right). \quad (4.34)$$

Each point in RD and RDA map is considered as originated from either a target or clutter, if its value is over the estimated threshold or not, respectively. The threshold $\xi_{\mathbf{p}}$ for a point at the coordinate $\mathbf{p}$ can be defined based on the learning (i.e., clutter distribution inference), performed during the offline phase, as follows

$$\xi_{\mathbf{p}} = \bar{\mathbf{n}}_{\mathbf{p}} + \alpha \sigma_{\mathbf{p}}^2 \quad (4.35)$$

where $\alpha \in \mathbb{R}^+$. The parameter α tunes the false alarm and miss-detection probability. In fact, an higher α increases the number of miss-detection, but reduces the false alarm probability and vice-versa.

The target detection consist in processing the RD map to reveal moving elements

³For notational convenience, we denote $\mathbf{p} \in \mathbb{R}^2$ a RD map coordinate and $\mathbf{p} \in \mathbb{R}^3$ a RDA map coordinate.

in the environment. Then, given the range-velocity coordinate of a detection, the corresponding AOA is obtained leveraging the RDA map. In particular, from the RDA map, given a range-velocity coordinate, it can be obtained a list of angles, each characterized by a frequency shift. The estimated AOA is the one providing the largest frequency deviation.

The target presence produces a set of over-threshold points in the RD map due to the high radar resolution and the multipath propagation. Multiple targets generate multiple clouds of range-velocity coordinates. Clustering techniques allow to group the over-threshold range-velocity coordinates in clusters. Each cluster can be reduced to a SVE evaluating its own centroid. The DBSCAN algorithm allows to group detections depending on their density in the RD map. It expands a cluster until new elements within a certain radius are found. By denoting as Ω_j the j -th detected cluster, with $j \in \{1, 2, \dots, M\}$ where M is the number of clusters, the corresponding estimated centroid $\boldsymbol{\mu}_j$ is given by

$$\boldsymbol{\mu}_i = \frac{\sum_{\mathbf{p} \in \Omega_j} \mathbf{p} P_{\mathbf{p}}}{\sum_{\mathbf{p} \in \Omega_j} P_{\mathbf{p}}} \quad (4.36)$$

where $\mathbf{p}$ is the range-velocity coordinate and $P_{\mathbf{p}}$ is the associated received power, i.e., the frequency shift in the RD map at $\mathbf{p}$.

A further improvement in the detection consists in reducing the miss-detection probability exploiting the knowledge of the positional state of each target at the previous time instant. Specifically, around each positional prediction, a gating region is defined for extracting a sub-tensor from the RDA map to process using DBSCAN. By the union of centroids coming from both RD map and each RDA sub-tensor, a set of M_k detections is obtained. For each detection, range, velocity, and AOA define a positional measurement. Given a range and an angle, the Doppler shifts generated by the target movements are extracted from the SS dimension of the RDA map and exploited as a classification metric. Therefore, M_k positional measurements and classification analytics are obtained and compose the measurement vector $\mathbf{z}_k$, which is used as input to the JTC filtering.

4.2.2 Tracking and classification technologies

The closed-form solution proposed in Section 4.1.1 allows to approximate the JTC filtering and develop computational efficient algorithms to track and classify multiple targets, simultaneously. The JTC implementation, first, requires to find the most likely data association hypotheses considering both positional and class information. Then, independent estimators perform tracking and classification of the tracked targets.

At the current discrete time instant k , consider M_k measurements and N_k targets.

On one hand, each measurement is characterized by positional information and a classification features. On the other hand, tracked targets are described by the estimated positional state and the history of its associated classification features until the previous time instant $k - 1$.

The equation (4.29c) gives the JTC cost of associating a measurement j to a target i , with $j \in \{0, 1, \dots, M_k\}$ and $i \in \{1, 2, \dots, N_k\}$, and is expressed as the sum of two contributions, i.e., $l_T^{(i,j)}$ and $l_C^{(i,j)}$. The term $l_T^{(i,j)}$ is the statistical distance between a positional state and a positional measurement. It is the conventional metric used for data association in MTT and, under certain assumptions, allows a closed-form solution as expressed in (2.30) and (2.31). The term $l_C^{(i,j)}$ is the classification score and is computed as described in the following.

The collected classification features $\mathbf{z}_k^{(j)}$, with $j \in \{0, 1, \dots, M_k\}$ can be associated to each target to generate $(M_k + 1) \times N_k$ association hypotheses. In particular, each tracked target is described by its positional state and its sequence of classification features $\mathbf{z}_{1:k-1}^{(\theta_k^{(i)})}$. An association hypothesis $\mathbf{z}_{1:k}^{(\theta_k^{(i)})}$, with $i \in \{1, \dots, N_k\}$ and $\theta_k^{(i)} \in \{0, 1, 2, \dots, M_k\}$, is a sequence of positional states and classification features. To evaluate an association hypothesis $\mathbf{z}_{1:k}^{(\theta_k^{(i)})}$ consists in finding the probability that $\mathbf{z}_k^{(j)}$ is generated by i -th target.

By employing a classifier, the probability to belong to each class $c \in \mathcal{C}$ can be defined for an association hypothesis $\mathbf{z}_{1:k}^{(\theta_k^{(i)})}$. It generates $|\mathcal{C}|$ classification scores, which sum to 1. The classification of all the association hypotheses generates the probabilistic cube depicted in fig. 4.2. The normalized version of the probabilistic cube is the PMF of associating a measurement $\theta_k^{(i)} = j$ to the target i and that $c_k^{(l)}$, with $l \in \{1, 2, \dots, |\mathcal{C}|\}$, is the class of the classification metric sequence $\mathbf{z}_{1:k}^{(\theta_k^{(i)})}$, i.e., the probabilistic cube is the graphical representation of $f(c_k^{(l)}, \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)})$. Therefore, $\sum_{c_k^{(l)} \in \mathcal{C}_t} f(c_k^{(l)}, \mathbf{z}_{1:k}^{(\theta_k^{(i)})}, \theta_k^{(i)})$ provides the probabilistic score of associating the measurement $\theta_k^{(i)} = j$ to the i -th target by marginalizing with respect to the classes that can be assigned to a target.

To complete the data association cost matrix $\mathbf{L}$, both $l_C^{(i,j)}$ and $l_T^{(i,j)}$ need to be computed for all the association hypotheses. The processing of $\mathbf{L}$ through an M -best assignment algorithm (e.g., the Murty's algorithm) provides the M -best data association hypotheses [137].

The JTC data association can be fully employed in conventional MTT tracking algorithms, such as GNN and JPDA, providing information superiority guaranteed by the classification. The mathematical frameworks developed to complete the MTT filtering can be used in JTC filtering as positional estimator.

Chapter remark

Joint tracking and classification extends Bayesian filtering by simultaneously estimating position and class of multiple device-free targets. Introducing classification variables into the recursive model benefits both kinematic prediction thanks to the employment personalized model for each target and position inference thanks to information superiority for data association. The closed-form derivation of prediction, posterior, and data association enables real-time computation. Joint inference improves reliability and accuracy for achieving situation awareness even in complex wireless environments.

Chapter 5

Microservice-based architecture for multimodal sensing

Multimodal wireless sensing of non-collaborative targets represents a key enabler for achieving situation awareness in next-generation (xG) wireless networks. By leveraging data diversity, heterogeneous sensing technologies, such as automotive radars and camera systems, provide information superiority and enhance environmental perception. However, the integration of sensors characterized by different resolutions, sampling rates, and data representations requires the development of flexible and scalable frameworks for information fusion and analytic extraction.

This chapter introduces a microservice-based architecture that enables distributed and scalable processing of multimodal sensing data. Specifically, this chapter presents (i) a microservice-based design that integrates sensing of connected and unconnected targets within an asynchronous message-oriented framework, ensuring modularity, scalability, and interoperability across multiple services; and (ii) a methodology for multimodal detection, localization, and classification of multiple unconnected targets by exploiting radar and camera data through distributed inference.

5.1 Microservice-based architecture

Consider an xG network composed of N_b base stations (BSs) and N_u user equipments (UEs). The BSs support the deployment of multiple transmission-reception points (TRPs) to improve reliability and coverage. To enhance localization capabilities, we consider that each UE is equipped with an UWB radio and that N_a UWB anchors with known positions are deployed in the environment. Furthermore, we consider that the operation environment is monitored via wireless sensing. More specifically, we consider an automotive radar and a camera-based system to detect, track, and identify unconnected targets. The goal is to develop a descriptive network digital twin

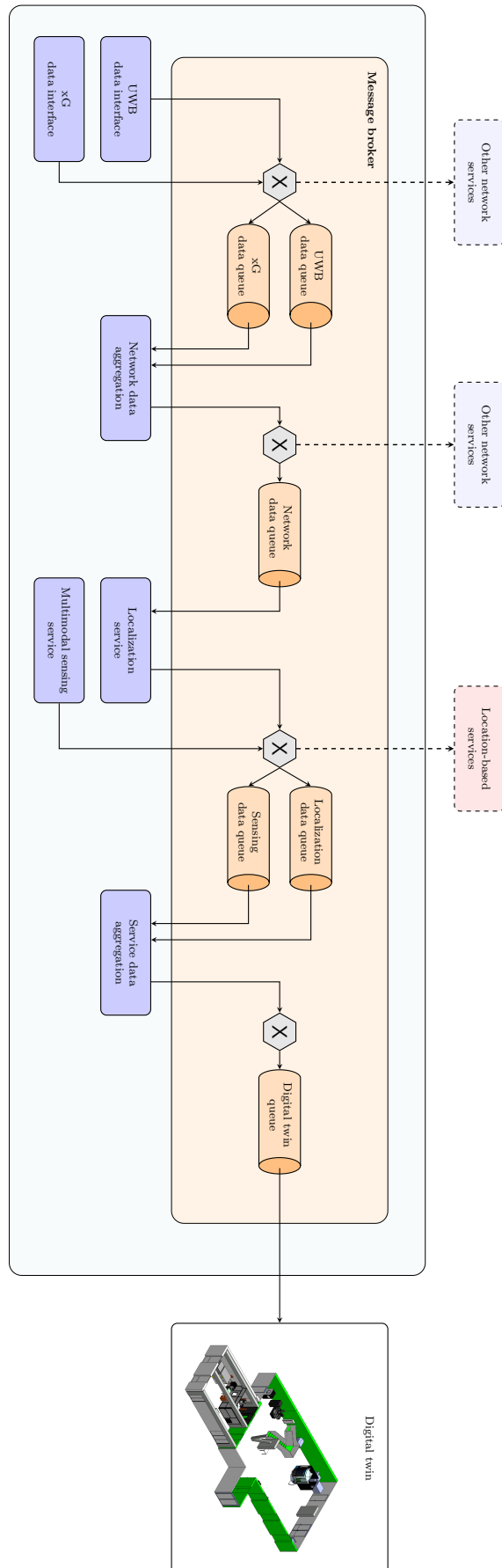


Figure 5.1: Block diagram of the proposed microservice-based architecture.

(NDT) allowing to monitor the network and its operation environment by harnessing situational information of both connected devices and unconnected targets.

Fig. 5.1 shows the proposed microservice-based architecture for exposing situation awareness in NDT. In this architecture, network services are seamlessly connected through a message broker. Each network service has a specific functionality, *publishing* and/or *consuming* data at different stages of the processing pipeline. The role of the message broker is to route data from publishers to consumers. Since the message broker is based on a network protocol, it allows services to be deployed on different servers in a distributed fashion, e.g., by considering edge cloud components. Specifically, we consider advanced message queuing (AMQ) protocol for the messaging middleware (i.e., RabbitMQ) and data parsed in JavaScript object notation (JSON) format.

The proposed architecture spans a processing pipeline to provide localization and sensing capabilities. Such a pipeline includes: (i) data interfaces to collect measurements; (ii) data aggregation modules to parse information from multiple sources; and (iii) localization and sensing modules implementing inference algorithms. The data is routed through exchanges and queues that allow published data to be consumed and used at the corresponding stage. Note that the data can also be consumed by external network services connected to the exchange points at the different stages.

5.1.1 Architecture components

The components of the microservice-based architecture are described in the following.

xG data interface: This microservice collects and publishes data of the xG network. Specifically, we consider metrics from the xG radio access network (RAN), including reference signal received power (RSRP), block error rate (BLER), and modulation and coding scheme (MCS) for each connected UE. Furthermore, we also consider that this data interface provides time-of-arrival (TOA) measurements obtained from uplink sounding reference signal (SRS) for localization.

UWB data interface: This microservice collects and publishes data of the UWB network. We consider both time-based and angle-based measurements obtained from each UWB anchor for each UE.

Network data aggregation: This microservice consumes the data from the xG and UWB data interfaces, and aggregates it to publish a single packet containing all the measurements.

Localization service: This microservice performs inference for xG localization. We consider data fusion with UWB for enhanced localization accuracy. This microservice consumes the aggregated network data to obtain the xG and UWB measurements. For the inference, we consider soft information (SI)-based localization [17], a probabilistic approach that enables seamless fusion of heterogeneous measurements and whose

```

1  {
2    "ID_0": {
3      "type": "",
4      "source": "5G+UWB",
5      "timestamp": "09:00:00",
6      "position": {
7        "x": 0,
8        "y": 0,
9        "z": 0,
10       "prj": "RHZU",
11       "timestamp": "09:00:00",
12       "origin": {
13         "x": 0,
14         "y": 0,
15         "z": 0,
16         "prj": "RHZU"
17       }
18     },
19     "measures": [
20       {
21         "metric": "",
22         "unit": "",
23         "value": 0,
24         "timestamp": ""
25       },
26     ],
27     "additional_data": {}
28   },
29   "ID_1": {}
30 }

```

Figure 5.2: Data model template in JSON format to share data across the developed microservice-based architecture.

benefits have been demonstrated for 5G and beyond networks [139–143]. This service publishes a packet aggregating the location information, i.e., position estimates $\hat{\mathbf{p}}_i$ for $i \in \{1, 2, \dots, N_u\}$, with the network measurements.

Sensing service: This microservice performs wireless sensing to obtain information of unconnected targets present in the environment. We consider detection, tracking, and identification of multiple targets employing an automotive radar and a camera-based system. In this service, the radar and camera measurements are directly employed without publishing them to the message broker. This service publishes the sensing data for N_d detected targets at a given time instant. In particular, the data contains the estimated positional information $\hat{\mathbf{x}}_i$ and class $\hat{c}_i \in \mathcal{C}$ for $i \in \{1, 2, \dots, N_d\}$. The positional information $\hat{\mathbf{x}}_i$ contains the coordinates and velocities of the target i . The class of the target identifies its type (e.g., human or robot) from a predefined set $\mathcal{C}$ containing all the possible target classes.

Service data aggregation: This microservice consumes the data from the localization and sensing services, and aggregates them to publish a single packet containing all the measurements and situational information of N_u UEs and N_d unconnected targets.

Digital twin service: This microservice consumes the aggregated service data for building the NDT. We consider a digital twin of the environment to visualize the

network behavior as a descriptive NDT. This service can be extended to include further NDT functionalities such as simulation.

The proposed architecture is flexible and scalable to expand its functionalities, including multimodal sensing and other NDT services. The design based on a message broker allows to implement a distributed architecture harnessing edge cloud servers to allocate compute-intensive tasks such as inference algorithms. The proposed architecture also allows to disaggregate hardware components that provide measurements from the software components that process the measured data.

