Università degli Studi di Modena e Reggio Emilia

Dipartimento di Ingegneria "Enzo Ferrari"

International Doctorate in
Information and Communication Technologies

XXXV Cycle

FMCW Radars
For Automotive Applications

Supervisor:
Prof. Giorgio Matteo Vitetta

Candidate:
Giorgio Guerzoni

PhD School Coordinator:
Prof. Sonia Bergamaschi

Abstract

While radar technology has been around for decades, the size and cost of radar devices has limited its use to niche applications. In recent years, however, *radio frequency integrated circuits* (RFIC) working at 24 or 77 GHz have become available to the consumer market, empowering engineers to build smaller and cheaper radars. At the same time, a deep interest in autonomous driving vehicles has arisen, pushing *integrated circuit* (IC) manufacturers to release a vast amount of *multiple-input multiple-output* (MIMO) radar frontends specifically designed for the automotive industry. The aim of this thesis is to explore the possibilities offered by such devices, as well as those brought by the availability of wider antenna arrays, and to design novel techniques for analyzing the data that these devices produce. First, a novel deterministic algorithm for frequency and complex amplitude estimation is presented and compared against state-of-the-art algorithms; its use in target range, azimuth and elevation estimation is illustrated. Then, *machine learning* (ML) techniques are applied to radar data in order to perform both human activity recognition and target tracking. Next, human heart rate and breath rate monitoring techniques exploiting MIMO radars are investigated. All the numerical results provided in this thesis are based on both synthetically generated data and various measurements acquired through commercial radars in different scenarios.

Contents

List of acronyms

AAF Anti Aliasing Filter.
AD Arctangent Demodulation.
ADC Analog to Digital Converter.
AE Auto-Encoder.
AI Artificial Intelligence.
AM Alternating Minimization.
ANC Adaptive Noise Cancellation.
API Application Programming Interface.
ASIC Applied Specific Integrated Circuit.
AWGN Additive White Gaussian Noise.

Bi Bipolar.
BN Batch Normalization.
BPF Band Pass Filtering.
BPTT Back-Propagation Through Time.
BR Breath rate.

CAE Convolutional Auto-Encoder.
CFAR Constant False Alarm Rate.
CMOS Complementary Metal Oxide Semiconductor.
CNN Convolutional Neural Network.
CPU Central Processing Unit.
CS Compressed Sensing.
CSD Complex Signal Demodulation.
CV Correlation of Variation.
CVD Cadence Velocity Diagram.
CW Continuous Wave.
CWT Continuous Wavelet Transform.

DACM Differentiate And Cross Multiplier.
DBF Digital Beamforming.
DBSCAN Density Based Spatial Clustering of Applications with Noise.
DC Direct Current.
DCNN Deep Convolutional Neural Network.
DCT Discrete Cosine Transform.
DFT Discrete Fourier Transform.
DL Deep Learning.
DoA Direction of Arrival.
DSP Digital Signal Processor.
DT Decision Tree.

ECG Electrocardiogram.
EMD Empirical Mode Decomposition.
ERM Empirical Risk Minimization.
ESPRIT Estimation Signal Parameters Rotational Invariance Technique.

FC Fully Connected.
FCN Fully Convolutional Network.
FDM Frequency Division Multiplexing.
FFT Fast Fourier Transform.
FM Frequency Modulation.
FMCW Frequency-Modulated Continuous Wave.
FoV Field of View.
FPGA Field Programmable Gate Array.
FPS Frames Per Second.

GAN Generative Adversarial Network.
GPU Graphic Processing Unit.
GUI Graphical User Interface.

HCI Human Computer Interface.
HGR Human Gesture Recognition.
HMM Hidden Markov Model.
HR Heart Rate.
HRV Heart Rate Variability.

IAA Iterative Adaptive Approach.

ICs integrated circuits.
IDFT Inverse DFT.
IF Intermediate Frequency.
IFFT Inverse FFT.
IMF Intrinsic Mode Function.
IOU Intersection Over Union.
IR-UWB Impulse Radio Ultra-Wideband.
ISM Industrial Scientific and Medical.
IWR Industrial mmWave Radar.

KNN K - Nearest Neighbour.

LB Leaning Based.
LDA Linear Discriminant Analysis.
LNA Low Noise Amplifier.
LO Local Oscillator.
LP lowpass.
LPC Linear Predictive Coding.
LRR Long Range Radar.
LS Least Square.
LSTM Long Short Term Memory.

MAE Mean Absolute Error.
MEMS Micro Electro-Mechanical System.
MF Matched Filter.
MIMO Multiple Input Multiple Output.
ML Machine Learning.
MLP Multi-Layer Perceptron.
MMIC Monolithic Microwave Integrated Circuit.
MRR Medium Range Radar.
MSE mean square error.
MUSIC Multiple Signal Classification.

NB Naive Bayes.
NN Neural Network.
NN Normal to Normal.

OFDM Orthogonal Frequency Division Multiplexing.

P2G Position To Go.
PA Power Amplifier.
PAE Peak Absolute Error.
PCA Principal Component Analysis.
PCFEA Parallel Cancellation & Fine Estimation Algorithm.
PCR Pulsed Coherence Radar.
PGA Programmable Gain Amplifier.
PLL Phase Locked Loop.
PMCW Phase Modulated Continuous Wave.
PRI Pulse Repetition Interval.

R Region.
RADAR Radio Detection And Ranging.
RASCA Range & Angle Serial Cancellation Algorithm.
RBM Random Body Movement.
RCS Radar Cross Section.
RF Radio Frequency.
RHS Right Hand Side.
RMSE Root Mean Square Error.
RMSSD Root Mean Square Successive Difference.
RNN Recursive Neural Network.
ROI Region of Interest.
RPE Range Profile Estimation.
RX Receive.

SAMME Stagewise Additive Modeling Multi-class Exponential.
SCCEA Serial Cancellation & Coarse Estimation Algorithm.
SDNN Standard Deviation of Normal to Normal.
SFCW Stepped Frequency Continuous Wave.
SFE Single Frequency Estimator.
SGD Stochastic Gradient Descent.
SiGe Silicon-Germanium.
SISO Single Input Single Output.
SNR Signal-to-Noise Ratio.
SPE SPatial Estimator.
SRR Short Range Radar.
STDAEC Single Target Detection, Angular Estimation and Cancellation.
STDREC Single Target Detection, Range Estimation and Cancellation.
STFT Short Time Fourier Transform.
SVM Support Vector Machine.

TDM Time Division Multiplexing.
TI Texas Instrument.
TOF Time Of Flight.
TRI Triangular Index.
TRP Target Range Profile.
TX Transmit.

ULA Uniform Linear Array.
URA Uniform Rectangular Array.
UWB Ultra-wideband.

VA Virtual Antenna.
VCO Voltage Controlled Oscillator.
VGA Voltage Gain Amplifier.
VHSIC Very High Speed Integrated Circuit.

XAI eXplainable Artificial Intelligence.

YOLO You Only Look Once.

1. Introduction

This chapter provides an introduction to the world of colocated MIMO radar systems. After illustrating some basic information about their characteristics and outlining their evolution in the last two decades, the architecture of a colocated MIMO radar system is described. Finally, the received signal model is briefly analysed, and essential information about various detection and estimation algorithms that can be employed in colocated radar systems is provided.

In the second chapter of this thesis a collection of novel algorithms for range and angle estimation of multiple targets is presented, together with some simulated and experimental results. Their application to the automotive industry is obvious: the detection of extended objects that leads to the generation of pointclouds; moreover, in this scenario, the detection of weak targets that may be shadowed by objects with a greater *radar cross section* is of paramount importance in autonomous driving.

The third chapter is devoted to describing the application of machine learning and deep learning techniques to radar systems: in particular, in an autonomous driving perspective, it is once again important to detect whether or not a target is a static object, a pedestrian, a runner or a cyclist in order to allow a vehicle to react accordingly.

In the fourth and last chapter, radar application to vital sign monitoring, i.e. to heart rate and breath rate estimation, is explored. The most important use case for this technology is in the hospitals, but a certain interest in performing in-cabin sensing to assess the health status of car occupants also exists; radars are particularly suited for this task because they do not require the application of sensors to the body of their users.

1.1 Basic principles and classification

The initial excitement about the use of antenna arrays at both transmit and receive sides (i.e., briefly, about MIMO) in wireless systems has been sparked by the pioneering work of J. H. Winters [1], G. J. Foschini [2], Foschini and M. J. Gans [3], and E. Telatar [4]; these researchers predicted huge capacity gains in *wireless communications* affected by multipath fading [5]. A few years later, the exploitation of antenna arrays has been also investigated in the radar field for the potential improvements it could provide in terms of *signal-to-noise ratio* (SNR), resolution and detection capability. In fact, in principle, the availability of multiple transmit/receive antennas allows to (e.g., see [6–8])

1) increase the SNR characterizing target echoes and make it more stable;

2) implement *spatial filtering* (i.e., *beamforming*) for directional signal transmission/reception and, consequently, achieve a large *field of view* (FOV);

3) increase the overall number of degrees of freedom and, consequently, the maximum number of targets that can be detected at a given range;

4) improve the *angular resolution* with respect to traditional radar systems;

5) exploit *spatial diversity*, so that uncorrelated aspects of a given target can be perceived.