5.1.2 Architecture data model

This section introduces a data model to provide situational information of generic assets. The proposed microservice-based architecture requires a data model for describing generic assets and to exchange their information among the different network services. We focus on establishing a data model that allows to deliver and seamlessly parse data for situation awareness across the processing pipeline of the microservice-based architecture.

Fig. 5.2 depicts the data model template in JSON format. We consider that assets are represented by objects listed as `IDi` where i denotes the index. Each object represents a generic asset, either a connected device or unconnected target, and includes situational information and generic measurements. The types of values for the object keys in the data model are:

- **type**: string describing the type of asset.
- **source**: string providing the data source.
- **timestamp**: string providing the timestamp of the data in format “HH:MM:SS”.
- **position**: object describing the position information. Such an object includes the coordinates of the asset position, the projection convention [e.g., right-handed Z up (RHZU)], the timestamp, and the local map origin.
- **measures**: array of objects describing generic measurements related to the asset.
- **additional_data**: object with arbitrary format for specifying additional data related to the asset.

The object representing the position information allows to specify the projection convention and the origin of the local coordinate system. The types of values for the keys in this object are:

- **x**, **y**, and **z**: float numbers representing the coordinates.

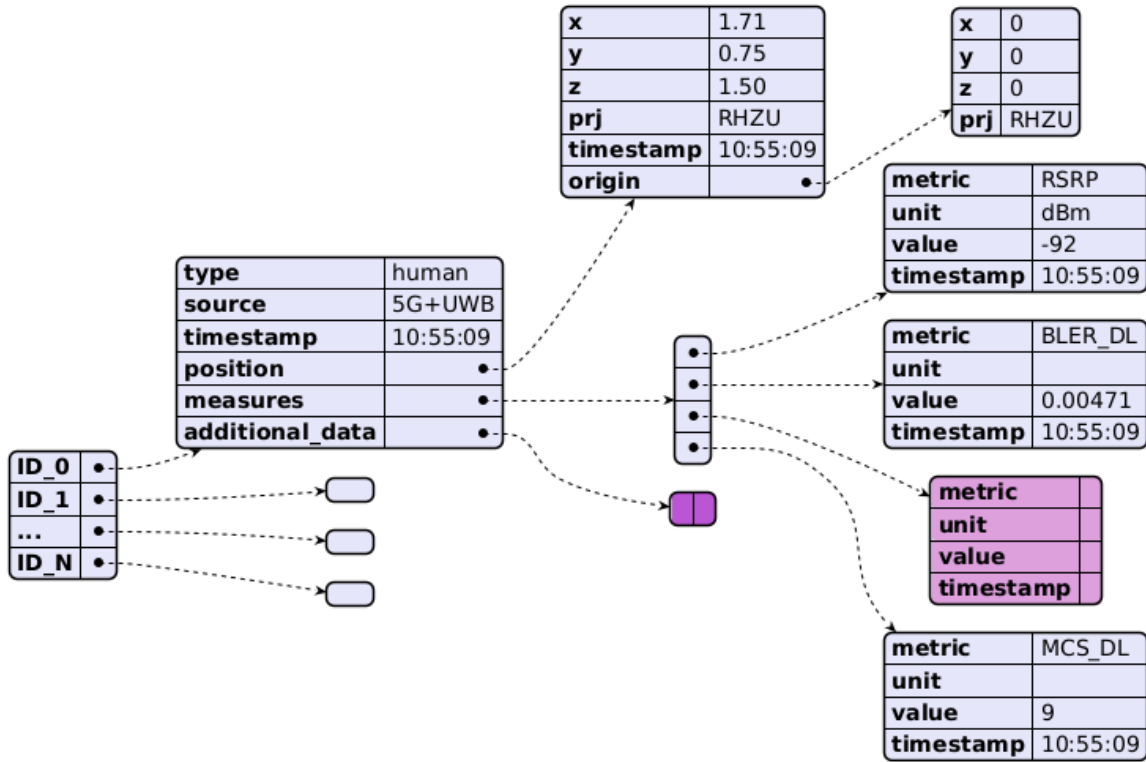


Figure 5.3: Data model template in JSON format to share data across the developed microservice-based architecture.

- **prj**: string with the projection convention, e.g., RHZU.
- **timestamp**: string providing the timestamp of the position information.
- **origin**: object with keys **x**, **y**, **z**, and **prj** formatted as those in the main body providing the coordinates of the local origin.

This data model allows to aggregate arbitrary measurements related to the asset. We define the following keys and values for the measurements:

- **metric**: string with the name of the measurement.
- **unit**: string providing the unit of measure, e.g., m.
- **value**: float number for the measurement value.
- **timestamp**: string providing the timestamp.

Fig. 5.3 depicts an example of the data model in unified modeling language (UML) format. Note that the data model allows to provide different timestamps in its components to support asynchronous data. This graphical representation highlights the flexibility of the proposed model to support reporting generic measurements and additional data in arbitrary objects.

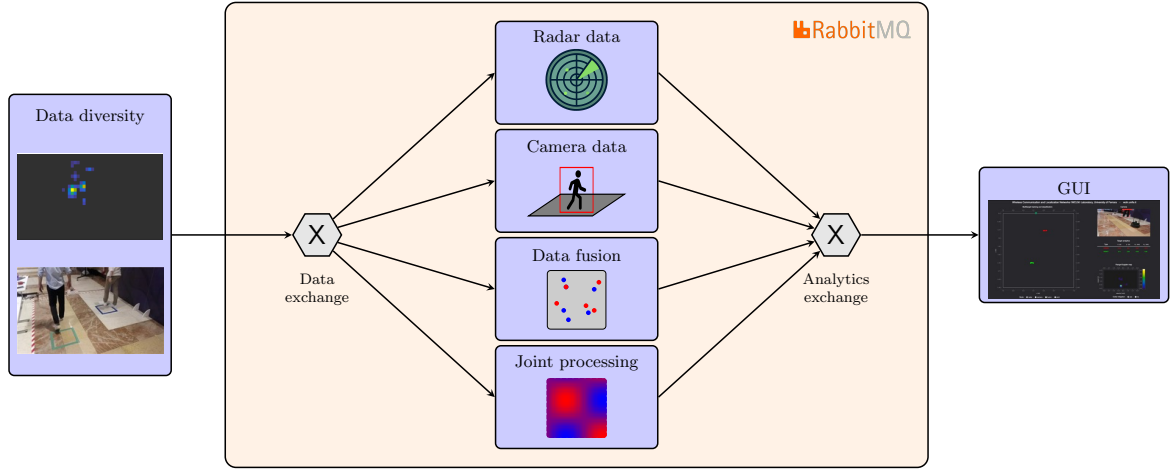


Figure 5.4: Block diagram of the proposed microservice-based architecture for multimodal sensing.

5.2 Multimodal wireless sensing

This section presents the implementation of the multimodal sensing block illustrated in Fig. 5.1. The developed microservice-based architecture for multimodal wireless sensing enables distributed and scalable processing of heterogeneous data collected by the automotive radar and camera devices. Fig. 5.4 shows the block diagram of the proposed architecture, where radars and cameras act as data sources, while multiple microservices are responsible for extracting, processing, and publishing sensing information. Each sensor operates independently and provides its measurements to the corresponding service that perform detection, tracking, and classification.

In the proposed architecture, the data sources (i.e., radar and camera) publish collected data to a RabbitMQ exchange, referred to as the *Data Exchange*, using dedicated routing keys. Each service then subscribes to the Data Exchange and consumes only the data of interest by specifying the corresponding routing key. RabbitMQ allows multiple services to simultaneously consume the same published data, thus supporting parallel processing and modular scalability. Processed sensing analytics are then published on the *Analytics Exchange*, which populates the *Sensing Data Queue* shown in Fig. 5.1, while also providing real-time data for the developed graphical user interface (GUI).

The multimodal sensing services proposed in the developed architecture are *radar data*, *camera data*, *data fusion*, and *joint processing* and their implementation is described in the following.

- **Radar data service** implements the integrated tracking and classification of multiple targets using only radar data, as detailed in Section 3.
- **Camera data service** implements tracking and classification of multiple targets

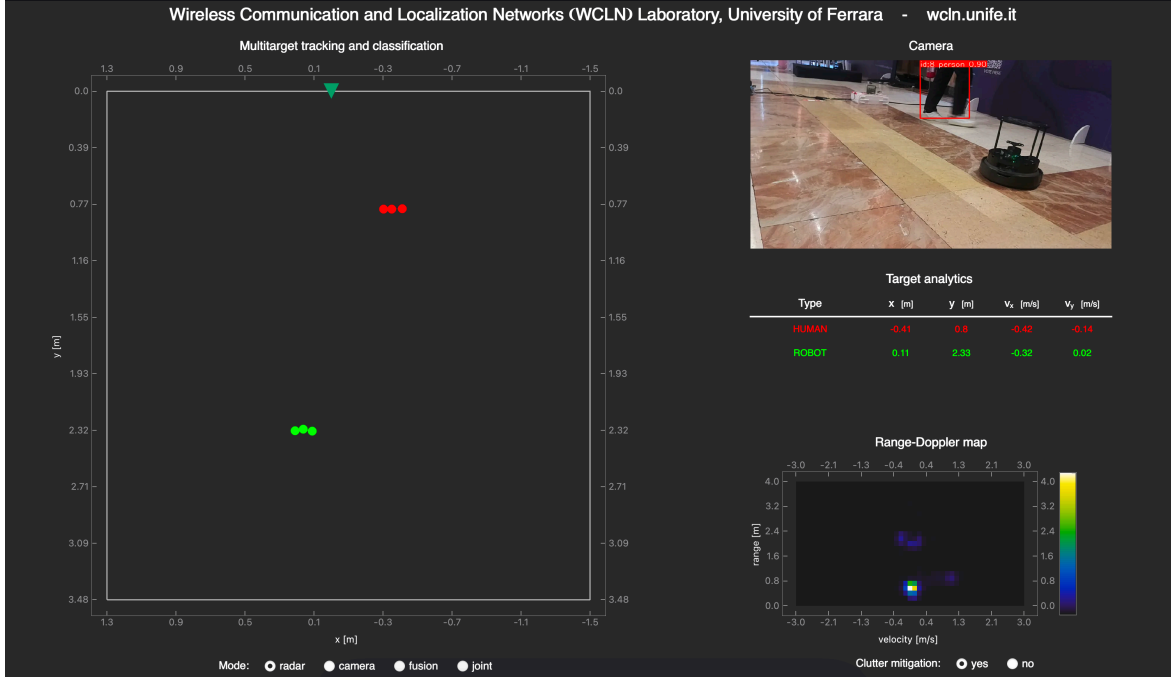


Figure 5.5: Multimodal sensing GUI in the proposed microservice-based architecture.

based on camera data, following the framework described in Section 2.3.

- **Data fusion service** implements integrated tracking and classification of multiple targets by combining measurements from both radar and camera. Therefore, $\mathbf{z}_k = \mathbf{z}_{r,k} \cup \mathbf{z}_{c,k}$, where $\mathbf{z}_{r,k}$ and $\mathbf{z}_{c,k}$ are measurements obtained by radar and camera, respectively
- **Joint processing service** implements the JTC approach to perform tracking and classification of multiple targets jointly relying on both radar and camera measurements, as described in Section 4.

A dedicated GUI has been developed to visualize the sensing information generated by the different services. Specifically, the GUI includes: (i) a 2D map of the monitored environment displaying target trajectories in real time; (ii) a dynamic table summarizing the extracted sensing analytics (e.g., target class and position); and (iii) a synchronized visualization of RD maps either before or after clutter mitigation and camera frames after YOLO processing. The user can select which service to visualize through the interface controls and can switch between displaying the RD map either before or after clutter mitigation. Fig. 5.5 shows the developed GUI for multimodal wireless sensing operating in real-time. Target trajectories and related analytics are displayed on the 2D environmental map and the corresponding data table, while the RD map after clutter mitigation and the camera frames after YOLO detection are shown simultaneously to provide complete situational awareness.

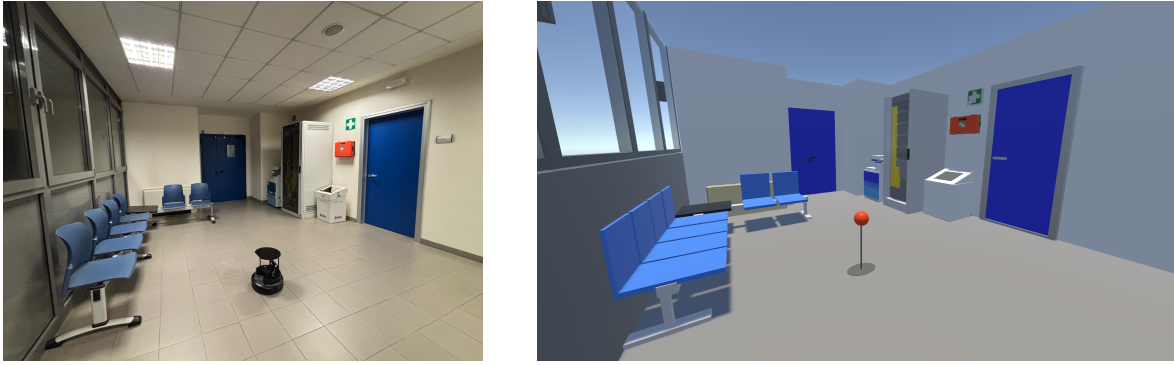


Figure 5.6: Digital twin interface consuming sensing analytics through the microservice-based architecture in an indoor office environment.

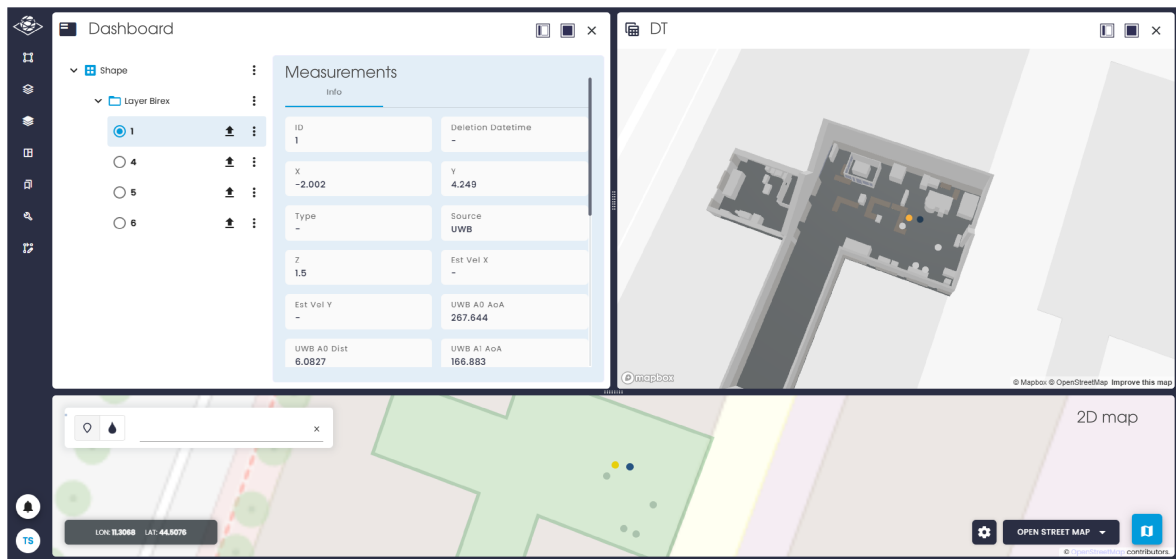


Figure 5.7: Digital twin dashboard consuming sensing analytics through the microservice-based architecture in an industrial environment.

The microservice-based architecture has been deployed considering different indoor environments, including office and industrial areas, as well as different visualization tools. Fig. 5.6 shows a digital twin interface that consumes sensing analytics through the microservice-based architecture in an indoor office environment. Fig. 5.7 shows a digital twin dashboard that consumes sensing analytics through the microservice-based architecture in an industrial environment. The developed microservice-based architecture allows the deployment of different visualization services exploiting only the sensing analytics of interest. Note that visualization and processing tools can operate simultaneously and distributedly providing efficient resource utilization and processing scalability.

Chapter remark

A microservice-based architecture supports scalable and distributed multimodal sensing. Integrating radar and camera data within containerized services allows modularity, interoperability, and efficient computation. The multimodal approach improves accuracy and resilience to environmental variations, while the microservice paradigm ensures flexibility for deployment across heterogeneous hardware and network infrastructures.

Chapter 6

Field-trials and experimentation

This chapter reports the experimental validation and field trials conducted to assess the performance of the proposed tracking and classification frameworks. Experimentations are carried out using off-the-shelf FMCW radar and camera devices in both indoor and outdoor environments, considering different targets such as humans and mobile robots. The chapter details the system setup, experimental methodology, and performance evaluation in terms of tracking and classification accuracy. The obtained results confirm the effectiveness, scalability, and robustness of the proposed approaches under realistic operating conditions, demonstrating their suitability for xG wireless networks.

6.1 Demonstration of integrated tracking and classification

The proposed framework to perform tracking and identification of multiple targets is validated via experimentation using a mmWave MIMO radar in the pilot line of the BI-REX, an Italian competence center for Industry 4.0.

6.1.1 System setting

The considered experimentation area is $8.3 \times 6.5 \text{ m}^2$ and presents many scatterers, such as industrial machines, tables, and other assets. Fig. 6.1 shows the scenario during the experimentation. The sensor employed to perform both tracking and identification is an off-the-shelf FMCW MIMO radar with 2 transmitting and 16 receiving antennas forming a linear array. The transmitted signals are linear chirps operating in the frequency range $76 - 77 \text{ GHz}$ with a duration $T_c = 120 \mu\text{s}$. Each radar measurement consists in transmitting 128 chirp signals and collecting echoes generated by reflections at all the receiving antennas. Each reflected signal is composed by 512 samples. The

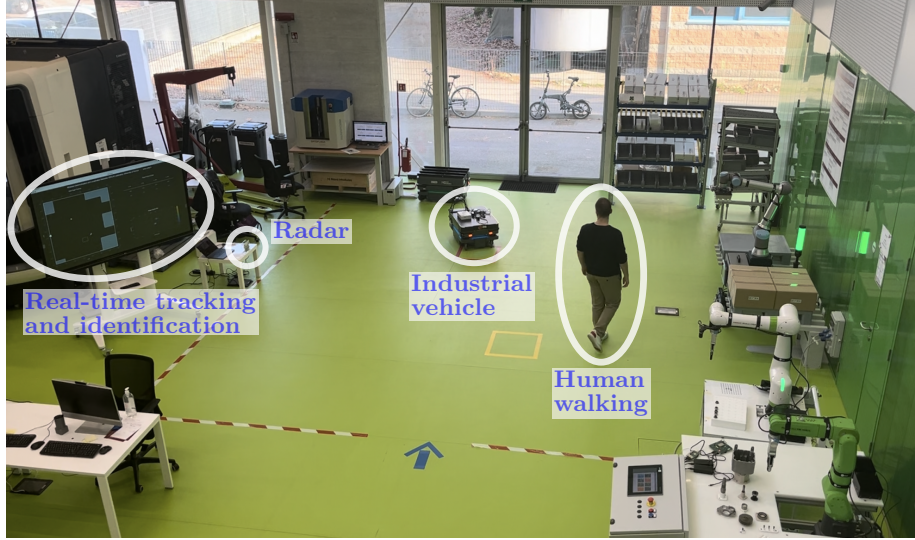


Figure 6.1: Experimentation area in the pilot line of the BI-REX with a human walking and an industrial vehicle moving in the monitored area.

entire chirp sequence is collected every 0.1 s by multiple antennas and compose the radar data cube.

The collected radar data cubes are processed via DFT to obtain RD and RDA maps. All the elements in the RD and RDA maps are compared with a specific threshold to detect the target presence. The thresholds are defined during the offline phase, in which 1000 measurements are collected with no target in the environment. The processing of such measurements produces RD and RDA maps, that are exploited to estimate the Gaussian distribution parameters that best approximate the noise distribution at each range-Doppler and range-Doppler-angle coordinate. Then, for a point at the coordinate $\mathbf{p}$ in the RD and RDA maps, the threshold $\xi_{\mathbf{p}}$ is designed according to equation (4.35) by using the estimated mean $\bar{n}_{\mathbf{p}}$ and standard deviation $\sigma_{\mathbf{p}}$, together with a scaling parameter $\alpha_{\mathbf{p}} = 10$. The over threshold elements of the RD map are clustered by employing the DBSCAN algorithm, with a neighborhood searching radius $\epsilon = 0.05$ and considering clusters with a minimum of 2 elements. Centroids are obtained by equation (3.2) and compose the set of positional measurements.

6.1.2 Tracking algorithm

Target positional states and positional measurements are 4-dimensional vectors (i.e., $n_x = m_z = 4$), each one characterized by a 2D position and a 2D velocity. Track initialization requires that a track hypothesis achieves a length of $K = 3$ measurements, while both track hypotheses and confirmed tracks are deleted for values of the corresponding covariance matrix trace above $\xi_d = 2.5$. Data association is performed considering a gating region $\xi_g = 0.5$ m around the predicted positional state of each target. In particular, the target positional state prediction is performed considering the

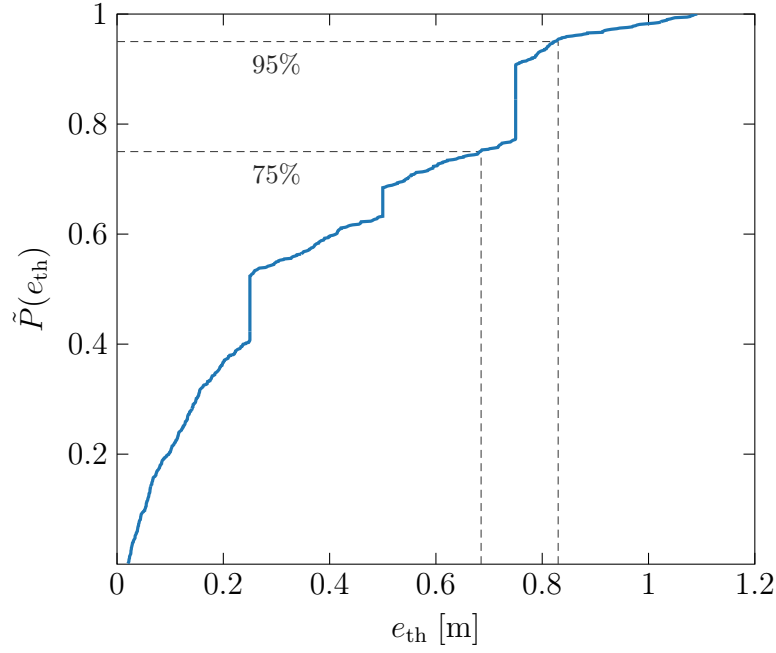


Figure 6.2: ECDF of the tracking error evaluated using the GOSPA metric.

constant velocity model, which assumes no velocity variations in the interval between two measurements. The linear measurement model employed is the identity matrix, i.e., $\mathbf{H}_k = \mathbf{I}_4$.

The tracking accuracy is quantified using the generalized optimal subpattern assignment (GOSPA) metric of order $p = 1$, cut off $c = \xi_g$, and normalization term $\alpha = 2$ [144]. In particular, localization error is evaluated computing the Euclidean distance and each misdetection introduces a penalty of 0.25 m in the GOSPA metric. When a misdetection occurs the range-angle coordinate employed for collecting classification analytics is given by the predicted positional state. The target position is estimated every 0.1 s, while the target class every 2 s aiming to determine if the target is a human or an industrial vehicle.

Fig 6.2 shows the empirical cumulative distribution function (ECDF) of the tracking error based on the GOSPA metric, which accounts both localization accuracy and misdetections. The proposed framework has been validated monitoring an area with moving humans and industrial vehicles by collecting 1200 measurements to cover a time frame of 120 s. The 99% of the measurements considers the presence of at least of 1 target, while the 85% of exactly 3 targets. At the 95th percentile, the presented approach for detection and tracking provides a localization error of 0.83 m, while, at the 75th percentile, the error is reduced to 0.68 m. The average number of false-alarms per radar measurement is 1.23. In particular, for 98% of the measurements at most 3 false-alarms are collected.

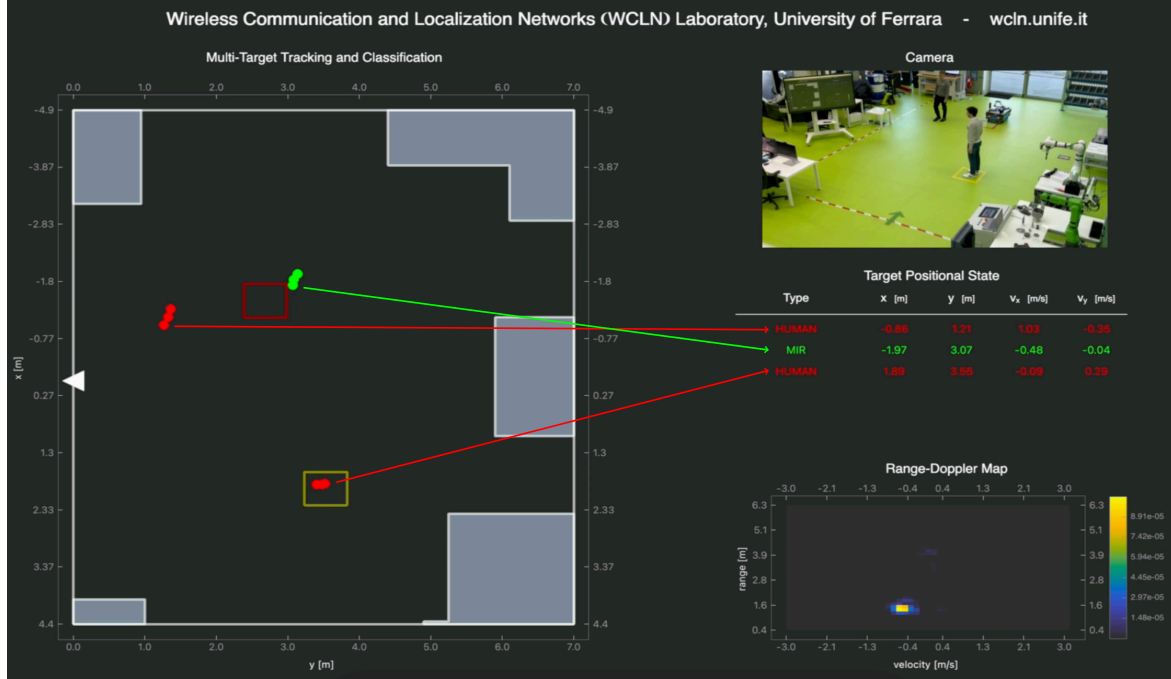


Figure 6.3: Representation of real-time multitarget classification in the developed GUI for integrated tracking and classification.

6.1.3 Identification algorithm

Classification is performed employing a CNN, which inputs $W = 20$ sequences of $N_c = 128$ samples collected relying on the estimated target positions and organized as a linear array for preserving the time correlations. Fig. 6.3 shows real-time multitarget classification in the developed GUI for integrated tracking and classification. The CNN architecture employed for processing the 1D time series consist of 6 layers: (i) 1D convolutional input layer with 128 filters and kernel size of 2; (ii) max pooling layer with pool size of 2; (iii) 1D convolutional layer with 32 filters and kernel size of 2; (iv) max pooling layer with pool size of 2; (v) a fully connected layer with 16 units; and (vi) the output layer. The activation function in the CNN inner layers is the rectified linear unit function, while the output layer employs the sigmoid function. Training is performed employing the categorical cross-entropy as empirical loss function and the Adam optimizer with a learning rate of 10^{-3} .

The training set consists of 1800 time series collected in the same experimentation environment, but with a different deployment of the industrial machines. The 1800 instances equally represent humans and industrial vehicles and are divided in 70% of training and 30% of validation. We apply data augmentation on the training set by performing a cyclic permutation of each training instance.

Identification has been performed for 137 time-domain Doppler signatures. The achieved identification accuracy during the testing phase is 92%. Wrong classification



Figure 6.4: Experimentation environment for activity recognition.

occurs according to misdetections and significative localization errors. In fact, the averaged localization error of sequences leading to a wrong classification is 30% greater than that of sequences producing a correct classification.

6.1.4 Activity recognition

The experimentation for activity recognition considers a binary classification problem in which we are interested in distinguishing between an arm movement and a fall. For the experimentation performed, we employ the Mexican hat mother wavelet given by

$$\psi(t) = \frac{2}{\sqrt{3s^4\sqrt{\pi}}} \left(1 - \frac{t^2}{s^2}\right) e^{-\frac{t^2}{2s^2}} \quad (6.1)$$

with 1024 scales s in $(0, 4]$. We trained a CNN on the time-frequency images obtained using the CWT. The deep NN architecture proposed is composed by convolutional layers, dropout and max pooling layers, and a fully connected layer. The empirical loss for training the model is the binary cross-entropy loss. We use Adam as optimization algorithm. In particular, we evaluate the training performance for different divisions of the training data due to the small dataset collected. We collected 100 measurements for each activity and we divide in 34% for training, 16% for validation, and 50% for testing. The dataset was divided randomly ensuring a balanced number of classes in each subset. To divide the data, ten seeds was chosen in order to make repeatable the dataset division and the final accuracy for a single seed was averaged over three trained models.