Generally speaking, MIMO radar systems can be divided in *statistical* MIMO radars [9, 10] and *colocated* MIMO radars [11], on the basis of the distance between their transmit and receive arrays [8]. In fact, the transmit and receive antennas of the radar systems belonging to the first class are widely separated; on the contrary, in radar systems of the second class, transmit antennas are close to the receive ones and, in particular, are usually placed on the same shield. Colocated MIMO radars can be further classified as: a) *mono-static radars*, where transmit and receive arrays share their antenna elements; b) *pseudo-bistatic* radars, where transmit and receive arrays are made of distinct antenna elements, placed at different positions. It is important to keep in mind that, in statistical MIMO radars, spatial diversity originates from the fact that distinct receive antennas, being well separated, can observe uncorrelated parts of the same target. In colocated MIMO radars, instead, a large spatial aperture is achieved by radiating *orthogonal* waveforms. Based on the way these waveforms are generated, colocated MIMO radars can be divided in: a) *time division multiplexing* (TDM) radars [12], b) *frequency division multiplexing* (FDM) radars [13] and c) *orthogonal frequency division multiplexing* (OFDM) radars [14]. On the one hand, in TDM (FDM) radars, orthogonality is achieved by transmitting through distinct antennas over disjoint time (frequency) intervals; on the other hand, in OFDM radars, any transmit antenna can be used to radiate multiple orthogonal waveforms at the same time. A further classification of colocated MIMO radars, commonly adopted in the automotive field, is based on the maximum measurable range. According to this classification, these systems are divided in (see Table 1.1, where, for each type of radar, the achievable range, the transmission frequency and the typical applications are listed):

1. *Short range radars* (SRRs) - These are able to measure a maximum range of about 30 m and offer the highest angular resolution.

2. *Medium range radars* (MRRs) - These are characterized by a maximum range of about 100 m, offer a quite large azimuthal FOV and achieve a reasonable angular resolution.

3. *Long range radars* (LRRs) - These are characterized by the largest maximum range (250 m) and the thinnest FOV.

In the last paragraph of this section, the architecture of a pseudo-bistatic colocated MIMO radar operating in TDM mode is described in some detail. Our interest in this specific architecture is motivated by its wide use in various civilian applications, and by its capability of detecting multiple targets and accurately estimating their position.

Radar type	Max range (m)	Freq. (GHz)	Typical applications
Short range	30	5-77	Park assist, pre-crash
Mid range	100	24-77	Blind spot detection Rear collision avoidance Cross traffic alert
Long range	250	40-77	Adaptive cruise control

Table 1.1: *Classification of colocated MIMO radars on the basis of their maximum measurable distance.*

1.2 A brief history of the colocated MIMO radar technology

The birth of *radio detection and ranging* (briefly, radar) systems dates back to 1904, when the German inventor *Christian Hulsmayer* built a simple ship detection device for avoiding collisions in fog [15]. However, the first practical radar system was developed by the British physicist Sir *Robert Watson-Watt* in 1935, and was employed by the British army in World War II to detect air and sea aggressors [16]. Another fundamental step in the evolution of radar technology is represented by the early studies on *optimal filtering*; the rigorous formulation of this problem and its solution are due to the American scholar *Norbert Wiener* and date back to the 40' [17]. Since then, many advancements have been made in military and civilian radar systems, thanks to the development of signal processing techniques and to the evolution in electronic technology. The most significant advances in signal processing methods applicable to radar systems equipped with antenna arrays have involved both the transmit side and the receive side, and can be summarised as follows.

As far as the transmit side is concerned, substantial research efforts have been devoted to the study of *analog beamforming* (ABF) and *digital beamforming* (DBF) methods for controlling *phased arrays*; both types of methods allow to obtain electronic *beam steering*, i.e. to *steer* the main lobe of the array radiation pattern without any movement of the antennas forming it. It is worth stressing that phased arrays have been around for more than fifty years [18], and that a radar equipped with a phased array is much simpler than a MIMO radar. In fact, a radar system endowed with a phased array generates a single waveform feeding each transmit antenna with a different phase (or, equivalently, with a different delay); consequently, the waveforms radiated by distinct antennas are highly correlated. Moreover, analog beamforming represents the earliest method for electronic beam steering; in this case, each of the signal feeding a transmit antenna is first amplified and then delayed through a phase shifter in a *radio frequency* (RF) stage; an important drawback of this method is represented by the fact that the shape of the resulting beam is fixed. On the other hand, DBF is based on the idea of implementing beam steering in the (digital) baseband portion of the radar hardware by multiplying each signal by a complex gain [19]. This procedure allows to digitally customize the radiated beam, adapting its direction to channel conditions.

This technique, also known as *adaptive beamforming* [20], plays an important role in the presence of severe path loss. However, it should be always kept in mind that any radar transmitter exploiting beamforming requires some time (in practice, multiple dwells) to scan the area of interest. On the contrary, if a MIMO radar is employed, the entire observed area is illuminated in a single dwell and beamforming is obtained through the use of different orthogonal waveforms [21].

Another important research area concerning the transmit side of radar systems equipped with antenna arrays concerns the design of the radiated waveforms [8]. Despite the fact that significant theoretical results have been achieved in the field of optimal design of waveforms (e.g., see [22]), few modulation techniques have been employed in commercial MIMO radars until now. These include the *frequency modulated continuous wave* (FMCW) technique [23] (also known as *chirp signal modulation*) and the *stepped frequency continuous wave* (SFCW) technique [24]. In the last years, considerable attention has been also paid to the use of the OFDM technique [25, 26] and to the *phase modulated continuous wave* (PMCW) technique [27].

Early research work regarding the receive side of radar systems endowed with antenna arrays has focused on the development of beamforming methods [28]. One of the most important contributions to this area is represented by the so called *Capon beamformer*, which can provide good resolution and interference rejection capability [29, 30]. Other fundamental contributions about the processing of multiple signals acquired by a radar systems through its antenna array concern the estimation of the *direction of arrival* (DOA) of the electromagnetic waves impinging on the array itself. Here, we limit ourselves to mention the MUSIC [31] and the *estimation of signal parameters via rotational invariance* (ESPRIT) techniques [32, 33].

The development of signal processing methods for MIMO radars started after the end of 2003; in fact, in that period, the concept of MIMO radar, defined as a device able to probe a wireless channel by transmitting multiple signals and receiving their echoes with similar multiplicity, was proposed for the first time [7]. Since the beginning, it was clear that MIMO technology could have represented an important tool to improve the SNR of received signals and to increase radar aperture [6, 8, 21, 34]. Since then, the exploitation of known DOA estimation strategies, developed in the previous years for antenna arrays (like MUSIC and ESPRIT), has been widely investigated for this new type of radars [35–37]. However, the availability of MIMO radars able to radiate wideband signals by a large number of antennas and to acquire their echoes through an even larger number of antennas have raised various problems, whose solution requires substantial research efforts. In fact, on the one hand, these devices allow to acquire a rich set of information about the surrounding propagation environment; on the other hand, they require storing and processing large datasets. This has motivated the investigation of *compressed sensing* (CS) and statistical sparsity-based techniques, since these can be exploited to perform signal detection and parameter estimation on the basis of a much smaller dataset than that available in the case in which the received waveforms undergo Nyquist sampling [38, 39]; various examples of CS-based estimation algorithms can be found in ref. [40].

As far as the advancement in electronic technology is concerned, in the remaining part of this paragraph we focus on some important results achieved in the development of compact integrated radar devices employed in the automotive field, since this

is one of the first commercial markets in which MIMO radars have been playing a fundamental role. The first generation of commercial *ultra-wideband* (UWB) automotive radar sensors operating in the 77 GHz band has become available in 1999. These devices were not endowed with antenna arrays and their implementation was based on discrete electronic components (in particular, gallium-arsenide Gunn diodes mounted inside a waveguide cavity were employed in the generation of RF waveforms). However, electronic technology progressed quickly in this field and, after few years, MMICs employing high-performance *silicon-germanium* (SiGe) transistors became available for the implementation of fully integrated radars. Pioneering work in the development and manufacturing of such a technology has been accomplished by the *Infineon* company, that has started its production in 2004 [41]. It is also worth mentioning that, in the same year, a description of the first fully integrated 24-GHz eight-element phased array receiver in SiGe and of the first fully integrated 24-GHz four-element phased array transmitter with integrated power amplifiers in *complementary metal-oxide semiconductor* (CMOS) has appeared [42]; these devices were able to accomplish beamforming and could be used for communication, ranging, positioning, and sensing applications. Other examples of phased arrays operating in X and Ku-band have been described later in ref. [43]. The first FMCW MIMO radar transceiver operating at 77 GHz has been implemented in SiGe technology in 2008 [44], whereas the production of the first MIMO FMCW radar, operating according to a TDM strategy and equipped with an array of colocated antennas, started in 2009 [45,46]. As far as we know, the last device represents the first compact MIMO radar system based on a MMIC in SiGe, operating at 77 GHz and radiating ultra-wideband signals. In this system, wide-band and high-frequency patch antennas are built on a RF substrate [47], while the base-band MIMO signal processing is accomplished off-chip by a *field programmable gate array* (FPGA) board. Moreover, the *analog-to-digital converters* (ADCs) at the receive side are implemented in CMOS technology and embedded in the transceiver chip; this has been made possible by the SiGe Bi-CMOS process, which has allowed to integrate multiple functions on a single chip and at low cost. In the last decade, radar designers working on the development of new integrated radar devices have investigated the use of the more scalable CMOS RF technology [48]. An important trend in the technological evolution of MIMO radar systems is also represented by the attempt of exploiting the same hardware for both radar and communications [49]. Some milestones achieved in the evolution of the signal processing methods and of the technology employed in colocated radar systems during the last two decades are summarized in Fig. 1.1.

1.3 Architecture of a colocated TDM MIMO radar

In the remaining part of this chapter, we always refer to a colocated and bistatic MIMO radar system; its architecture is illustrated in Fig. 1.2. Moreover, we first assume that: a) the considered radar system is equipped with a *two-dimensional* (2D) array, consisting of N_T *transmit* (TX) and N_R *receive* (RX) antennas; b) it employs a TDM strategy; c) it exploits all the available transmit diversity (i.e., all the available TX antennas). Consequently, if a time slot of T_0 s is assigned to each TX antenna, transmission from all the TX antennas is accomplished over an interval lasting $T_F \triangleq N_T T_0$ s; this interval represents the duration of a single *transmission frame*.

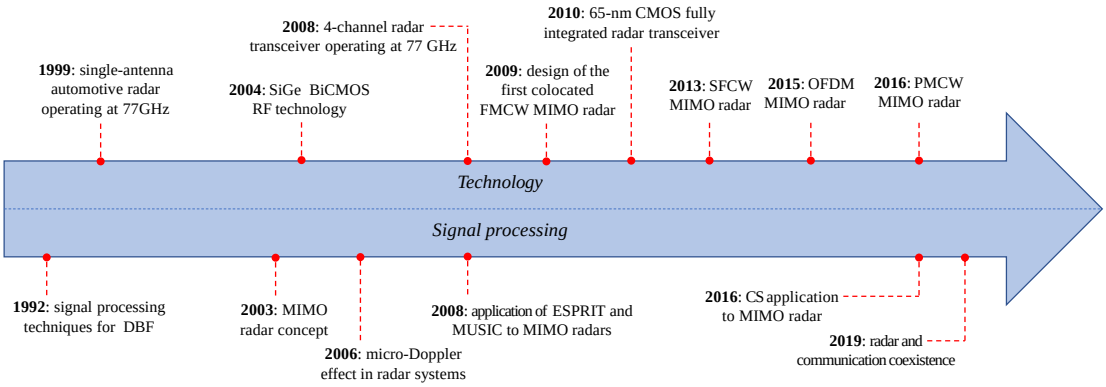


Figure 1.1: *Milestones in the evolution of colocated radars.*

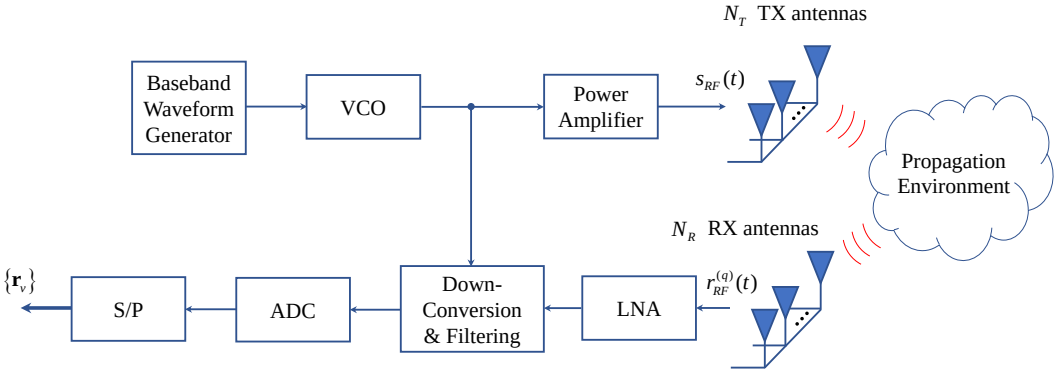


Figure 1.2: *MIMO radar transmitter (upper part) and receiver (lower part).*

In the following, we consider the *voltage controlled oscillator* (VCO) of the radar transmitter to be fed by a *periodic ramp generator*; this produces a chirp FM signal, whose instantaneous frequency evolves periodically, as illustrated in Fig. 1.3; it is then radiated by its transmit array after power amplification.

In this figure, the parameters T, T_R and T_0 represent the *chirp interval*, the *reset time* and the *pulse period* (or *pulse repetition interval*), respectively [50], whereas the parameters f_0 and B are the *start frequency* and the *bandwidth*, respectively, of the transmitted signal. For this reason, if we focus on the time interval $(0, T)$ and assume that, in that interval, the p-th TX antenna is employed by the considered radar system (with $p \in \{0\ ,\ 1,\ ...,\ N_T - 1\}$), the radiated signal can be expressed as

$$s_{RF}(t) = A_{RF} \Re\{s(t)\}, \tag{1.1}$$

where A_{RF} is its amplitude,

$$s(t) \triangleq \exp[j\theta(t)], \tag{1.2}$$

$$\theta(t) \triangleq 2\pi \left(f_0 t + \frac{\mu}{2} t^2\right) \tag{1.3}$$

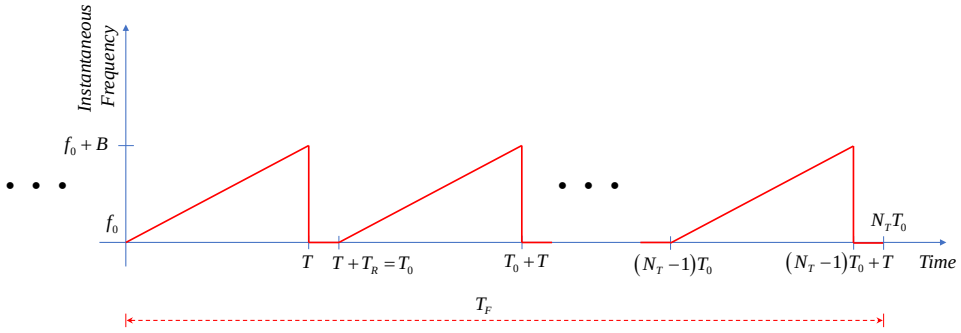


Figure 1.3: *Representation of the instantaneous frequency of the RF signal generated by the VCO in a FMCW radar system.*

and

$$\mu = \frac{B}{T} \tag{1.4}$$

is the *chirp rate*, i.e. the steepness of the generated frequency chirp.

Let $r_{RF}^{(q)}(t)$ denote the signal available at the output of the q-th receive antenna, with $q = 0, 1, ..., N_R - 1$ (see Fig. 1.2); this signal feeds a *low noise amplifier* (LNA), whose output undergoes downconversion, filtering and analog-to-digital conversion at a frequency $f_s = 1/T_s$, where T_s denotes the sampling period of the employed ADC. This will lead us to also state that

$$T = NT_s \tag{1.5}$$

If we assume that the radiated signal $s_{RF}(t)$ (1.1) is reflected by L *static point targets*, the useful component of $r_{RF}^{(q)}(t)$ consists of the superposition of L echoes, each originating from a distinct target. In this case, if the propagation environment undergoes slow variations, a simple mathematical model can be developed to represent the sequence of samples generated by the ADC in a single chirp interval. In deriving this model, the couple of the involved *physical* TX and RX antennas (namely, the p-th TX antenna and the q-th RX antenna) of the considered *bistatic radar* is often replaced by a single *virtual antenna* of an equivalent *monostatic radar*. In particular, the abscissa x_v and the ordinate y_v of the v-th *virtual antenna element* associated with the p-th TX antenna and the q-th RX antenna are computed as[1]

$$x_v = \frac{x_p + x_q}{2} \tag{1.6}$$

and

$$y_v = \frac{y_p + y_q}{2}, \tag{1.7}$$

[1]This is not the only rule adopted in the technical literature to compute the coordinates of the v-th virtual antenna element. For instance, in ref. [21, Par. 4.3.1, pp. 159-161], the abscissa (ordinate) of this element is evaluated as $2x_v$ ($2y_v$), where x_v and y_v are expressed by eqs. (1.6) and (1.7), respectively. Keep in mind, however, that, if the last rule is adopted, all the following formulas involving such coordinates must be changed accordingly.