Fig. 6.4 shows the experimentation environment and Table 6.5 shows the accuracy obtained for each data division and corresponding trained models. We observe

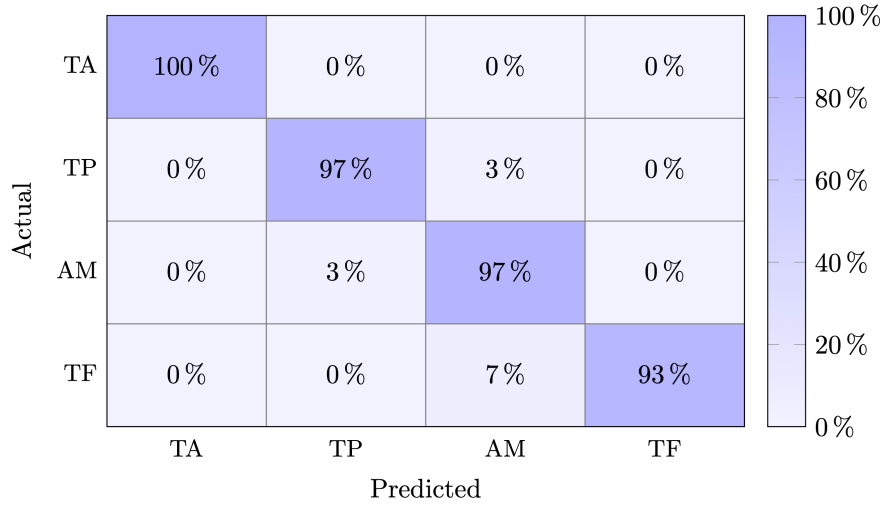


Figure 6.5: Confusion matrix for the activity recognition accuracy. The considered activities are human presence (HP), human falling (HF), no human (NH), and arm movement (AM).

that similar performance is obtained regardless of the dataset division and stochastic optimization during training. The accuracy obtained in activity recognition is 97% averaging over all the learning experiments. The developed approach performs high-accuracy classification even with the small training set.

6.2 Experimentation of joint tracking and classification

The JTC approach has been tested via two experimentation in an indoor environment at the Engineering Department, University of Ferrara. The first experimentation is performed in a scenario crowded by multiple people, while the second considers both a robot and humans. In particular, in the first experiment, classification aims to recognize whether a single collected measurement $z_k^{(j)}$, with $j \in \{0, 1, \dots, M_k\}$, is an observation or a clutter measurement. In the second one, classification aims, first, to distinguish an observation from a clutter measurement and, then, whether a tracked target is a human or a robot.

The reminder of this section describes the experimental setup and shows the experimental results obtained.

6.2.1 System setting

The experimental room is a $7.5 \times 6.5 \text{ m}^2$ area with height of 2.8 m and characterized by furnitures, walls, windows, and other scatters. The device employed to perform both

Table 6.1: Parameters employed for the experimentations.

Parameter definition at any time k	Notation	Value
DFT points for the FS dimension	N_s	512
DFT points for the SS dimension	N_c	128
DFT points for the angular dimension	N_a	64
Motion model covariance matrix	$\mathbf{V}_k$	$2.5 \cdot 10^{-2} \mathbf{1}_{4 \times 4}$
Observation model covariance matrix	$\mathbf{W}_k$	$0.5 \cdot 10^{-2} \mathbf{1}_{4 \times 4}$
Detection probability	P_d	0.98
Survival probability	P_s	0.98
Clutter Intensity	λ_k	10^{-3}
Deleter threshold	t_d	0.3
Door size	G_d	2.0 m
Gating size in the initiator	G_i	0.5 m

tracking and classification is an off-the-shelf FMCW MIMO radar with 2 transmitting and 16 receiving antennas organized as a linear array. The transmitted signals are linear chirps with 1 GHz of bandwidth, in the frequency range from $f_0 = 76$ GHz to $f_1 = 77$ GHz, with a duration $T_c = 120 \mu s$. A radar measurement is characterized by the transmission of 128 chirp signals and is repeated 3 times per second. For each collected radar measurement the 128 chirps are processed via DFT to create the RD and RDA maps.

The employed clutter mitigation technique has been introduced in Section 4.2.1 and is implemented considering an α such that

$$\bar{n}_{\mathbf{p}} + \alpha \sigma_{\mathbf{p}}^2 > \max_{n_{\mathbf{p}} \in \mathcal{A}} n_{\mathbf{p}} \quad (6.2)$$

for each coordinate $\mathbf{p}$ in the RD, or RDA, map. The set $\mathcal{A}$ contains all measurements collected during the offline phase and used to estimate the parameters $\bar{n}_{\mathbf{p}}$ and $\sigma_{\mathbf{p}}^2$ of the distribution defined in (4.34). Then, given α , by the equation (4.35), we can define a threshold $\xi_{\mathbf{p}}$ for each coordinate $\mathbf{p}$ of the RD, or RDA, map. By clustering via DBSCAN the over-threshold elements of the RD and RDA maps, we obtain a set of centroids representing the positional measurements. Specifically, the employed DBSCAN neighborhood searching radius is $\epsilon = 0.05$ and the minimum neighbors of points for creating a cluster is $N_{\text{DBSCAN}} = 2$. We define both positional measurements and positional states as 4-dimensional vectors (i.e., $m_z = n_x = 4$), each one characterized by a bi-dimensional position and a bi-dimensional velocity.

Multi-target tracking is performed considering linear and Gaussian models. In particular, the positional prediction employs the constant velocity motion model, which assumes that the target velocity is constant between two subsequent measurements. To experiment the JTC approach considering both hard and soft-decision based data association, we implemented the JPDA and the GNN algorithms. On one hand, the GNN algorithm update the target position using only the greediest measurement provided by completing data association. On the other hand, the JPDA algorithm assigns a data association weight to each measurement and, then, consider all the weighted contributions to update the target position. For both the approaches, data association is performed considering only the measurements inside a gating region with dimension $G = 1$ m.

The track initialization relies on the GNN algorithm. At the first time instant, all the measurements are supposed potential tracks, i.e., each generates a track hypothesis. At the next time instants, the measurements that are not assigned to any existing track are either associated to a track hypothesis or considered as new potential tracks. If a track hypothesis achieves a length of $K = 4$ measurements is validated as real track. Both tracking and initialization exploit a track deleter, which evaluates the trace of the covariance matrix that characterizes a target positional state. Until the trace of the covariance matrix is below the defined threshold, the track is kept. Since an environment is, typically, composed of entrance and exit points, we define entrance and exit spatial regions, referred to as *doors*. The proposed initiator and deleter evaluates only tracks and measurements inside a door.

Classification is performed using a CNN for processing 1D time series. In particular, applying the DFT to the FS and angular dimensions of a collected radar data cube, we obtain range-angle coordinates characterized by a sequence of 128 samples, i.e., one sample for each received chirp. Such a sequence of 128 samples is the time domain representation of the Doppler shifts generated by the target movements and is the classification metric that we employ. Thus, each measurement $\mathbf{z}_k^{(j)}$, with $j \in \{0, 1, \dots, M_k\}$ is obtained as follows. Firstly, we detect positional measurements processing RD and RDA maps, then, we extract the classification metric that correspond to the range-angle coordinate of the position identified.¹

Consider the i -th tracked target, which is characterized by its estimated positional states and its history of classification metrics $\mathbf{z}_{1:k-1}^{(\theta_{1:k-1}^{(i)})}$. The j -th measurement can be associated to the considered target to generate an association hypothesis $\tilde{\mathbf{z}}_{1:k}^{(\theta_{1:k}^{(i)})}$, where $\theta_k^{(i)} = j$. Targets classification is performed considering only the last W measurements. In particular, the parameter W defines the size of the time window and is chose accord-

¹Note that $\mathbf{z}_k^{(j=0)}$ is the measurement provided by the prediction density. The classification metric is extracted relying on the range-angle coordinate given by the predicted positional state.

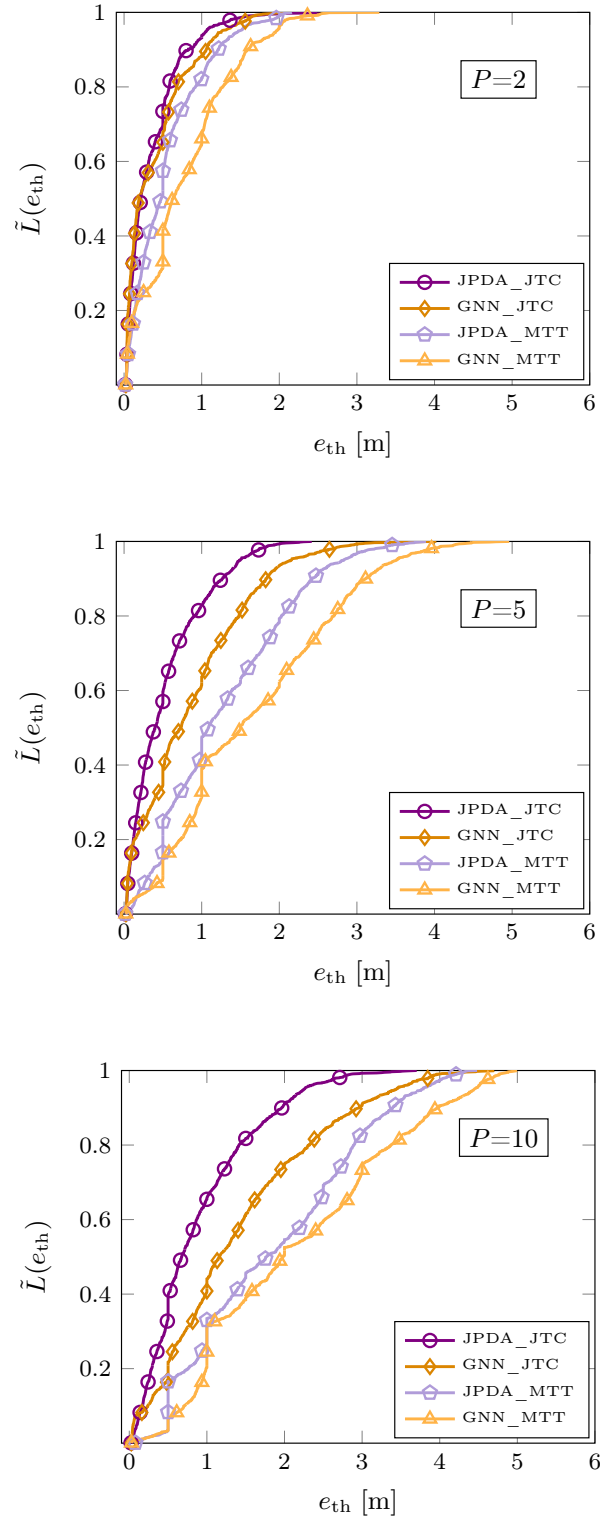


Figure 6.6: ECDF of the localization errors evaluated using the GOSPA metric and obtained tracking 5 humans in a complex wireless environment. Comparison between conventional MTT and the JTC approach, leveraging JPDA and GNN algorithms.

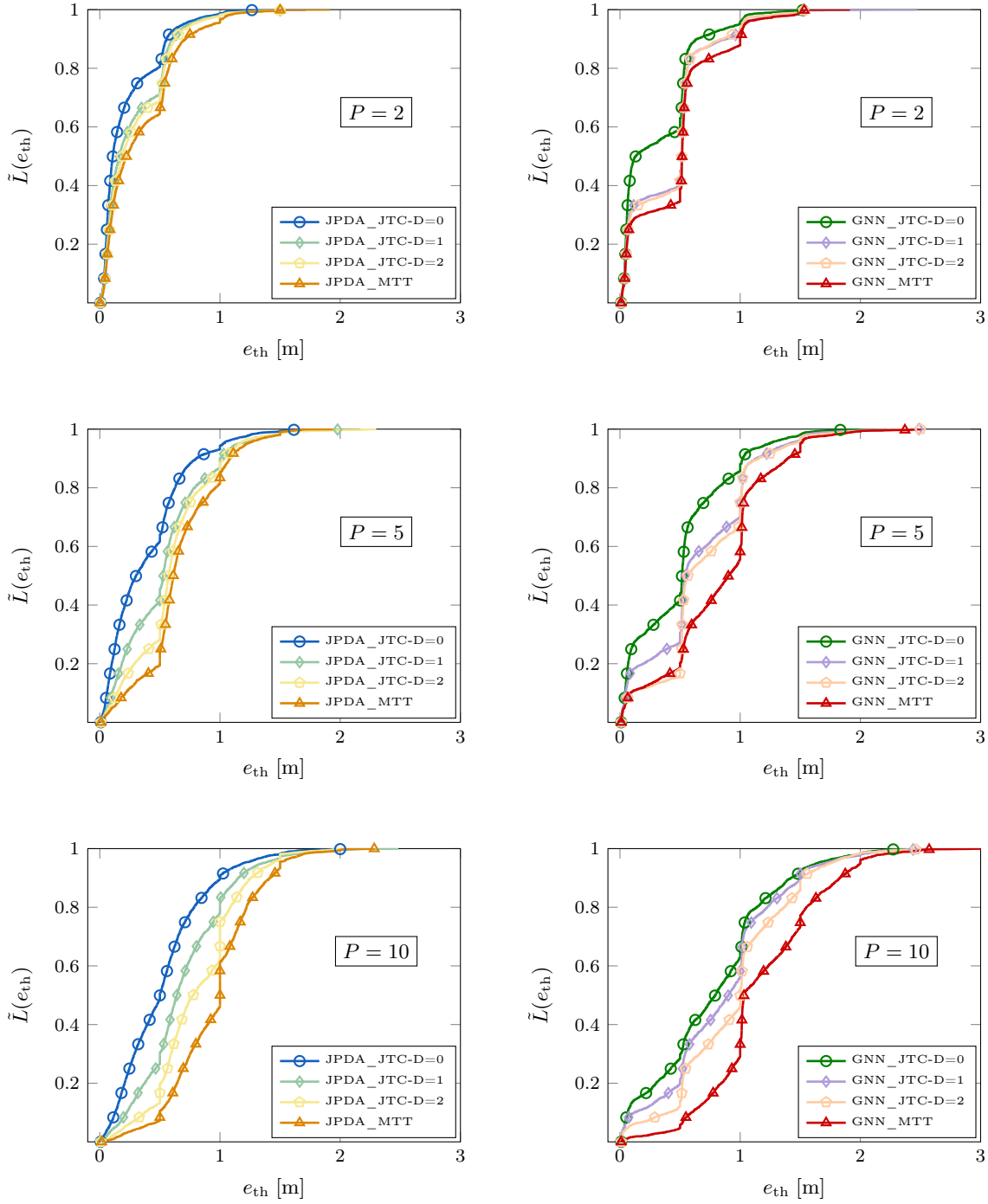


Figure 6.7: ECDF of the localization errors evaluated using the GOSPA metric and obtained tracking a robot and multiple humans in a complex wireless environment. Comparison between conventional MTT and the JTC approach varying the overlap percentage of the classification metric employed and leveraging JPDA and GNN algorithms.

ing to the type of classification to perform. Therefore, the association hypothesis can be reduced to $\tilde{\mathbf{z}}_{k-W:k}^{(\theta_{1:k}^{(i)})}$, with $W \in \{0, 1, \dots, k-2\}$. The identification performance varying W and the consequent effects for the JTC approach are discussed in Section 6.2.2. Table 6.1 summarizes further parameters used in the two experimentations performed.

In both the experimentations, classification is performed employing the same CNN. The implemented CNN is composed of a (i) 1D convolutional input layer with 128 filters and kernel size of 2; (ii) max pooling layer with pool size of 2; (iii) 1D convolutional layer with 64 filters and kernel size of 2; (iv) max pooling layer with pool size of 2; (v) a fully connected layer with 16 units; and (vi) an output layer. The inner layers employ the rectified linear unit (ReLU) as activation function, while the output layer the *softmax* function. Training is performed employing the *categorical crossentropy* as empirical loss function and Adam optimizer with a learning rate of 10^{-3} .