respectively, with $v = 0, 1, ..., N_V - 1$; here, (x_p, y_p) ((x_q, y_q)) are the coordinates[2] of the TX (RX antenna) and

$$N_V \triangleq N_T \cdot N_R \tag{1.8}$$

represents the overall size of the resulting *virtual array*. Based on these assumptions, the n-th received signal sample acquired through the v-th virtual antenna element (with $v = 0, 1, ..., N_V - 1$) can be expressed (e.g., see [51, Par. 4.6, eq. (4.27)])

$$\begin{aligned} x_{r,n}^{(v)} &= \sum_{l=0}^{L-1} a_l \cos\left(2\pi n\, F_l^{(v)} + \psi_l^{(v)}\right) + w_{r,n}^{(v)} \\ &= \sum_{l=0}^{L-1} \left[C_l^{(v)} \exp\left(j2\pi n\, F_l^{(v)}\right) + \right. \\ &\quad \left. + (C_l^{(v)})^* \exp\left(-j2\pi n\, F_l^{(v)}\right)\right] + w_{r,n}^{(v)}, \end{aligned} \tag{1.9}$$

where

$$C_l^{(v)} \triangleq \frac{1}{2} a_l \exp\left(j\psi_l^{(v)}\right) \tag{1.10}$$

represents the *complex amplitude* of the real tone appearing in the *right hand side* (RHS) of (1.9), or, if the radar system is able to receive both *in phase* and *quadrature* components of the signal,

$$x_{c,n}^{(v)} = \sum_{l=0}^{L-1} A_l^{(v)} \exp\left(j2\pi n\, F_l^{(v)}\right) + w_{c,n}^{(v)}, \tag{1.11}$$

must be adopted in place of its real counterpart (1.9) for any n; here,

$$A_l^{(v)} \triangleq a_l \exp\left(j\psi_l^{(v)}\right) \tag{1.12}$$

for any v and l with $n = 0, 1, ..., N - 1$; here, N is the overall number of samples acquired over a chirp period, a_l is the amplitude of the l-th component of the useful signal (this amplitude depends on both the range R_l and the reflectivity of the l-th target, but is assumed to be independent of v for simplicity),

$$F_l^{(v)} \triangleq f_l^{(v)} T_s \tag{1.13}$$

is the normalized version of the frequency

$$f_l^{(v)} \triangleq \mu\, \tau_l^{(v)}, \tag{1.14}$$

characterizing the l-th target detected on the v-th virtual receive antenna,

$$\tau_l^{(v)} = \frac{2}{c} \left[R_l + x_v \cos(\theta_l) \sin(\phi_l) + y_v \sin(\theta_l)\right] \tag{1.15}$$

is the delay of the echo generated by the l-th target and observed on the v-th virtual channel, R_l, ϕ_l and θ_l denote the range of the l-th target, its azimuth and its elevation, respectively,

$$\psi_l^{(v)} \cong 2\pi f_0 \tau_l^{(v)}, \tag{1.16}$$

[2] A reference system lying on the physical antenna array is assumed.

and $w_{z,n}^{(v)}$ is the n-th sample of the *additive white Gaussian noise* (AWGN) sequence affecting the received signal (the noise variance is denoted σ_w^2 in the following and is assumed to be independent of v)[3]. The samples $\{x_n^{(v)};\ n = 0,\ 1,\ ...,\ N-1\}$ can be collected in the N-dimensional vector

$$\mathbf{x}_z^{(v)} \triangleq \left[x_{z,0}^{(v)}, x_{z,1}^{(v)}, ..., x_{z,N-1}^{(v)}\right]^T, \tag{1.17}$$

which is processed by the next stages of the radar receiver for target detection and estimation. As it can be easily inferred from eq. (1.9), in a FMCW radar system, the problem of target detection and range estimation on the v-th virtual channel is equivalent to the classic problem of estimating the *frequencies* of multiple overlapped sinusoids (multiple overlapped complex exponentials) in the presence of AWGN [52]. In fact, if an estimate $\hat{f}_l^{(v)}$ of the frequency $f_l^{(v)}$ (1.14) and an estimate $\hat{F}_l^{(v)}$ of the normalised frequency $F_l^{(v)}$ (1.13) are available for the v-th virtual channel, an estimate of the range R_l can be computed as (see eqs. (1.14) and (1.15))

$$\hat{R}_l^{(v)} = \frac{1}{2}\frac{\hat{f}_l^{(v)}}{\mu}c \tag{1.18}$$

for any v and l. Information about the angular coordinates (namely, the azimuth and the elevation) of the l-th target, instead, can be acquired through the estimation of the set of N_V phases $\{\psi_{v,l};\ v = 0,\ 1,\ ...,\ N_V - 1\}$ observed over the available virtual antennas. In fact, since (see eqs. (1.15) and (1.16))

$$\psi_l^{(v)} \cong 4\pi\frac{f_0}{c}\left[R_l + x_v\cos(\theta_l)\sin(\phi_l) + y_v\sin(\theta_l)\right] \tag{1.19}$$

where

$$\lambda \triangleq \frac{c}{f_0} \tag{1.20}$$

is the wavelength associated with the frequency f_0, the sequence $\{\psi_l^{(v)};\ v = 0,\ 1,\ ...,\ N_V - 1\}$ exhibits a periodic behavior characterized by the *normalised horizontal spatial frequency*

$$F_{H,l} \triangleq 2\frac{d_H}{\lambda}\cos(\theta_l)\sin(\phi_l), \tag{1.21}$$

if the considered virtual elements form an horizontal *uniform linear array* (ULA), whose adjacent elements are spaced d_H m apart. Dually, if a vertical ULA is assumed, the periodic variations observed in the same sequence of phases are characterized by the*normalised vertical spatial frequency*

$$F_{V,l} \triangleq 2\frac{d_V}{\lambda}\sin(\theta_l), \tag{1.22}$$

where d_V denotes the distance between adjacent elements of the vertical virtual array. Consequently, angle finding can be easily accomplished by DBF, i.e. by performing FFT processing on the estimated phases taken across multiple elements of the virtual array

[3] In the following, when the letter z will be used in a subscript, it will be implicitly assumed, unless differently stated, that it can be equal to r or c.

in a single frame interval [53, 54]. Note, however, that other angle estimation methods, achieving a better resolution than FFT processing are also available; here, we limit to mention the so called subspace-based methods (such as MUSIC and ESPRIT), sparse sensing-based methods [55, 56] and the *iterative adaptive approach* (IAA) developed in ref. [57]. Subspace-based methods require computing an accurate estimate of the array covariance matrix; consequently, the measurements acquired over multiple snapshots must be processed. Moreover, they do not allow to estimate the amplitude of the echo associated with each detected target and require prior knowledge of the size of the useful signal subspace (i.e., of the number of detectable targets). On the contrary, sparse sensing-based methods and IAA can generate angle estimates on the basis of a single snapshot of the received signal; however, this result is obtained at the price of a significant computational effort.

The received signal model (1.9) holds if all the observed targets are static. Let us focus now on a FMCW radar system operating in the presence of L moving point targets and having the following characteristics: a) it is equipped with a single TX antenna and a single RX antenna (i.e., $N_T = N_R = 1$); b) its reset time T_R is equal to 0, so that $T_0 = T$ (see Fig. 1.3); c) its transmission frame consists of N_c chirps, so that the duration T_F of the transmission frame is equal to $N_c T_0 = N_c T$ s; d) N distinct ADC samples are acquired in each chirp interval at the receive side. Then, it is not difficult to prove that, if the ranges of all the targets are much larger than their displacements observed during the considered transmission frame, the n-th sample of the signal acquired in the k-th chirp interval (with $k = 0, 1, ..., N_c - 1$) can be expressed in a similar way as eq. (1.9), namely as (e.g., see [50, eq.(5)])

$$r_n^{(k)} \cong \sum_{l=0}^{L-1} a_l \cos(2\pi n \ (F_l + F_{D,l}) \ + \psi_l^{(k)}) + w_n^{(k)}, \tag{1.23}$$

where $F_l = \mu \, \tau_l \, T_s$ (see eqs. (1.13) and (1.14)), $\tau_l = 2R_l/c$ is the delay of the echo generated by the l-th target and observed in the first chirp interval (in this interval, the target range is assumed to be equal to R_l),

$$F_{D,l} = \frac{2 \, v_l}{\lambda} T_s \tag{1.24}$$

is the normalised Doppler frequency, v_l is the radial velocity[4] of the l-th target,

$$\psi_l^{(k)} \cong \frac{4\pi}{\lambda} R_l^{(k)}, \tag{1.25}$$

$$R_l^{(k)} = R_l + v_l \, kT \tag{1.26}$$

is the target range observed in the k-th chirp interval and $w_n^{(k)}$ is the n-th sample of the AWGN sequence affecting the received signal in the same chirp interval.

The rate of change observed in the sequence of phases $\{\psi_l^{(k)}; \, k = 0, 1, ..., N_c - 1\}$ is proportional to v_l, since (see eqs. (1.25) and (1.26))

$$\psi_l^{(k+1)} - \psi_l^{(k)} \cong \frac{4\pi}{\lambda} \left(R_l^{(k+1)} - R_l^{(k)} \right) = 4\pi \frac{T}{\lambda} v_l \tag{1.27}$$

[4]This speed is positive (negative) if the target is approaching (is moving away) from the radar.

with $k = 0, 1, ..., N_c - 1$. Therefore, target velocity can be easily assessed by means of FFT processing after computing an estimate of the above mentioned phases.

In the technical literature, range and speed information of the moving targets detected in a given propagation environment are usually condensed in a 2D plot, called *range-Doppler map* [50, 58]. In a FMCW radar system equipped with a single TX antenna and a single RX antenna, this map is generated as follows. Let $\mathbf{r}^{(k)}$ denote the N-dimensional (column) vector consisting of the real measurements acquired in the k-th chirp of a transmission frame, with $k = 0, 1, ..., N_c - 1$, where N_c is the overall number of chirps forming the frame itself. The N_c vectors $\{\mathbf{r}^{(k)}; k = 0, 1, ..., N_c - 1\}$ are collected in the matrix

$$\mathbf{R} = [\mathbf{r}^{(0)}\, \mathbf{r}^{(1)}\, \cdots\, \mathbf{r}^{(N_c-1)}]\,, \tag{1.28}$$

having size $N \times N_c$. This matrix undergoes zero-padding, that turns it into a matrix $\mathbf{R}_{\text{ZP}}$ of size $N_0 \times N_0^{'}$. The last matrix feeds a $N_0 \times N_0^{'}$-th order FFT, that generates the *range-Doppler* (complex) *matrix*

$$\mathbf{D} = [d_{p,q}] \triangleq \text{FFT}_{N_0 \times N_0^{'}}\,[\mathbf{R}]\,, \tag{1.29}$$

where $\text{FFT}_{X\times Y}[\cdot]$ denotes 2D FFT operator of size $X \times Y$; note that the index p (q) labelling the elements of the matrix $\mathbf{D}$ refers to the range (Doppler) domain. Representing, on a Cartesian plane, the absolute value of the elements of the matrix $\mathbf{D}$ yields the above mentioned range-Doppler map.