Data for training have been collected in a different indoor environment from the one proposed for the experimentation, using the same robot, but different people. In particular, the training area is a $4 \times 5 \text{ m}^2$ space with several metallic surfaces and other scatters, e.g., chairs and cabinets. We collected 17328 clutter measurements, 11577 humans observations, and 5800 robot observations, which we equally divided in 70 % of training and 30 % of validation. To reduce the overfitting issues, we implement early stopping by employing, during the testing, the trained model, which has achieved the best accuracy over the validation data.

6.2.2 Experimental results

The accuracy improvements in tracking, provided by the JTC approach, are quantified via the GOSPA metric comparing the estimated positions with a camera-based ground-truth [144]. In particular, the GOSPA metric allows to quantify the localization errors for a detected target and to penalize miss-detection and false alarms. In the proposed experimentation we set the GOSPA parameters as follows. The order of the metric equal to 2, the maximum allowable localization error equal to the gating region size G , and any miss-detection, or false alarm, costs is $G^2/2 \text{ m}$.

In order to make tracking and classification even more challenging to experiment the JTC approach independently from the proposed clutter mitigation and detection techniques, we increase the number of measurements at any time step. We introduce number of detections, uniform distributed inside a gating region, which allows to achieve P measurements per target. This results in challenging data association characterized by at least $P - 1$ clutter measurements for each target at any time step. In fact, when targets are closer to each other, their gating regions overlap, leading to an even increased number of clutter measurements.

Fig. 6.6 shows the ECDF of the localization error obtained tracking 5 humans in

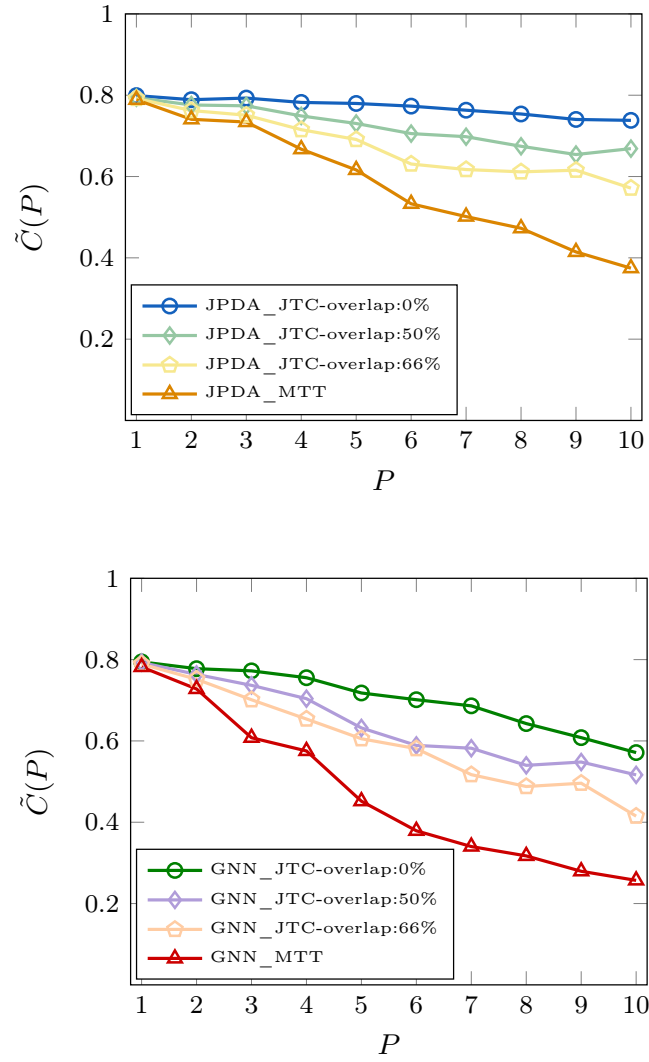


Figure 6.8: Accuracy in human and robot identification, exploiting training data collected in a different complex wireless environment for the one proposed for the experimentation.

a complex wireless environment. We collected measurements for 60s and, defined a P , we repeated 10 times the performance evaluation varying randomly the position of the introduced detections. Classification aims to distinguish whether a single collected measurement $\mathbf{z}_k^{(j)}$, with $j \in \{0, 1, \dots, M_k\}$, is an observation or a clutter measurement. Then, the probability of a measurement-to-target association, given by the JTC data association, is independent of the previous associations, i.e., $W = 0$ and, therefore, $\mathbf{z}_{k:k}^{(\theta_{k:k}^{(i)})}$, with $i \in \{1, \dots, N_k\}$. The trained model employed for completing classification achieved 94.5 % of accuracy on the validation data. The JTC approach outperforms conventional MTT by reducing the localization errors thanks to an enhanced data association. Both GNN and JPDA algorithms demonstrate improved tracking accuracy by exploiting the JTC association scores for completing hard and soft-decision data

association, respectively.

Fig. 6.7 shows the ECDF of the localization error obtained tracking a robot and multiple humans in a complex wireless environment. We collected measurements for 90 s and, defined a P , we repeated 10 times the performance evaluation varying randomly the position of the introduced detections. Classification evaluates an association hypothesis $\tilde{\mathbf{z}}_{k-W:k}^{(\theta_{1:k}^{(i)})}$, where $W \in \{0, 1, 2\}$. The JTC approach outperforms the conventional MTT for all the considered value of W . The choice of W defines the number of measurements considered to complete classification and impact on the tracking accuracy gain obtained employing the JTC approach.

Consider the NN proposed as classifier, the parameters W determines the weight that the measurement collected at the current time instant has in the evaluation of the classification score. Since the classification metric is a sequence of measurements, if $W = 0$, the collected measurement $\mathbf{z}_k^{(j)}$ represents the 100 % of the classification metric, while with $W = 1$ the 50 %, and with $W = 2$ the 33 %. Increasing W allows to extend the number of measurements that compose a classification metric and then the observation time. A larger observation time allows to achieve a more reliable classifier. However, as W increases, the contribution given by the measurement $\mathbf{z}_k^{(j)}$ to the classification score decreases.

The proposed classification task aims to distinguish among human, robot and clutter. The trained models employed achieved, on validation data, an accuracy of 86.5 %, 91.2 % and 92.9 % for $W = 0, 1, 2$, respectively. The experimental results show that increasing the observation time the classifier results more accurate, while the JTC tracking accuracy tends to the conventional MTT. Greater W , increases the overlap between time-subsequent sequence of classification features reducing the impact of the new measurements on the JTC data association score.

Fig. 6.8 shows the classification performance to distinguish between human and robot. On the same training data employed for performing JTC data association, we trained a model achieving 94.7 % of accuracy on validation data, by considering $W = 4$. Then, given the positional states estimated over time using either JTC or MTT, we extract the classification features to process using the trained model. The JTC approach allows to consider classification since the inference of the target positional state and outperforms classification of signal features extracted using positional states estimated based on conventional MTT.

6.3 Field-trials and technology readiness

The proposed approaches for multitarget tracking and classification have been validated via experimentation in multiple environments considering different types of robots. The rest of this section describes videos of experimentations performed as outcome of national and international projects. The proposed demonstration achieved a technology readiness level (TRL) of 5, since it has been performed field trials in real operation environments.

6.3.1 Tracking of people in an indoor environment

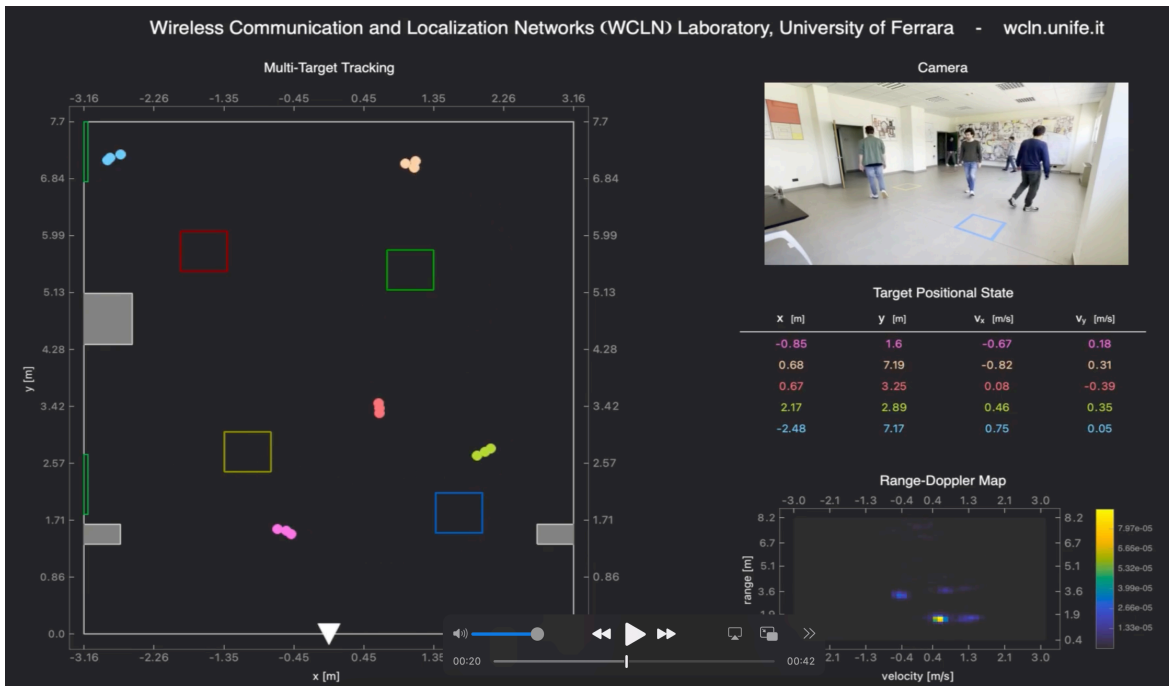


Figure 6.9: The video showing tracking of people in an indoor environment is available at <https://wcln.unife.it/multitarget-tracking-indoor-environment>.

This video presents wireless sensing for tracking multiple targets in an indoor environment at the Department of Engineering, University of Ferrara. The video GUI is organized as follows. On the left, the 2D map of the environment is designed to show the target trajectories. On the right, from top to bottom, the GUI contains the camera-based ground-truth, a table summarizing target positional information, and the real-time RD map.

The experimentation has been performed employing an off-the-shelf automotive MIMO radar operating at 76 GHz. Processing latency requires 100 ms. Tracking is performed using JPDA algorithm. The research activity has been conducted within the ONR Global projects nCOA (non-Collaborative Object Analytics).

6.3.2 Tracking of people in an outdoor environment

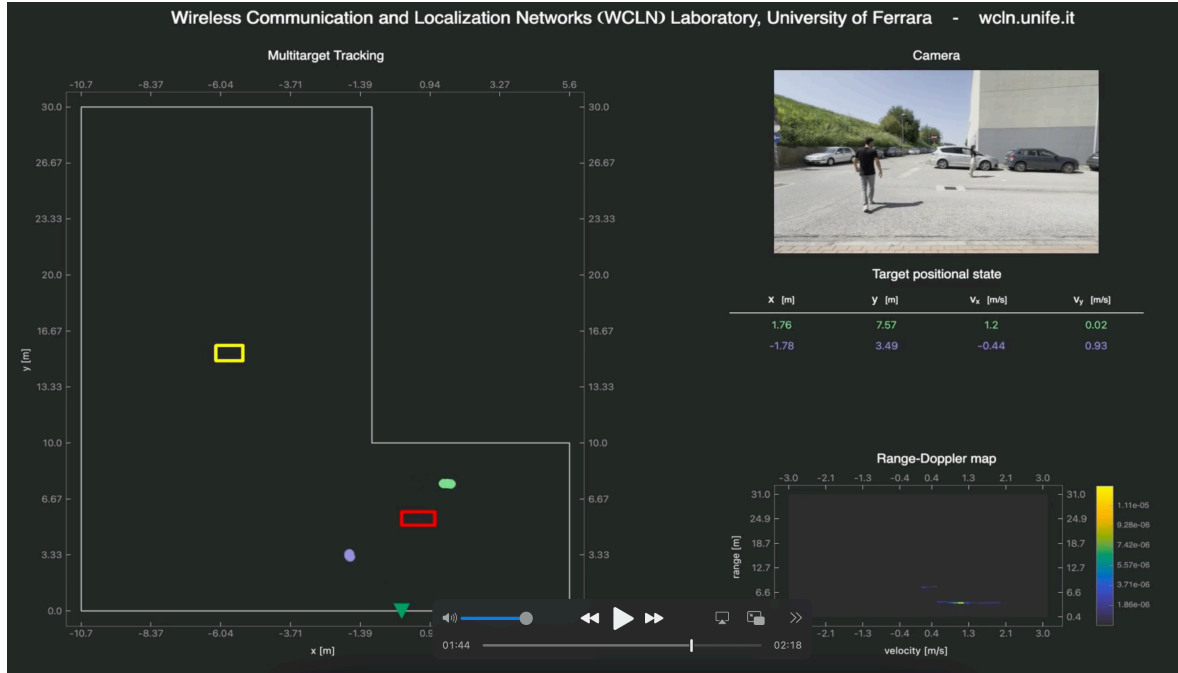


Figure 6.10: The video showing tracking of people in an outdoor environment is available at <https://wcln.unife.it/multitarget-tracking-outdoor-environment>.

This video presents wireless sensing for tracking multiple targets in an outdoor environment at the Department of Engineering, University of Ferrara.

The video GUI is organized as follows. On the left, the 2D map of the environment is designed to show the target trajectories. On the right, from top to bottom, the GUI contains the camera-based ground-truth, a table summarizing target positional information, and the real-time RD map.

The experimentation has been performed employing an off-the-shelf automotive MIMO radar operating at 76 GHz. Processing latency requires 100 ms. Tracking is performed using JPDA algorithm. The research activity has been conducted within the Programma di supporto alle tecnologie emergenti 5G project CTE (Casa delle Tecnologie Emergenti).

6.3.3 Multitarget tracking and classification in an indoor environment

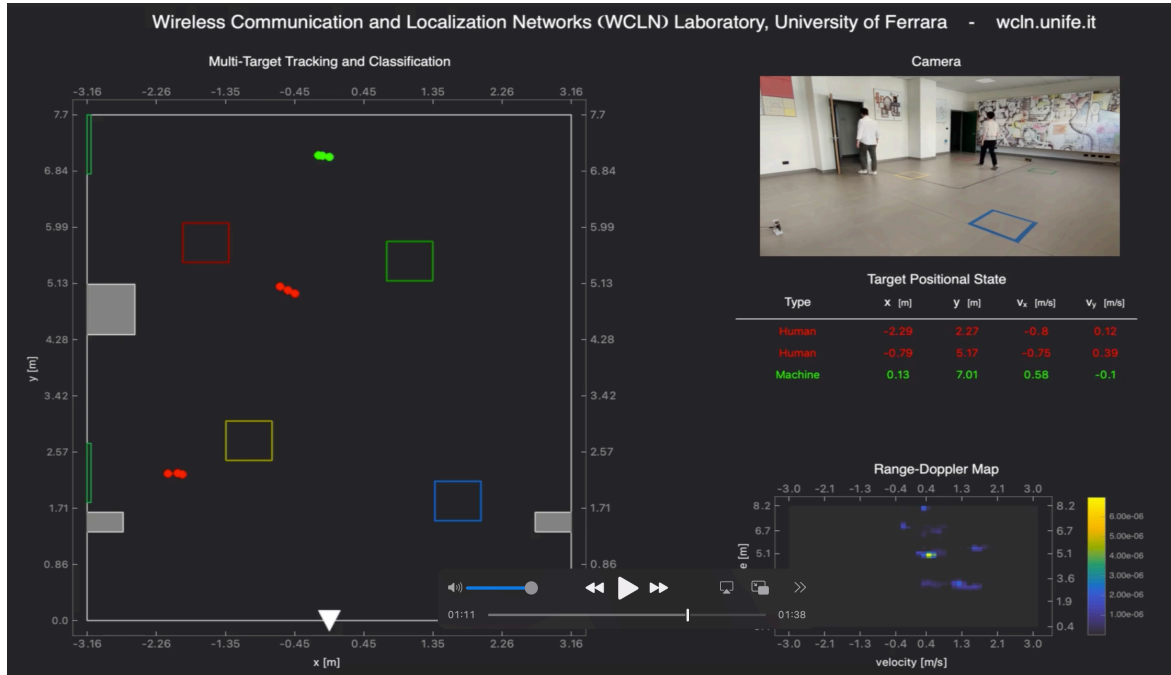


Figure 6.11: The video showing multitarget tracking and classification in an indoor environment is available at <https://wcln.unife.it/multitarget-tracking-classification-indoor-environment>.