In the last fifteen years, substantial attention has been also paid to the problem of estimating the *micro-movements* of detected targets; such movements usually originate from mechanical vibrations or rotations (overlapping to a bulk translation) and may generate a frequency modulation in the received signal; the last phenomenon is known as *micro-Doppler*. The recent interest in micro-Doppler is motivated by the fact that it can be exploited to establish the dynamic properties of targets [59] and, consequently, can be used to *classify* them or identify specific properties related to their motion. In a FMCW radar system equipped with a single TX antenna and a single RX antenna, the micro-Doppler phenomenon can be analysed as follows. Let us assume that N_f consecutive frames (each consisting of N_c chirps) are transmitted by the considered radar system and that the range-Doppler matrix $\mathbf{D}$ (1.29) is evaluated for each frame (the matrix referring to the m-th frame is denoted $\mathbf{D}_m = [d_{p,q}^{(m)}]$, with $m = 0, 1, ..., N_f - 1$). Relevant information about the micro-Doppler fluctuations, also known as the micro-Doppler *signatures*, characterizing a certain range interval can be acquired through the real matrix $\mathbf{E} = [E_{m,q}]$, having size $N_f \times N_0^{'}$ and whose element on its m-th row and q-th column is evaluated as

$$E_{m,q} \triangleq \sum_{p=p_{\min}}^{p_{\max}} |d_{p,q}^{(m)}|^2 \tag{1.30}$$

with $m = 0, 1, ..., N_f - 1$ and $q = 0, 1, ..., N_0^{'} - 1$; here, $p_{\min}$ ($p_{\max}$) denotes the value of the index p associated with the minimum (maximum) range of interest. Representing the elements of the matrix $\mathbf{E}$ on a Cartesian plane produces the so called *spectrogram* [59], that shows the time evolution of the Doppler phenomenon.

Additional information about the dynamical properties of a moving target can be acquired through another diagram, known as *cadence velocity diagram* (CVD). This diagram allows us to identify the most relevant frequency components associated with a given motion (e.g., if a walking pedestrian is considered, the speed of his arms can be extracted from the associated CVD). Moreover, its generation is based on the complex matrix $\mathbf{G} = [G_{l,q}]$, having size $N_f^{'} \times N_0^{'}$ and computed as the $N_f^{'} \times N_0^{'}$-th order FFT of the matrix $\mathbf{E}_{ZP} = [E_{m,q}^{(ZP)}]$, that results from zero padding of the matrix $\mathbf{E}$ defined above; therefore, we have that

$$G_{l,q} \triangleq \frac{1}{N_f} \sum_{m=0}^{N_f^{'}-1} E_{m,q}^{(ZP)} \exp\left(-j2\pi \frac{m}{N_f^{'}} \bar{f}_l \, T_F\right) \tag{1.31}$$

with $l = 0,\ 1,\ ...,\ N_f^{'} - 1$ and $q = 0,\ 1,\ ...,\ N_0^{'} - 1$; here, T_F is the duration of a single transmission frame, $E_{m,q}^{(ZP)} = E_{m,q}$ for $m = 0,\ 1,\ ...,\ N_f - 1$ and $E_{m,q}^{(ZP)} = 0$ for $m > N_f - 1$, and

$$\bar{f}_l \triangleq \frac{l}{T_F} \tag{1.32}$$

is the l-th *cadence frequency*. The CVD results from representing, on a Cartesian plane, the absolute value of the elements of the matrix $\mathbf{G}$.

Finally, it is important to note that, in the development of detection and estimation algorithms for colocated MIMO radar systems operating at millimeter waves, the following technical issues need to be taken carefully into account:

1. These radar systems often operate at short ranges and in the presence of extended targets. Each radar image is a *cloud of point targets* whose mutual spacing can be very small [60]. For this reason, the accuracy of these images depends, first of all, on the frequency resolution (*delay* resolution) achieved by the detection and estimation algorithm employed on each virtual antenna in a FMCW radar system. In fact, this makes the radar receiver able to separate point targets characterized by similar ranges.

2. Distinct radar echoes can be characterized by substantially different *signal-to-noise ratios* (SNRs), because of relevant differences among the amplitudes of the L overlapped oscillations forming the useful component of the received signal (see (1.9) and (1.11)).

3. The number N of samples acquired over each virtual channel usually ranges from few hundreds to few thousands.

The last two issues explain why significant attention must be paid to the accuracy achieved by the adopted detection and estimation algorithms at low SNRs and/or for relatively small values of N, since this can appreciably influence the quality of the generated radar image.

2. Range & Angle Serial Cancellation Algorithm

2.1 Introduction

It is well known that achieved by any colocated MIMO radar system depends not only on some important characteristics of its hardware (e.g., the operating frequency, the number of transmit and receive antennas, the configuration of the transmit and receive arrays, etc.), but also on the techniques employed in the generation of its radiated waveforms and in the processing of the measurements acquired through its receive array. As far as the last issue is concerned, it is worth stressing that *optimal* (i.e., *maximum likelihood*, ML) techniques for the estimation of the overall number of targets and of their spatial coordinates cannot be employed in practice, since they require solving complicated multidimensional optimization problems and, consequently, entail a huge computational effort, even in the presence of a small number of targets [50]. This has motivated the development of various *sub-optimal* estimation techniques able to achieve good estimation accuracy at a manageable computational cost. A well known sub-optimal technique employed in real world radar systems is the one described in ref. [61] for *frequency modulated continuous wave* (FMCW) radar systems; this requires:

a) the computation of a multidimensional *Fast Fourier Transform* (FFT) of the matrix collecting the time-domain samples of the signals acquired through the receive array of the employed radar device;

b) the search for the peaks of the resulting amplitude spectrum over a range-azimuth-elevation domain or a range-azimuth domain in *three-dimensional* (3D) and *two-dimensional* (2D) imaging, respectively. Despite the practical importance of this technique, the computational effort it requires is still significant, since it involves multidimensional spectral analysis of the acquired signals. Moreover, it suffers from the following relevant drawback: it can miss targets whose electromagnetic echoes are weaker than those generated by other spatially close targets; this is due to the fact the spectral contribution due to weak echoes is usually hidden by the leakage originating from stronger echoes. This drawback may substantially affect the overall quality of radar imaging in the presence of extended targets, since such targets can be usually modelled as a cluster of point targets characterized by different radar cross sections [60].

Alternative sub-optimal techniques available in the technical literature are based on the idea of turning a complicated multidimensional optimization problem into a series of simpler and interconnected optimization sub-problems, each of which involves a search for the local maxima of a specific cost function over a limited *one-dimensional*

(1D) or 2D parameter space. Examples of this approach are offered by [62] and [63], and by [64], where range-azimuth estimators for FMCW MIMO radars are derived. More specifically, on the one hand, target delays are first estimated by applying the *multiple signal classification* (MUSIC) algorithm to a temporal auto-correlation matrix or by identifying the beat frequencies in the downconverted received signal through spectral analysis (in particular, through the FFT algorithm) in [62] and in [63], respectively; then, the acquired information are exploited to estimate the *direction of arrival* (DOA) of the echoes originating from detected targets. On the other hand, a different approach is proposed in [64], where various *iterative deterministic methods* applicable to a 2D propagation scenario are derived. These methods have the following relevant features: 1) they process a single snapshot of the received signal (acquired over the whole antenna array); 2) they estimate a new target in each iteration; 3) they do not require prior knowledge of the overall number of targets; 4) they involve 1D or 2D maximizations only; 5) they achieve a good accuracy at a reasonable computational cost; 6) the computational effort they require can be easily controlled by setting a threshold on the maximum number of targets to be detected.

In the following, three new detection and estimation algorithms for 2D and 3D radar imaging are developed. They represent different versions of the same algorithm, called *range & angle serial cancellation algorithm* (RASCA), and can be interpreted as embodiments of a general approach to target detection and estimation. In addition, they share some important features with the detection and estimation techniques developed in [63] and [64]. In fact, similarly as the techniques illustrated in [64], they are deterministic, process a single snapshot, operate in an iterative fashion and are computationally efficient; the last feature can be related to the fact they require the evaluation of 1D FFTs only and the search for the global maximum of proper cost functions over 1D (frequency, azimuth or elevation) domains. Moreover, similarly as [63], they first extract the range of each detected target from the spectra of the received signals and, then, fuse the information originating from such spectra to extract DOA information. In addition, they exploit the iterative estimation techniques developed in [65] and based on a serial cancellation approach for evaluating the parameters of multiple overlapped sinusoids or multiple overlapped complex exponentials in the presence of additive noise. The devised algorithms are able to accurately detect and estimate multiple close targets, and to solve the problem of *merged measurements* or *unresolved measurements* [8, 66–68], in the sense that targets generating merged measurements in the range domain are resolved in the estimation of their angular coordinates. The general approach to target detection and estimation on which our algorithms are based is illustrated in Section 2.2, whereas the algorithms themselves are described in Sections 2.3, which are devoted to FMCW radars. Various important details about these algorithms are provided in Section 2.4, whereas some technical limitations encountered in their implementation in real world radar systems are discussed in Section 2.5. A description of other FFT-based and MUSIC-based detection and estimation algorithms with which our algorithms are compared is provided in Section 2.6. The computational cost of all the considered algorithms is illustrated in Section 2.7, whereas their performance is analysed in Section 2.8, where various numerical results, based on both synthetically generated data and experimental measurements, are illustrated. Finally, some conclusions are offered in Section 2.9.

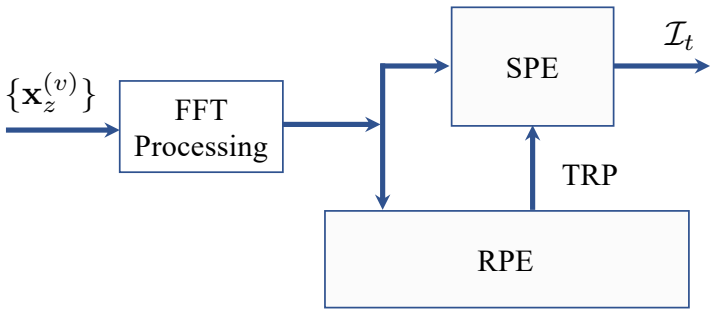


Figure 2.1: *Block diagram describing the general approach to target detection and estimation adopted in this work.*

2.2 Description of the proposed approach to the detection and estimation of multiple targets

All the algorithms developed in the following section can be considered as specific instances of a general approach to target detection and estimation; this approach is described by the block diagram shown in Fig. 2.1. The processing accomplished by the blocks which this diagram consists of can be summarized as follows. Each vector of the set $\{\mathbf{x}_z^{(v)}\}$, collecting N_{VR} vectors (see (1.17)), undergoes FFT processing, so that, in a FMCW radar system, the analysis of the acquired measurements is moved from the time-domain to the frequency-domain (time-domain). The output of the FFT block is processed by the *range profile estimator* (RPE), that generates the so called *target range profile* (TRP), i.e. a collection of: a) the ranges at which the relevant echoes are detected; b) the associated energies. Note that the last quantities allow us to rank each range on the basis of its perceptual importance. The output of the FFT processing block and the target range profile are processed by the *spatial estimator* (SPE). This block detects all the targets associated with each range appearing in the TRP and estimates their angular parameters; moreover, it may generate a finer estimate of their range. The SPE output is represented by the set

$$\mathcal{I}_t \triangleq \left\{ \left(\hat{R}_l, \hat{\theta}_l, \hat{\phi}_l, \left| \hat{C}_l \right| \right) ; l = 0, 1, \ldots, \hat{L} - 1 \right\} \tag{2.1}$$

or the set

$$\mathcal{I}_t \triangleq \left\{ \left(\hat{R}_l, \hat{\theta}_l, \left| \hat{C}_l \right| \right) ; l = 0, 1, \ldots, \hat{L} - 1 \right\} \tag{2.2}$$

in the case of 3D and 2D imaging, respectively; here, $\hat{L}$ represents an estimate of the parameter L (i.e., of the overall number of point targets), whereas $\hat{R}_l$, $\hat{\theta}_l$, $\hat{\phi}_l$ and $|\hat{C}_l|$ represent an estimate of the range R_l, azimuth θ_l, elevation ϕ_l and amplitude $|C_l|$, respectively, of the l-th target (with $l = 0$, 1, ..., $\hat{L} - 1$).

It is important to point out that:

1. If this approach is adopted, range estimation is decoupled from angular estimation, so that a 3D (2D) detection and estimation problem is turned into a) a 1D detection/estimation problem involving the detection of multiple targets and the estimation of their ranges only plus b) a 2D (1D) estimation problem concerning

the targets detected at the same range and the estimation of their azimuth and elevation (azimuth only). Consequently, the overall problem of detecting multiple targets and estimating their range and angles is turned into a couple of simpler detection and estimation problems.

2. The SPE exploits the range information generated by the RPE in order to *concentrate its computational effort on a set of well defined ranges*; this allows to reduce the size of the search space involved in spatial estimation. This explains also why the processing accomplished by the SPE cannot start before that at least a portion of the range/energy information (i.e., a portion of the TRP) generated by the RPE becomes available.

3. Various techniques can be exploited in the RPE and in the SPE to develop computationally efficient embodiments of the proposed approach.

As far as the last point is concerned, the following techniques can be adopted by the RPE to mitigate its complexity:

a) *Antenna selection* – This consists in feeding the RPE with a subset of the outputs of the FFT processing block; such outputs are generated on the basis of N_A of the N_{VR} VAs. Note that, on the one hand, a larger N_A allows to compute a more accurate TRP; on the other hand, selecting a smaller N_A results in a reduction of the overall effort required for the computation of the TRP.

b) *Antenna-by-antenna processing* – The measurements acquired through the selected N_A VAs can be efficiently processed by adopting a two-step procedure. In the first step, target range estimation is accomplished on each VA *independently of all the other* VAs, i.e. the acquired measurements are processed on an antenna-by-antenna basis; this is beneficial when parallel computing hardware is employed in the execution of the first step. In the second step, instead, the target range information extracted from each of the selected N_A VAs are fused to generate the TRP.

c) *Serial target cancellation in the range domain* – Target detection and range estimation on each VA represent a *multidimensional problem* since they aim at detecting multiple targets and estimating their ranges. In our method, this multidimensional problem is turned into a sequence of 2D estimation problems by adopting a *serial interference cancellation* (SIC) approach (e.g., see [64]). This means that the noisy signal observed on each VA is processed in an iterative fashion. In each iteration, a single (and, in particular, the strongest) target is detected, and its range and complex amplitude are estimated. Then, the contribution of this target to the received signal is estimated and subtracted from the signal itself (i.e., the detected target is treated as a form of *interference* to be cancelled), so generating a *residual signal*. The last signal represents the input of the next iteration. This procedure is repeated until the overall energy of the residual drops below a given threshold. Note also that the use of this SIC-based approach allows us to mitigate the impact of the spectral leakage due to strong targets, that can potentially hide weak targets having similar ranges.

d) *Alternating maximization* – The estimation of the normalised frequency (or delay) and the complex amplitude of a detected target requires searching for the maximum (or the minimum) of a proper cost function over a 2D domain. In our method, the *alternating maximization* (AM) technique is exploited to develop iterative algorithms that alternate the estimation of the normalised frequency (or delay) of a given target with that of its complex amplitude; for this reason, a 2D optimization problem is turned into a couple of interacting 1D optimization problems (e.g., see [69, Par. IV-A]).

In the SPE block, instead, the following techniques can be employed to reduce its overall computational complexity:

a) *Alternating maximization* – The AM technique is exploited to develop iterative algorithms that alternate the estimation of the elevation of a given target with that of its azimuth. This allows us to decouple the estimation of target elevation from that of its azimuth.

b) *Serial target cancellation in the angular domain* – Each of the ranges collected in the TRP is associated with an unknown number of targets; for this reason, the processing accomplished by the SPE aims at resolving all the targets associated with a given range and estimating their angular coordinates. In the technical literature about radar systems, the detection of an unknown number of targets characterized by the *same range* (or by ranges whose mutual differences are below the range resolution of the employed radar system) and the estimation of their angular parameters is known to be a difficult multidimensional problem (e.g., see [62, Par. III-C]). In our method, a SIC approach is exploited to turn this multidimensional problem into a sequence of 2D (1D) estimation problems in 3D (2D) imaging (see [64] and references therein). This means that the noisy data referring to a specific range and acquired on all the VAs are iteratively processed to detect a single (and, in particular, the strongest) target, and to estimate its angular coordinates and complex amplitude. Then, the contribution of this target to the outputs of the FFT processing block is estimated and subtracted from them, so generating a set of *residual data*. This *detection/estimation/cancellation* procedure is iteratively applied to the residual data until their overall energy drops below a given threshold. Moreover, in a 3D propagation scenario, this procedure is combined with the AM approach described in the previous point; this allows to detect and estimate the angular parameters of a single target (i.e., to solve a 2D optimization problem) by means of an iterative procedure alternating the estimation of its elevation with that of its azimuth (i.e., by means of an algorithm solving a couple of 1D optimization problems). Note also that, once again, the use of a serial cancellation approach allows us to mitigate the impact of the spectral leakage due to strong targets, that can potentially hide weak targets having similar spatial coordinates.

c) *Parallel processing of the data associated with different ranges* – The detection and the estimation of the targets associated with distinct ranges of the TRP can be accomplished in a *parallel fashion* or in a *sequential fashion*. The first approach

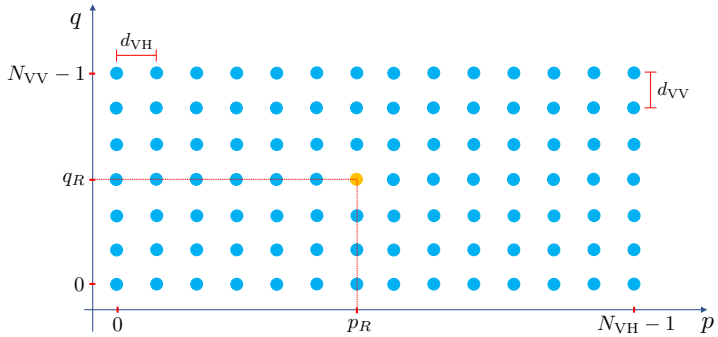


Figure 2.2: *Virtual antenna array considered in the description of our detection and estimation algorithms. The reference VA selected in this case is identified by a yellow circle.*

is more efficient than the second one if spatial estimation is executed on parallel computing hardware. In fact, in this case, multiple spatial estimation algorithms can be run *in parallel*, one for each of the ranges appearing in the TRP. Note, however, that the price to be paid for this is represented by the fact that the target information generated by all the parallel procedures need to be fused when they end. In fact, the analysis of the measurements referring to close ranges appearing in the TRP may lead to detecting the same target more than one time.