This video presents wireless sensing for tracking and classification of multiple targets in an indoor environment at the Department of Engineering, University of Ferrara. The video GUI is organized as follows. On the left, the 2D map of the environment is designed to show the target trajectories. In particular, target color defines the target type, i.e., red for humans and green for robots. On the right, from top to bottom, the GUI contains the camera-based ground-truth, a table summarizing target positional information and class, and the real-time RD map.

The experimentation has been performed employing an off-the-shelf automotive MIMO radar operating at 76 GHz. Processing latency requires 100 ms. Tracking is performed using JPDA algorithm and classification is based on NN for processing μ D features generated by target movements. The research activity has been conducted within the ONR Global projects nCOA (non-Collaborative Object Analytics).

6.3.4 Multitarget tracking and classification in an industrial pilot line

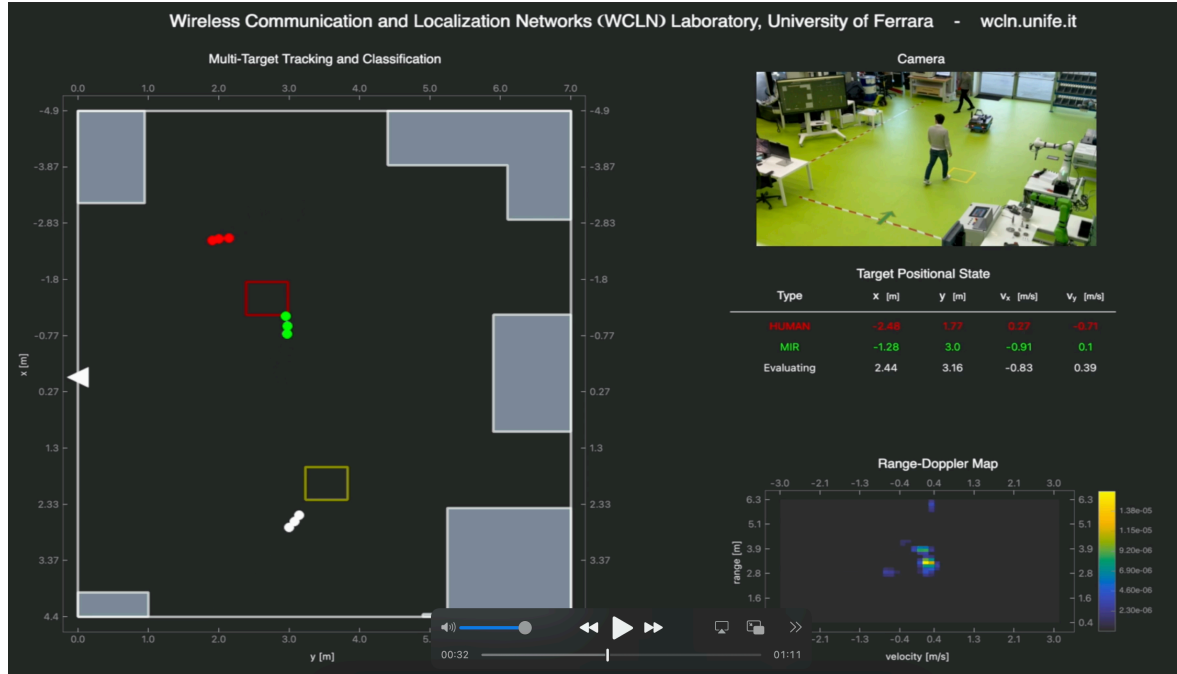


Figure 6.12: The video showing multitarget tracking and classification in an industrial pilot line is available at <https://wcln.unife.it/multitarget-tracking-classification-pilot-line>.

This video presents wireless sensing for tracking and classification of multiple targets in an industrial environment at Birex, the competence center for big data and Industry 4.0. The video GUI is organized as follows. On the left, the 2D map of the environment is designed to show the target trajectories. In particular, target color defines the target type, i.e., red for humans and green for robots. On the right, from top to bottom, the GUI contains the camera-based ground-truth, a table summarizing target positional information and class, and the real-time RD map.

The experimentation has been performed employing an off-the-shelf automotive MIMO radar operating at 76 GHz. Processing latency requires 100 ms. Tracking is performed using JPDA algorithm and classification is based on NN for processing μD features generated by target movements. The research activity has been conducted within the ONR Global projects nCOA (non-Collaborative Object Analytics) and the NextGenerationEU Italian NRRP RESTART Program project TAP-IN (Technologies, Algorithms, and Protocols for use cases of Industrial Networks)

6.3.5 Multitarget tracking and classification in an industrial environment

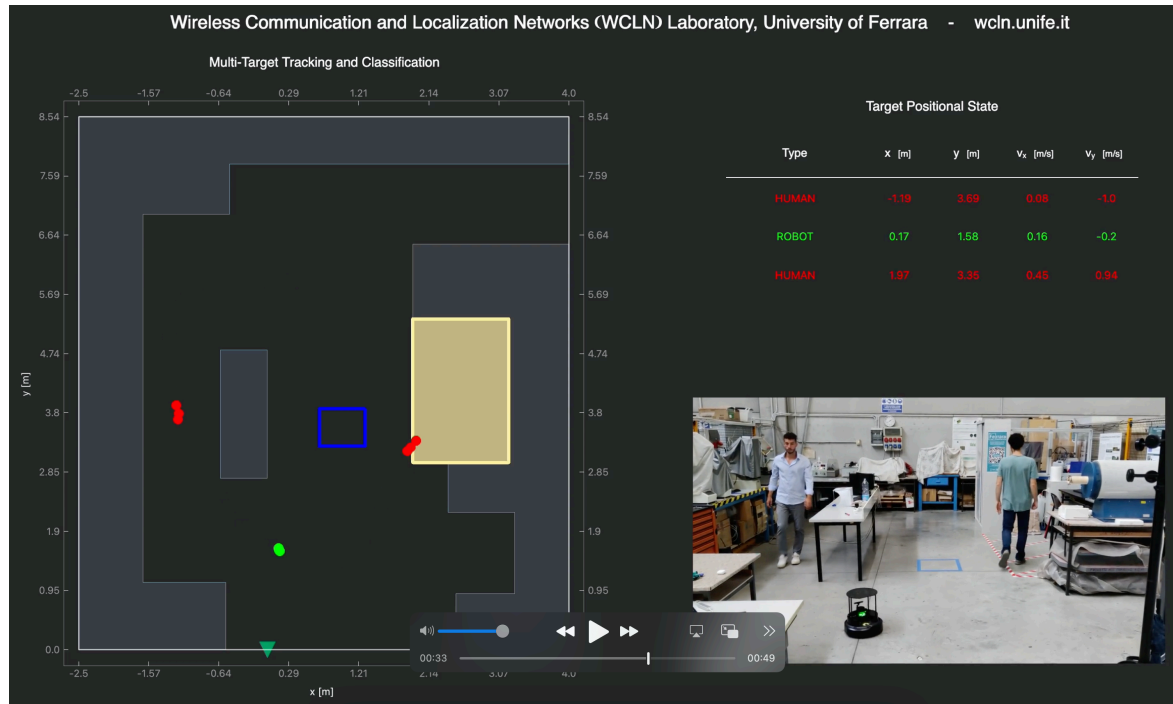


Figure 6.13: The video showing multitarget tracking and classification in an industrial environment is available at <https://wcln.unife.it/multitarget-tracking-classification-industrial-environment>.

This video presents wireless sensing for tracking and classification of multiple targets in an indoor environment at the Department of Engineering, University of Ferrara. The video GUI is organized as follows. On the left, the 2D map of the environment is designed to show the target trajectories. In particular, target color defines the target type, i.e., red for humans and green for robots. On the right, from top to bottom, the GUI contains the camera-based ground-truth, a table summarizing target positional information and class, and the real-time RD map.

The experimentation has been performed employing an off-the-shelf automotive MIMO radar operating at 76 GHz. Processing latency requires 100 ms. Tracking is performed using JPDA algorithm and classification is based on NN for processing μ D features generated by target movements. The research activity has been conducted within the NextGenerationEU Italian NRRP RESTART Program project TAP-IN (Technologies, Algorithms, and Protocols for use cases of Industrial Networks)

Chapter remark

The experimental results demonstrate that integrated tracking and classification provides accurate sensing performance in both indoor and industrial environments, achieving sub-meter localization precision and classification accuracy exceeding 90%. The JTC approach further enhances these results by jointly estimating position and class within a unified Bayesian framework. The JTC method outperforms the ITC framework in terms of accuracy and reliability, especially in complex wireless environments.

Chapter 7

Conclusion

This thesis addresses multimodal sensing of unconnected targets in complex wireless environments to enable situational awareness in xG wireless networks. As confirmed by 3GPP, tracking and classification of multiple targets is essential to achieve context awareness in ISAC systems. This thesis introduces novel approaches for multimodal tracking and classification of unconnected targets based on probabilistic frameworks and AI-based tools. In addition, the development of a microservice-based architecture guarantees distributed and scalable processing for information fusion and analytic extraction. The contributions of this dissertation are summarized in the following.

Chapter 3 proposes a radar-based framework for integrated tracking and classification of multiple targets. The framework combines clutter mitigation, target detection, multitarget tracking and NN-based classification of Doppler signatures. Experimental results using an FMCW MIMO radar operating at millimeter-wave frequencies validate the proposed approach and demonstrate high localization accuracy and reliable class inference, even in complex wireless environments.

Chapter 4 introduces an approach for JTC of multiple targets. The proposed approach jointly estimates target positions and classes with a derived closed-form solution to ensure computational tractability for real-time applications. Experimental validation shows that the joint approach improves accuracy of both tracking and classification compared to integrated methods.

Chapter 5 designs a microservice-based architecture for multimodal sensing to enable distributed processing of heterogeneous data sources, including radar and camera sensors. The architecture integrates asynchronous communication via RabbitMQ to support modular and scalable services.

Chapter 6 proposes field-trials and experimentation performed to validate the developed frameworks in complex wireless environments. The experimental results show the high-accuracy of the proposed frameworks demonstrating a sub-meter localization error in 98% of cases and a classification accuracy above 90% for target identification.

In addition, a latency of 100 ms guarantees real-time operation.

This thesis proposes novel approaches for the deployment of sensing-based services in xG wireless networks. Simultaneous localization and classification of multiple unconnected targets enable the development of innovative applications, including those oriented to the fragile assistance and the smart surveillance. The capability of tracking and classifying, jointly, humans, vehicles, and objects support the development of use cases fulfilling the 3GPP requirements in terms of accuracy, availability, and latency. Multimodal wireless sensing implemented within a microservice-based architecture enable the deployment of distributed and scalable services, providing information superiority to achieve situation awareness. The outcomes of this thesis paves the way for the development of use cases for the xG of wireless networks capable of achieving context awareness, even in complex wireless environments.

List of contributions

The contributions of the author are listed in the following.

Publications

- C. A. Gómez-Vega, A. Vaccari, F. Ferrari, R. Marini, M. Z. Win, A. Conti, "Microservice-based Architecture for Situation Awareness in Network Digital Twin," in *IEEE Networking Lett.*, *Early Access*.
- H. Park, A. Vaccari, A. Conti, S. Kim, M. Z. Win, "Semantic Near-field Beam Focusing and Tracking for Integrated Sensing and Communication," *Proc. IEEE Global Telecomm. Conf.*, Taipei, Taiwan, Dec. 2025, pp. 1–6.
- A. Vaccari, M. Z. Win, and A. Conti, "Tracking and identification of targets via mmWave MIMO radar," in *Proc. Military Commun. Conf.*, Washington, DC, USA, Oct. 2024, pp. 336–341.
- A. Vaccari, C. A. Gómez-Vega, G. Torsoli, M. Cloâtre, M. Z. Win, and A. Conti, "Activity recognition from Doppler signatures via mmWave MIMO radar," in *Proc. The Korean Inst. Commun. and Inf. Sciences Winter Conf.*, Jeju, South Korea, Jun. 2024, pp. 149–150, **Best Paper Award**.
- A. Conti, G. Torsoli, C. A. Gómez-Vega, A. Vaccari, G. Mazzini, and M. Z. Win, "3GPP-compliant datasets for xG location-aware networks," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 473–484, Apr. 2024.

Dataset

- A. Conti, G. Torsoli, C. A. Gómez-Vega, A. Vaccari, and M. Z. Win, xG-Loc: 3GPP-compliant datasets for xG location-aware networks, *IEEE Dataport*, Dec. 2023. [Online]. Available: <https://dx.doi.org/10.21227/rper-vc03>

Dissemination

- A. Vaccari, C. A. Gómez-Vega, E. Sangiorgi, and A. Conti, AI-based localization and sensing of things, *one6G Summit*, Bologna, Italy, Sep. 2025, Live Demonstration.
- A. Vaccari, C. A. Gómez-Vega, E. Sangiorgi, and A. Conti, Multimodal wireless sensing of things in industrial environments, *Orizzonte 2030: l'Italia tra 5G e 6G*, Bologna, Italy, Sep. 2025, Live Demonstration.
- A. Vaccari and A. Conti, "Real-time tracking and classification of multiple targets via wireless sensing," *RESTART Plenary Dissemination Workshop*, Naples, Italy, Jun. 2025, Live Demonstration.
- A. Vaccari and A. Conti, Real-Time Tracking and Classification of Multiple Targets, *one6G Summit*, Valencia, Spain, Sep. 2024, Live Demonstration.
- A. Vaccari and A. Conti, Multi-target Tracking and Classification via MIMO Radar at mmWaves, *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, Valencia, Spain, Sep. 2024, Live Demonstration, **Best Demo Award**.
- A. Vaccari and A. Conti, Multi-target Tracking and Classification via MIMO radar at mm-Waves, *one6G Summit*, Munich, Germany, Nov. 2023, Live Demonstration.