Based on the general approach outlined above and on the techniques listed for the RPE and the SPE, four specific algorithms, called *range & angle serial cancellation algorithms* (RASCAs) are developed in the following. These algorithms apply to colocated MIMO FMCW radar systems and are called RASCA-FR2 (RASCA-FC2) and RASCA-FR3 (RASCA-FC3), since they generate 2D and 3D radar images, respectively, on the basis of real (complex) measurements. In the description of these algorithms we assume, without loosing generality, that the available measurements are acquired through the $N_{VH} \times N_{VV}$ virtual *uniform rectangular array* (URA) represented in Fig. 2.2 in the case of 3D imaging and through an *horizontal* ULA (HULA), consisting of N_{VH} virtual antennas, in the case of 2D imaging. In the first case, the horizontal (vertical) spacing between adjacent antennas is denoted d_{VH} (d_{VV}), whereas, in the second one, is denoted d_{VH}. Moreover, in our considerations, we assume that a *reference* VA, identified by $(p, q) = (p_R, q_R)$ $(p = p_R)$ in the 3D (2D) case, is selected in the virtual array, as exemplified by Fig. 2.2.

In the following sections, all the RASCAs are described. In Section 2.3 we first focus on the RASCA-FR3 and RASCA-FC3, i.e. on the algorithms to be employed in a MIMO FMCW radar system equipped with a 2D antenna array (in particular, with the URA shown in Fig. 2.2). Then, we show how to adapt these algorithms to the case in which this radar system is equipped with a 1D antenna array (in particular, with an ULA); this leads to the RASCA-FR2 and the RASCA-FC2.

Table 2.1: *Most relevant parameters, sets, vectors and matrices appearing in our descrption of RASCA-FC3.*

Parameter	Description
$\mathbf{x}_{0,\mathrm{ZP}}^{(v)}$	Zero-padded version of the vector $\{\mathbf{x}_z^{(v)}\}$.
$\mathbf{x}_{m,\mathrm{ZP}}^{(v)}$	Zero-padded version of the vector $\{\mathbf{x}_m^{(v)}\}$ (with $m = 1$ and 2).
$\mathbf{X}_m^{(v)}$	DFT of the vector $\mathbf{x}_{m,\mathrm{ZP}}^{(v)}$ (with $m = 0$, 1 and 2).
$\mathscr{S}_{\mathrm{FFT}}$	Set collecting the N_{VR} triads $(\mathbf{X}_0^{(v)}, \mathbf{X}_1^{(v)}, \mathbf{X}_2^{(v)})$ such that $v = 0, 1, \ldots, N_{\mathrm{VR}} - 1$.
N_A	Number of virtual antennas exploited by the RPE.
$\mathscr{S}_A$	Set collecting the N_A indices of the virtual antennas exploited by the RPE.
$\bar{\mathscr{S}}_{\mathrm{FFT}}$	Set collecting the N_A triads $(\mathbf{X}_0^{(v)}, \mathbf{X}_1^{(v)}, \mathbf{X}_2^{(v)})$ such that $v \in \mathscr{S}_A$.
T_{STDREC}	Threshold set in the STDREC algorithm to limit its overall number of iterations.
$\mathrm{X}_m^{(v_k)}[i]$	Residual spectrum available for the v_k-th antenna at the end of the $(i-1)$-th iteration of the STDREC algorithm.
$E_i^{(v_k)}$	Energy associated with the residual spectrum $\mathrm{X}_m^{(v_k)}[i]$ and computed by the STDREC algorithm for the v_k-th antenna at the end of its $(i-1)$-th iteration.
$\hat{C}_i^{(v_k)}$	Estimate of the complex amplitude $C_i^{(v_k)}$ computed for the v_k-th antenna in the i-th iteration of the STDREC algorithm.
$\hat{F}_i^{(v_k)}$	Estimate of the normalised frequency $F_i^{(v_k)}$ computed for the v_k-th antenna in the i-th iteration of the STDREC algorithm.
$\mathbf{C}_{X_m}^{(v_k)}[i]$	Estimate of the contribution given by the i-th target detected on the v_k-th antenna to the vector $\mathrm{X}_m^{(v_k)}[i]$ (with $m = 0$, 1 and 2).
$\mathscr{S}_{\mathrm{RPE}}$	Set collecting the information generated by the RPE and feeding to the SPE.
50 T_{STDAEC}	Threshold set in the STDAEC algorithm to limit its overall number of recursions.
$\mathbf{X}^{(i)}[l]$	Matrix representing the spectral contribution given by the l-th frequency bin and available at the beginning of the i-th iteration of the STDAEC algorithm.
$E^{(i)}[l]$	Energy associated with the residual spectrum $\mathbf{X}^{(i)}[l]$ made available by the STDAEC algorithm for the l-th frequency bin at the end of its $(i-1)$-th iteration.
$\mathbf{s}_{\mathrm{VULA},k}^{(i)}[l]$	k - th vector collecting the spectral information computed for the reference VULA and the l-th frequency bin in the i-th iteration of the STDAEC algorithm (with $k = 0, 1$ and 2).
$\hat{C}_{\mathrm{V},i}[l]$	Estimate of the complex amplitude $C_{\mathrm{V},i}[l]$ computed for the target detected in the l-th frequency bin within the i-th iteration of the STDAEC algorithm.
$\hat{F}_{\mathrm{V},i}[l]$	Estimate of the normalised vertical spatial frequency $F_{\mathrm{V},i}[l]$ computed for the target detected in the l-th frequency bin within the i-th iteration of the STDAEC algorithm.
$R_i^{(\mathrm{VF})}[l,q]$	Phase rotation factor computed for the q-th VULA and the l-th frequency bin within the i-th iteration of the STDAEC algorithm (vertical folding).
$\mathbf{X}_i^{(\mathrm{VF})}[l]$	Vertically folded spectrum computed for the l-th frequency bin within the i-th iteration of the STDAEC algorithm.
$\hat{C}_{\mathrm{H},i}[l]$	Estimate of the complex amplitude $C_{\mathrm{H},i}[l]$ computed on the basis of a vertically folded spectrum within the i-th iteration of the STDAEC algorithm for the l-th frequency bin.
$\hat{F}_{\mathrm{H},i}[l]$	Estimate of the normalised horizontal spatial frequency $F_{\mathrm{H},i}[l]$ computed on the basis of a vertically folded spectrum within the i-th iteration of the STDAEC algorithm.
$R_i^{(\mathrm{HF})}[l,p]$	Phase rotation factor computed for the p-th HULA and the l-th frequency bin within the i-th iteration of the STDAEC algorithm (horizontal folding).
$R_i^{(\mathrm{HV})}[l,p,q]$	Phase rotation factor computed for the (p,q) VA and the l-th frequency bin within the i-th iteration of the STDAEC algorithm (overall folding).
$\{\mathbf{X}_{m,\mathrm{OF}}^{(i)}[l]\}$	Overall folded spectrum computed for the l-th frequency bin within the i-th iteration of the STDAEC algorithm.
$\hat{C}_i[l]$	Estimate of the complex amplitude $C_i[l]$ computed on the basis of a overall folded spectrum within the i-th iteration of the STDAEC algorithm.
$\hat{F}_i[l]$	Estimate of the normalised frequency $F_i[l]$ computed on the basis of a overall folded spectrum within the i-th iteration of the STDAEC algorithm.
$\mathbf{C}_{X_0}^{(i)}[l]$	Estimate of the contribution, given by the target detected within i-th iteration of the STDAEC algorithm, to the vector $\mathbf{X}^{(i)}[l]$.
$\hat{R}_i[l]$, $\hat{\theta}_i[l]$, $\hat{\phi}_i[l]$	Estimates of the range, azimuth and elevation made available by the SPE for the i-th target detected in the l-th frequency bin.

2.3 Range & angle serial cancellation algorithms for a frequency modulated continuous wave radar system

In this paragraph, we provide a short description of the architecture of the RASCAs for FMCW radar systems and comment on the method we developed for target detection and cancellation in the angular domain. Then, we illustrate RASCA-FR3 and RASCA-FC3 in detail. Finally, we show how to derive the RASCA-FR2 and RASCA-FC2 from them.