Bibliography

- [1] J. Kim, A. Conti, and M. Z. Win, “Travel demand modeling and estimation for high-dimensional mobility,” *IEEE Trans. Mobile Comput.*, vol. 24, no. 3, pp. 1264–1277, Mar. 2025.
- [2] N. Kuxdorf-Alkirata, G. Maus, M. Gemci, and D. Brckmann, “A passive fingerprinting approach for device-free surveillance and localization applications using a bluetooth low energy infrastructure,” in *Int. Conf. Intell. Eng. Syst.*, Reykjavk, Iceland, Jul. 2020, pp. 31–36.
- [3] Z. Wang, Z. Liu, Y. Shen, A. Conti, and M. Z. Win, “Holographic localization with synthetic reconfigurable intelligent surfaces,” *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 4, pp. 603–618, May 2024, special issue on *Recent Advances in Wideband Signal Processing for Classical and Quantum Synthetic Apertures*.
- [4] Z. Wang *et al.*, “From personalized medicine to population health: A survey of mHealth sensing techniques,” *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15 413–15 434, Sep. 2022.
- [5] B. M. Masini *et al.*, “Toward the integration of ADAS capabilities in V2X communications for cooperative driving,” in *Int. Conf. Elect. Electron. Techn. Autom.*, Turin, Italy, Nov. 2020.
- [6] M. Z. Win, Z. Wang, Z. Liu, Y. Shen, and A. Conti, “Location awareness via intelligent surfaces: A path toward holographic NLN,” *IEEE Veh. Technol. Mag.*, vol. 17, no. 2, pp. 37–45, Jun. 2022, special issue on *Backscatter and Reconfigurable Intelligent Surface Empowered Wireless Communications in 6G*.
- [7] R. Zheng, Y. Zhao, and B. Chen, “Device-free and robust user identification in smart environment using WiFi signal,” in *Int. Symp. Parallel Distrib. Process. Appl. and Int. Conf. Conference Ubiquitous Comput. Commun.*, Guangzhou, China, Dec. 2017, pp. 1039–1046.

- [8] K. Lin, M. Chen, J. Deng, M. M. Hassan, and G. Fortino, “Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings,” *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1294–1307, Jul. 2016.
- [9] M. Moussa and M. Youssef, “Smart devices for smart environments: Device-free passive detection in real environments,” in *Int. Conf. Pervasive Comput. Commun.*, Galveston, TX, USA, May 2009, pp. 1–6.
- [10] G. Kwon, W. Shin, A. Conti, W. C. Lindsey, and M. Z. Win, “Access-backhaul strategy via gNB cooperation for integrated terrestrial-satellite networks,” *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1403–1419, May 2024, special issue on Space Communications New Frontiers: From Near Earth to Deep Space.
- [11] F. Morselli, S. M. Razavi, M. Z. Win, and A. Conti, “Soft information based localization for 5G networks and beyond,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9923–9938, Dec. 2023.
- [12] Z. Wang, Z. Liu, Y. Shen, A. Conti, and M. Z. Win, “Location awareness in beyond 5G networks via reconfigurable intelligent surfaces,” *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2011–2025, Jul. 2022, special issue on *Integrated Sensing and Communication*.
- [13] M. Gareis, C. Carlowitz, and M. Vossiek, “A MIMO UHF-RFID SAR 3D locating system for autonomous inventory robots,” in *Int. Conf. Microw. Intell. Mobility*, Linz, Austria, Nov. 2020, pp. 1–4.
- [14] S. Hu, Q. Wu, and X. Wang, “Energy management and trajectory optimization for UAV-enabled legitimate monitoring systems,” *IEEE Wireless Commun.*, vol. 20, no. 1, pp. 142–155, Jan. 2021.
- [15] Y. Cui, F. Liu, X. Jing, and J. Mu, “Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges,” *IEEE Netw.*, vol. 35, no. 5, pp. 158–167, Oct. 2021.
- [16] A. Liu *et al.*, “A survey on fundamental limits of integrated sensing and communication,” *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 994–1034, 2nd Quart. 2022.
- [17] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, “Soft information for localization-of-things,” *Proc. IEEE*, vol. 107, no. 11, pp. 2240–2264, Nov. 2019.

- [18] G. Geraci *et al.*, “What will the future of UAV cellular communications be? A flight from 5G to 6G,” *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1304–1335, 3rd Quart. 2022.
- [19] Z. Liu, A. Conti, S. K. Mitter, and M. Z. Win, “Continuous-time distributed filtering via a Gaussian feedback channel,” *IEEE J. Sel. Areas Commun.*, vol. 43, pp. 1–16, 2025, special issue on *Co-Design of Communication, Computing, and Control in Industrial Cyber-Physical Systems*, to appear.
- [20] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, Jun. 2020.
- [21] S. Bartoletti, S. Mazuelas, A. Conti, and M. Z. Win, “Efficient localization via soft information with generic sensing measurements,” *IEEE Trans. Wireless Commun.*, vol. 24, no. 7, pp. 5400–5414, Jul. 2025.
- [22] W. Y. B. Lim *et al.*, “Federated learning in mobile edge networks: A comprehensive survey,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, thirdquarter 2020.
- [23] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, “6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions,” *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, Jul. 2020.
- [24] C. Huang *et al.*, “Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends,” *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.
- [25] *Technical Specification Group Services and System Aspects; Service requirements for the 5G system; Stage 1*, 3rd Generation Partnership Project 3GPP™ TS 22.261 V19.1.0, Dec. 2022, Release 19.
- [26] *Technical Specification Group TSG SA; Study on 6G Use Cases and Service Requirements; Stage 1*, 3rd Generation Partnership Project 3GPP™ TR 22.870 V0.3.1, Jun. 2025, Release 20.
- [27] *Technical Specification Group TSG SA; Service requirements for Integrated Sensing and Communication*, 3rd Generation Partnership Project 3GPP™ TS 22.137 V19.1.0, Mar. 2024, Release 19.

- [28] *Technical Specification Group TSG SA; Feasibility Study on Integrated Sensing and Communication*, 3rd Generation Partnership Project 3GPP™ TR 22.837 V19.4.0, Jun. 2024, Release 19.
- [29] *Technical Specification Group Services and System Aspects; Feasibility Study on Integrated Sensing and Communication*, 3rd Generation Partnership Project 3GPP™ TR 22.837 V19.0.0, Jun. 2023, Release 19.
- [30] L. Pucci, E. Paolini, and A. Giorgetti, “System-level analysis of joint sensing and communication based on 5G new radio,” *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2043–2055, Jul. 2022.
- [31] J. A. Zhang *et al.*, “Enabling joint communication and radar sensing in mobile networks a survey,” *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 306–345, 1st Quart. 2022.
- [32] G. Kwon, A. Conti, H. Park, and M. Z. Win, “Joint communication and localization in millimeter wave networks,” *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1439–1454, Nov. 2021, special issue on *Joint Communication and Radar Sensing for Emerging Applications*.
- [33] X. Li, Y. Cui, J. A. Zhang, F. Liu, D. Zhang, and L. Hanzo, “Integrated human activity sensing and communications,” *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 90–96, May 2023.
- [34] G. Kwon, Z. Liu, A. Conti, H. Park, and M. Z. Win, “Integrated localization and communication in 3GPP industrial environments,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Seoul, South Korea, Apr. 2024, pp. 9191–9195.
- [35] A. Liu *et al.*, “A survey on fundamental limits of integrated sensing and communication,” *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 994–1034, Secondquarter 2022.
- [36] J. Pegoraro, J. O. Lacruz, F. Meneghello, E. Bashirov, M. Rossi, and J. Widmer, “RAPID: Retrofitting IEEE 802.11ay access points for indoor human detection and sensing,” *IEEE Trans. Mobile Comput.*, 2023, textitely access.
- [37] J. A. Zhang, K. Wu, X. Huang, Y. J. Guo, D. Zhang, and R. W. Heath, “Integration of radar sensing into communications with asynchronous transceivers,” *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 106–112, Nov. 2022.

- [38] S. Bartoletti, A. Conti, and M. Z. Win, “Passive radar via LTE signals of opportunity,” in *Proc. IEEE Workshop on Advances in Network Localization and Navigation (ANLN), Int. Conf. Commun.*, Sydney, Australia, Jun. 2014, pp. 181–185.
- [39] Y. Zeng, R. Zhang, and T. J. Lim, “Wireless communications with unmanned aerial vehicles: opportunities and challenges,” *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.
- [40] J. Pegoraro, P. Saucedo, J. O. Lacruz, M. Rossi, and J. Widmer, “DISC: A dataset for integrated sensing and communications in mmWave systems,” *IEEE Commun. Mag.*, vol. 63, no. 10, pp. 94–100, Oct. 2025.
- [41] Z. Zhang, F. Shi, C. Jia, M. Zhao, and X. Cheng, “Collaborative association network for multi-view multi-human association and tracking using constraint optimization and object search,” in *Int. Conf. Acoust., Speech Sig. Process.*, Hyderabad, India, Apr. 2025, pp. 1–5.
- [42] F. Colone, D. Pastina, P. Falcone, and P. Lombardo, “WiFi-Based passive ISAR for high-resolution cross-range profiling of moving targets,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3486–3501, Jun. 2014.
- [43] L. Pucci, T. Bacchielli, and A. Giorgetti, “Cooperative maximum likelihood target position estimation for MIMO-ISAC networks,” *IEEE Wireless Commun. Lett.*, vol. 14, no. 5, pp. 1531–1535, May 2025.
- [44] E. Favarelli, E. Matricardi, L. Pucci, W. Xu, E. Paolini, and A. Giorgetti, “Sensor fusion and resource management in MIMO-OFDM joint sensing and communication,” *IEEE Trans. Veh. Technol.*, vol. 74, no. 6, pp. 9284–9298, Jun. 2025.
- [45] M. Lotti, G. Pasolini, A. Guerra, F. Guidi, R. D’Errico, and D. Dardari, “Radio SLAM for 6G systems at THz frequencies: Design and experimental validation,” *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 4, pp. 834–849, Jul. 2023.
- [46] A. Bazzi, B. M. Masini, A. Zanella, and G. Pasolini, “On the use of OFDMA for next generation vehicular ad hoc networks,” in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, London, UK, Sep. 2013, pp. 2223–2228.
- [47] L. Lin, M. Wang, Z. Shi, M. S. Greco, and F. Gini, “ISAC-enabled near-field high-mobility uav localization: Analysis, bounds and algorithm,” *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–15, Aug. 2025.

- [48] S. Bartoletti, Z. Liu, M. Z. Win, and A. Conti, "Device-free localization of multiple targets in cluttered environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 5, pp. 3906–3923, Oct. 2022.
- [49] F. Gini, M. Greco, and A. Farina, "Clairvoyant and adaptive signal detection in non-gaussian clutter: a data-dependent threshold interpretation," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1522–1531, Jun. 1999.
- [50] J. Billingsley, A. Farina, F. Gini, M. Greco, and L. Verrazzani, "Statistical analyses of measured radar ground clutter data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 2, pp. 579–593, Apr. 1999.
- [51] A. Quirini, G. P. Blasone, F. Colone, and P. Lombardo, "Supervised DPCA scheme based on reciprocal filter for clutter cancelation from moving OFDM radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 61, no. 3, pp. 6154–6172, Jun. 2025.
- [52] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
- [53] B.-N. Vo and W.-K. Ma, "The gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.
- [54] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. USA: Artech House, Inc., 2007.
- [55] R. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*. USA: Artech House, Inc., 2014.
- [56] F. Meyer *et al.*, "Message passing algorithms for scalable multitarget tracking," *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, Feb. 2018.
- [57] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-bernoulli mixture filter: Direct derivation and implementation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1883–1901, Aug. 2018.
- [58] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, Jan. 2004.
- [59] L. Svensson, D. Svensson, M. Guerriero, and P. Willett, "Set JPDA filter for multitarget tracking," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4677–4691, 2011.

- [60] S. Sato, M. Hashimoto, M. Takita, K. Takagi, and T. Ogawa, "Multilayer lidar-based pedestrian tracking in urban environments," in *IEEE Intell. Vehicles Symposium*, 2010, pp. 849–854.
- [61] P. Zhu, X. Si, and J. Liang, "A nonlinear waveform selection method for cognitive radar target tracking based on reinforcement learning," in *Int. Geosci. Remote Sens. Symp.*, Pasadena, CA, USA, Jul. 2023, pp. 4506–4509.
- [62] T. Thomas and S. S, "Comparison of nearest neighbor and probabilistic data association filters for target tracking in cluttered environment," in *Int. Conf. Comput., Commun. Autom.*, Arad, Romania, Dec. 2021, pp. 272–277.
- [63] T. Yoo, H. Bang, and W. Youn, "Adaptive DBSCAN-Based probabilistic data association filter for maritime object tracking using RADAR," *IEEE Sens. Lett.*, vol. 9, no. 6, pp. 1–4, Jun. 2025.
- [64] M. Sahal, Z. A. Said, R. E. Abdul Kadir, Z. Hidayat, Y. Bilfaqih, and A. Alkaff, "Tracking position of airborne target on SPx-Radar-Simulator using probabilistic data association filter," in *Int. Conf. Inf. Commun. Techn. Syst.*, Surabaya, Indonesia, Oct. 2021, pp. 258–263.
- [65] R. Su, J. Tang, J. Yuan, and Y. Bi, "Nearest neighbor data association algorithm based on robust kalman filtering," in *Int. Symp. Comput. Eng. Intell. Commun.*, Nanjing, China, Aug. 2021, pp. 177–181.
- [66] F. Meyer and M. Z. Win, "Scalable data association for extended object tracking," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 6, pp. 491–507, May 2020.
- [67] M. Chiani, A. Giorgetti, and E. Paolini, "Sensor radar for object tracking," *Proc. IEEE*, vol. 106, no. 6, pp. 1022–1041, Jun. 2018.
- [68] S. Bartoletti, A. Conti, A. Giorgetti, and M. Z. Win, "Sensor radar networks for indoor tracking," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 157–160, Apr. 2014.
- [69] C. A. Gómez-Vega, Z. Liu, C. A. Gutiérrez, M. Z. Win, and A. Conti, "Efficient deployment strategies for network localization with assisting nodes," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 6272–6287, May 2024.
- [70] Q. Zhang, C. Wang, and Y. Yang, "LSTM based data association for radar target tracking," in *Int. Conf. Elect. Autom. Artif. Intell.*, Guangzhou, China, Jan. 2025, pp. 1048–1052.

- [71] C. A. Gómez-Vega, G. Torsoli, M. Z. Win, and A. Conti, “Node activation for SI-based xG localization: 3GPP case studies using xG-Loc dataset,” in *Proc. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, Valencia, Spain, Sep. 2024, pp. 1–6.
- [72] Y. Liu, P. Qin, and J. Zeng, “Multi-object tracking based on feature fusion and hierarchical data association,” in *Chinese Control and Decision Conf.*, Kunming, China, May 2021, pp. 1742–1747.
- [73] X. Wang, D. Chen, and Y. Zhao, “An intelligent particle filter based joint probabilistic data association algorithm for long-duration multi-target tracking in clutter environment,” in *Int. Conf. Inform. Commun. Sig. Process.*, Zhoushan, China, Sep. 2024, pp. 823–827.
- [74] C. Ravindra, S. Rajkumar, and A. Vaidhyathan, “Probabilistic data association performance during launch vehicle tracking,” in *Int. Conf. Range Techn.*, Chandipur, Balasore, India, Feb. 2023, pp. 1–5.
- [75] W. Li, A. Yang, and L. Zhang, “Improved data association algorithm for airborne radar multi-target tracking via deep learning network,” in *Int. Geosci. Remote Sens. Symp.*, Kuala Lumpur, Malaysia, Jul. 2022, pp. 7417–7420.
- [76] R. Zhang, X. Jing, S. Wu, C. Jiang, J. Mu, and F. R. Yu, “Device-free wireless sensing for human detection: The deep learning perspective,” *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2517–2539, Feb. 2021.
- [77] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, and C. J. Spanos, “DeepSense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network,” in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, USA, May 2018, pp. 1–6.
- [78] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, and P. Willett, “Deep learning methods for vessel trajectory prediction based on recurrent neural networks,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 6, pp. 4329–4346, Dec. 2021.
- [79] I. Nirmal, A. Khamis, M. Hassan, W. Hu, and X. Zhu, “Deep learning for radio-based human sensing: Recent advances and future directions,” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 995–1019, 2nd Quart. 2021.
- [80] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, “Human action recognition from various data modalities: A review,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3200–3225, 2023.

- [81] A. Vaccari, C. A. Gómez-Vega, G. Torsoli, M. Clouâtre, M. Z. Win, and A. Conti, “Activity recognition from Doppler signatures via mmWave MIMO radar,” in *Proc. The Korean Inst. of Commun. and Inf. Sciences Winter Conf.*, Jeju, South Korea, Jun. 2024, pp. 149–150.
- [82] T. V. Nguyen, Y. Jeong, H. Shin, and M. Z. Win, “Machine learning for wideband localization,” *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1357–1380, Jul. 2015.
- [83] B. Dekker, S. Jacobs, A. Kossen, M. Kruithof, A. Huizing, and M. Geurts, “Gesture recognition with a low power FMCW radar and a deep convolutional neural network,” in *Eur. Radar Conf.*, Nuremberg, Germany, Oct. 2017, pp. 163–166.
- [84] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, “Micro-Doppler based human-robot classification using ensemble and deep learning approaches,” in *IEEE Radar Conf.*, Oklahoma City, OK, USA, Apr. 2018, pp. 1043–1048.
- [85] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, “HybridSN: Exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.
- [86] A. Vaccari, M. Z. Win, and A. Conti, “Tracking and identification of targets via mmWave MIMO radar,” in *Proc. Military Commun. Conf.*, Washington, DC, USA, Oct. 2024, pp. 336–341.
- [87] R. Mazzieri, J. Pegoraro, and M. Rossi, “Open-set gait recognition from sparse mmWave radar point clouds,” *IEEE Sens. J.*, vol. 25, no. 17, pp. 33 051–33 063, Sep. 2025.
- [88] D. Gaglione, G. Soldi, P. Braca, G. De Magistris, F. Meyer, and F. Hlawatsch, “Classification-aided multitarget tracking using the sum-product algorithm,” *IEEE Signal Process. Lett.*, vol. 27, pp. 1710–1714, Sep. 2020.
- [89] A. Azim and O. Aycard, “Detection, classification and tracking of moving objects in a 3D environment,” in *IEEE Intell. Vehicles Symposium*, 2012, pp. 802–807.
- [90] L. Pucci, E. Testi, E. Favarelli, and A. Giorgetti, “Human activities classification using biaxial seismic sensors,” *IEEE Sens. Lett.*, vol. 4, no. 10, pp. 1–4, Oct. 2020.
- [91] J. Pegoraro, J. O. Lacruz, F. Meneghello, E. Bashirov, M. Rossi, and J. Widmer, “RAPID: Retrofitting IEEE 802.11ay access points for indoor human detection and sensing,” *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4501–4519, May 2024.

-
- [92] J. Pegoraro and M. Rossi, “Human tracking with mmWave radars: a deep learning approach with uncertainty estimation,” in *Proc. IEEE Workshop on Signal Process. Advances in Wireless Commun.*, Oulu, Finland, Jul. 2022, pp. 1–5.
 - [93] M. Canil, J. Pegoraro, A. Shastri, P. Casari, and M. Rossi, “ORACLE: Occlusion-resilient and self-calibrating mmWave radar network for people tracking,” *IEEE Sens. J.*, vol. 24, no. 3, pp. 3157–3171, Feb. 2024.
 - [94] A. Shastri, M. Canil, J. Pegoraro, P. Casari, and M. Rossi, “mmSCALE: Self-calibration of mmWave radar networks from human movement trajectories,” in *IEEE Radar Conf.*, New York City, NY, USA, Mar. 2022, pp. 1–6.
 - [95] J. Pegoraro and M. Rossi, “Real-time people tracking and identification from sparse mm-Wave radar point-clouds,” *IEEE Access*, vol. 9, pp. 78 504–78 520, May 2021.
 - [96] S. Sun and Y. D. Zhang, “4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach,” *IEEE J. Sel. Areas Commun.*, vol. 15, no. 4, pp. 879–891, Jun. 2021.
 - [97] Y. Wang, T. Gu, T. H. Luan, M. Lyu, and Y. Li, “Heartprint: Exploring a heartbeat-based multiuser authentication with single mmWave radar,” *IEEE Internet Things J.*, vol. 9, no. 24, pp. 25 324–25 336, Dec. 2022.
 - [98] A. Shastri *et al.*, “A review of millimeter wave device-based localization and device-free sensing technologies and applications,” *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1708–1749, 3rd Quart. 2022.
 - [99] C. A. Gómez-Vega, O. Y. Kolawole, M. Hunukumbure, T. Mach, M. Z. Win, and A. Conti, “Device-free localization: Outdoor 5G experimentation at mm-Waves,” *IEEE Commun. Lett.*, vol. 27, no. 9, pp. 2353–2357, Sep. 2023.
 - [100] J. A. Zhang *et al.*, “An overview of signal processing techniques for joint communication and radar sensing,” *IEEE J. Sel. Areas Commun.*, vol. 15, no. 6, pp. 1295–1315, Nov. 2021.
 - [101] F. Abushakra *et al.*, “A miniaturized ultra-wideband radar for UAV remote sensing applications,” *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 3, pp. 198–201, Mar. 2022.
 - [102] Z. Li *et al.*, “Spatiotemporal processing for remote sensing of trapped victims using 4-D imaging radar,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, Apr. 2023.

- [103] F. J. Yanovsky, A. A. Pitertsev, C. M. H. Unal, and H. W. J. Russchenberg, "Multi-sensor data fusion based on millimeter radar and laser disdrometer," in *Int. Radar Conf.*, Rennes, France, Oct. 2024, pp. 1–5.
- [104] G. Adhikari, S. Halder, S. Banerjee, and S. S. Chaudhuri, "An effective use of synthetic aperture radar imaging technique over optical remote sensing and real aperture radar for mapping of earth surface features," in *Innovations Energy Management Renewable Resources*, Kolkata, India, Feb. 2021, pp. 1–6.
- [105] A. Quirini, F. Colone, and P. Lombardo, "Clutter suppression using thresholded reciprocal filter in OFDM radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 6, pp. 8316–8331, Dec. 2024.
- [106] A. Zaimbashi, M. S. Greco, and F. Gini, "Low-complexity LRT-based passive radar target detection algorithm in interfering conditions," *IEEE Trans. Radar Syst.*, vol. 2, pp. 549–560, May 2024.
- [107] J. Pegoraro, D. Solimini, F. Matteo, E. Bashirov, F. Meneghello, and M. Rossi, "Deep learning for accurate indoor human tracking with a mm-Wave radar," in *IEEE Radar Conf.*, Florence, Italy, Sep. 2020, pp. 1–6.
- [108] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. R. Liu, "ViMo: Multiperson vital sign monitoring using commodity millimeter-wave radio," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1294–1307, Feb. 2021.
- [109] B. R. Kiran *et al.*, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 4909–4926, Jun. 2022.
- [110] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58 443–58 469, Mar. 2020.
- [111] B. Vandersmissen *et al.*, "Indoor person identification using a low-power FMCW radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3941–3952, Apr. 2018.
- [112] D. Zhao, X. Zhang, X. Cui, Y. Xu, and X. Wang, "Design of MIMO imaging system by millimeter wave radar," in *Symp. on Antennas, Propag. EM Theory*, Hefei, China, Oct. 2024, pp. 1–4.
- [113] E. Dedic, D. Henry, M. Lihoreau, and H. Aubert, "3D tracking of small moving targets from a multi-target tracking algorithm applied to millimeter-wave radar images," in *European Radar Conf.*, Paris, France, Sep. 2024, pp. 176–179.

-
- [114] S. K. Joshi, S. V. Baumgartner, and G. Krieger, "Tracking and track management of extended targets in range-doppler using range-compressed airborne radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, Jun. 2022.
 - [115] K. Bell and B. Shapo, "Cognitive radar tracking with spectrum sensing and prediction," in *Int. Microw. Symp.*, Denver, CO, USA, Jun. 2022, pp. 638–640.
 - [116] Q. Zhou, Y. Yuan, H. Li, M. S. Greco, F. Gini, and W. Yi, "Target detection for distributed hybrid active-passive radars," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–18, Oct. 2025.
 - [117] Z. Hu, T. Su, D. Xu, G. Pang, and F. Gini, "A fast fourier-based near-field 3-D imaging algorithm for MIMO array," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, Jul. 2024.
 - [118] A. Zaimbashi, M. S. Greco, and F. Gini, "Integrated MIMO passive radar target detection," *IEEE Trans. Signal Process.*, vol. 72, pp. 2677–2691, May 2024.
 - [119] J. Pegoraro, F. Meneghello, and M. Rossi, "Multiperson continuous tracking and identification from mm-Wave micro-Doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 2994–3009, Apr. 2021.
 - [120] C. A. Gómez-Vega *et al.*, "Doppler spectrum measurement platform for narrow-band V2V channels," *IEEE Access*, vol. 10, pp. 27 162–27 184, 2022.
 - [121] B. I. Ahmad and S. Harman, "Tracking of target body and micro-Doppler components in drone surveillance radar," in *Int. Conf. Radar Syst.*, vol. 2022, Edinburgh, UK, Oct. 2022, pp. 260–265.
 - [122] A. G. Argello and D. Berges, "Radar classification for traffic intersection surveillance based on micro-Doppler signatures," in *Eur. Radar Conf.*, Madrid, Spain, Sep. 2018, pp. 186–189.
 - [123] Z. Chen, G. Li, F. Fioranelli, and H. Griffiths, "Personnel recognition and gait classification based on multistatic micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 669–673, Mar. 2018.
 - [124] F. Hoffmann, M. Ritchie, F. Fioranelli, A. Charlish, and H. Griffiths, "Micro-Doppler based detection and tracking of UAVs with multistatic radar," in *IEEE Radar Conf.*, Philadelphia, PA, May 2016, pp. 1–6.

- [125] J. Craley, T. S. Murray, D. R. Mendat, and A. G. Andreou, “Action recognition using micro-Doppler signatures and a recurrent neural network,” in *IEEE Annu. Conf. Inf. Sci. and Syst.*, Baltimora, MD, USA, Mar. 2017, pp. 1–5.
- [126] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, “Joint radar and communication design: Applications, state-of-the-art, and the road ahead,” *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.
- [127] T. Loran, A. Barros Cardoso da Silva, S. K. Joshi, S. V. Baumgartner, and G. Krieger, “Ship detection based on faster R-CNN using range-compressed airborne radar data,” *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, Dec. 2023.
- [128] Z. Cai, X. Bai, and J. Zhao, “Range-doppler processing for passive radar based on joint tracking and dynamic compressed sensing,” in *Int. Conf. Signal, Inf. Data Process.*, Zhuhai, China, Nov. 2024, pp. 1–6.
- [129] D. Xu *et al.*, “Micro-Doppler frequency extraction and scatterer classification for a smooth-surfaced cone-shaped precession target under narrowband radar,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 9368–9379, Mar. 2025.
- [130] J. Farhat, G. Pasolini, E. Paolini, M. Asad Ullah, and R. Demo Souza, “Doppler estimation and compensation techniques in LoRa direct-to-satellite communications,” *IEEE Open J. Commun. Soc.*, vol. 6, pp. 6759–6776, Aug. 2025.
- [131] *Zero-touch network and Service Management (ZSM); Network Digital Twin*, European Telecommunications Standards Institute ETSI™ GR ZSM 015 V1.1.1, Feb. 2024.
- [132] D. González-Sánchez, L. Bellido, I. D. Martínez-Casanueva, D. Martínez-García, D. Fernández, and D. R. Lopez, “Toward building a digital twin for network operations and management,” *IEEE Open J. Commun. Soc.*, vol. 6, pp. 2583–2598, Mar. 2025.
- [133] R. Poorzare, D. N. Kanellopoulos, V. K. Sharma, P. Dalapati, and O. P. Waldhorst, “Network digital twin toward networking, telecommunications, and traffic engineering: A survey,” *IEEE Access*, vol. 13, pp. 16 489–16 538, Jan. 2025.
- [134] S. Faye *et al.*, “A functional framework for network digital twins,” in *Proc. Joint European Conf. on Networks and Communications & 6G Summit*, Poznan, Poland, Jun. 2025, pp. 327–332.

- [135] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise.” AAAI Press, Aug. 1996, p. 226231.
- [136] J. LaViola, “A comparison of unscented and extended Kalman filtering for estimating quaternion motion,” in *Proc. American Control Conf.*, Denver, CO, USA, Jun. 2003, pp. 2435–2440.
- [137] K. G. Murty, “An algorithm for ranking all the assignments in order of increasing cost,” *Oper. Res.*, vol. 16, no. 3, pp. 682–687, Jun. 1968.
- [138] S. H. Rezatofighi, A. Milan, Z. Zhang, Q. Shi, A. Dick, and I. Reid, “Joint probabilistic data association revisited,” in *IEEE Int. Conf. Comput. Vision*, 2015, pp. 3047–3055.
- [139] A. Conti *et al.*, “Location awareness in beyond 5G networks,” *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 22–27, Nov. 2021, special issue on *Location Awareness for 5G and Beyond*.
- [140] A. Conti, G. Torsoli, C. A. Gómez-Vega, A. Vaccari, G. Mazzini, and M. Z. Win, “3GPP-compliant datasets for xG location-aware networks,” *IEEE Open J. Veh. Technol.*, vol. 5, pp. 473–484, Apr. 2024.
- [141] A. Conti, G. Torsoli, C. A. Gómez-Vega, A. Vaccari, and M. Z. Win, “xG-Loc: 3GPP-compliant datasets for xG location-aware networks,” *IEEE Dataport*, Dec. 2023. [Online]. Available: <https://dx.doi.org/10.21227/rper-vc03>
- [142] G. Torsoli, M. Z. Win, and A. Conti, “Blockage intelligence in complex environments for beyond 5G localization,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1688–1701, Jun. 2023, special issue on *3GPP Technologies: 5G-Advanced and Beyond*.
- [143] G. Torsoli, M. Z. Win, and A. Conti, “Beyond 5G localization via sidelinks in industrial IoT scenarios,” in *Proc. IEEE Int. Conf. Commun.*, Denver, USA, Jun. 2024, pp. 920–925.
- [144] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, “Generalized optimal sub-pattern assignment metric,” in *Int. Conf. Inf. Fusion*, Xi’an, China, Jul. 2017, pp. 1–8.