2.3.1 Architecture of the range & angle serial cancellation algorithms for a frequency modulated continuous wave radar system

The inner structure of the RASCAs for an FMCW radar system is described by the block diagram shown in Fig. 2.3, whereas the meaning of the most relevant parameters, sets, vectors and matrices appearing in the description of this algorithm is summarised in Table 2.1. The processing accomplished inside the blocks appearing in that figure can be summarized as follows. The FFT *processing* block turns the *time domain* information provided by the set of N_{VR} vectors $\{\mathbf{x}_z^{(v)}\}$ into the *frequency domain* information feeding both the RPE and the SPE blocks. This transformation requires the evaluation of $3N_{\mathrm{VR}}$ FFTs, all of order N_0. In fact, it consists in the evaluation of the triad ($\mathbf{X}_0^{(v)}$, $\mathbf{X}_1^{(v)}$, $\mathbf{X}_2^{(v)}$), collecting three N_0-dimensional vectors, on the basis of $\mathbf{x}_z^{(v)}$, for $v = 0$, 1, ..., $N_{\mathrm{VR}} - 1$. For this reason, the output of the considered block is represented by the set

$$\mathscr{S}_{\mathrm{FFT}} \triangleq \left\{ \left(\mathbf{X}_0^{(v)}, \mathbf{X}_1^{(v)}, \mathbf{X}_2^{(v)} \right) ; v = 0, 1, ..., N_{\mathrm{VR}} - 1 \right\}, \tag{2.3}$$

consisting of $3 \cdot N_{\mathrm{VR}}$ N_0-dimensional vectors. Note that, however, a portion of this set is discarded by the RPE, since this block processes the information originating from N_A distinct VAs only. The triads selected by the RPE form the subset

$$\bar{\mathscr{S}}_{\mathrm{FFT}} \triangleq \left\{ \left(\mathbf{X}_0^{(v)}, \mathbf{X}_1^{(v)}, \mathbf{X}_2^{(v)} \right) ; v \in \mathscr{S}_A \right\}, \tag{2.4}$$

of $\mathscr{S}_{\mathrm{FFT}}$ (2.3); here,

$$\mathscr{S}_A \triangleq \{v_0, v_1, ..., v_{N_A - 1}\}, \tag{2.5}$$

represents the set of the values of the VA index v identifying the elements of $\mathscr{S}_{\mathrm{FFT}}$ that belong to $\bar{\mathscr{S}}_{\mathrm{FFT}}$. Each of the triads of $\bar{\mathscr{S}}_{\mathrm{FFT}}$ is processed, independently of all the other ones, by a novel iterative estimation algorithm called *single target detection, range estimation and cancellation* (STDREC). This algorithm detects the most relevant targets on the selected antenna and estimates their ranges (i.e., the frequencies associated with these ranges; see (1.14) and (1.18)) and their complex amplitudes (see (1.16) and (1.12)). The name of this algorithm originates from the fact that, in each of its iterations, it *detects* a single target (namely, the strongest target), *estimates* its parameters (and, in particular, the frequency characterizing it, i.e. its range) and *cancels* the target contribution to the received signal; the residual signal resulting from target cancellation represents the input of the next iteration. The output of the STDREC algorithm that processes the raw data originating from the v_k-th VA is represented by

the set[1]

$$\mathscr{S}_{v_k} \triangleq \left\{ \left(\hat{\alpha}_i^{(v_k)}, \hat{C}_i^{(v_k)} \right), i = 0, 1, \ldots, L_k - 1 \right\}, \tag{2.6}$$

with $k = 0, 1, ..., N_A - 1$; here, L_k is the overall number of targets detected on the considered VA, whereas $\hat{C}_i^{(v_k)}$ and $\hat{\alpha}_i^{(v_k)}$ represent the estimate of the complex amplitude of the i-th target and the index of the *frequency bin*[2] in which this target has been detected. Finally, the information provided by the N_A sets $\{\mathscr{S}_{v_k}\}$ are merged to generate the single set

$$\mathscr{S}_{\text{RPE}} \triangleq \{ (\hat{\alpha}_l, E_{b,l}), l = 0, 1, \ldots, L_b - 1 \}, \tag{2.7}$$

where L_b is the overall number of targets detected on all the selected VAs, $\hat{\alpha}_l$ is the index of the *frequency bin* in which the l-th target has been detected and $E_{b,l}$ is the average energy estimated for it. Note that:

a) The cardinality L_b of the set $\mathscr{S}_{\text{RPE}}$ represents a preliminary estimate of the overall number of targets; in fact, multiple targets having the same range or ranges whose mutual differences are below the resolution of the employed radar system are detected as a single target and no effort is made at this stage to separate their contributions.

b) The energies $\{E_{b,l}\}$ represent the perceptual importance of the identified frequency bins, in the sense that a larger energy is associated with a more important frequency bin.

Both the sets $\mathscr{S}_{\text{FFT}}$ (2.3) and $\mathscr{S}_{\text{RPE}}$ (2.7) feed the SPE. The aim of this block is to analyse the spectral information associated with the ranges (i.e., with the frequency bins) identified by the RPE in order to: a) estimate the angular coordinates (i.e., azimuth and elevation) of the targets contributing to each frequency bin; b) detect additional targets associated with adjacent frequency bins and potentially hidden by the spectral leakage due to stronger targets; c) estimate the angular coordinates (i.e., azimuth and elevation) of such additional targets and compute a finer estimate of their range.

The first stage of the processing accomplished by the SPE involves the whole set $\mathscr{S}_{\text{FFT}}$ (2.3) and is executed on a bin-by-bin basis, since it aims at: a) detecting all the targets that contribute to the energy of each frequency bin contained in the TRP and b) estimating their angular coordinates. For this reason, this stage consists of L_b estimators running *in parallel*; each estimator focuses on one of the L_b frequency bins (i.e., ranges) appearing in the TRP (see Fig. 2.3). Moreover, each estimator executes a novel iterative estimation algorithm, called *single target detection, angular estimation and cancellation* (STDAEC). The l-th STDAEC algorithm processes the spectral information available on the whole virtual receive array and referring to the $\hat{\alpha}_l$-th frequency bin only (with $l = 0, 1, ..., L_b - 1$), detects $L[l]$ targets contributing to

[1]Note that the complex amplitude $\hat{C}_i^{(v_k)}$ appearing in the following equations is replaced by $\hat{A}_i^{(v_k)}$ if the received sequence is complex (see eqs. (1.10) and (1.12)). This consideration holds for various equations appearing in the remaining part of this chapter.

[2]Generally speaking, the evaluation of an FFT of order N_0 leads to partitioning the normalised frequency interval [0, 1/2) in N_0 *frequency bins*.

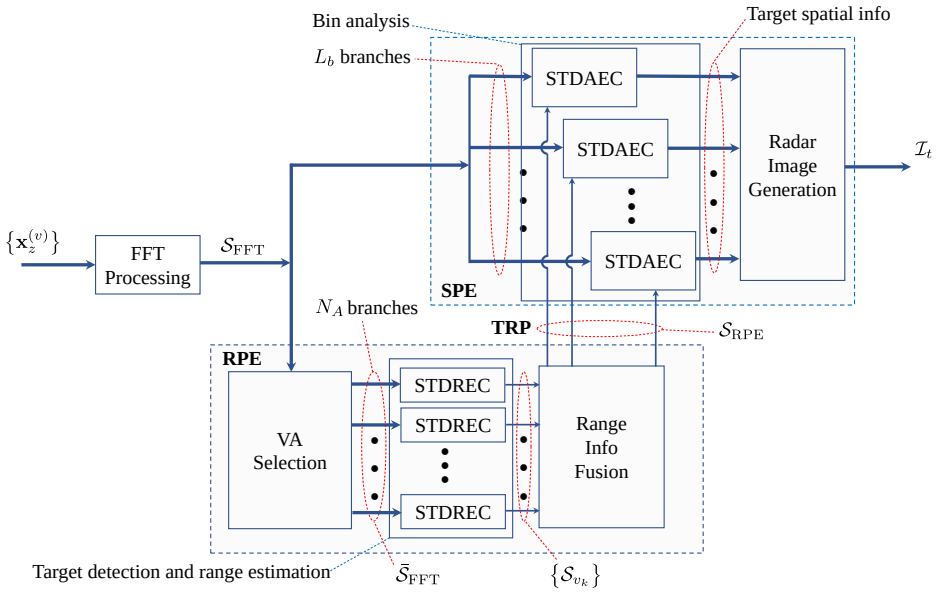


Figure 2.3: *Block diagram describing the inner structure of the RASCAs for an FMCW radar system.*

it and, for each detected target, computes: a) an estimate of its complex amplitude; b) an estimate of its angular coordinates (i.e., its azimuth and its elevation); c) a refined estimate of its range (do not forget that the preliminary estimate of this range is provided by the bin index $\hat{\alpha}_l$). If $D[l]$ iterations are accomplished by the l-th STDAEC algorithm, $D[l]$ distinct targets are detected in the $\hat{\alpha}_l$-th frequency bin, provided that none of them is classified as a *false* (i.e., *ghost*) *target*. In addition, all the estimates generated by this algorithm are collected in the set

$$\begin{aligned}\mathcal{T}_l \triangleq & \left\{\left(\hat{C}_i\left[l\right], \hat{F}_i\left[l\right], \hat{\alpha}_i\left[l\right], \hat{F}_{\mathrm{V},i}[l], \hat{F}_{\mathrm{H},i}[l]\right);\right. \\ & \left. i = 0, 1, \ldots, D[l]-1\right\}, \end{aligned} \tag{2.8}$$

or in the set

$$\begin{aligned}\mathcal{T}_l \triangleq & \left\{\left(\hat{C}_i\left[l\right], \hat{F}_i\left[l\right], \hat{\alpha}_i\left[l\right], \hat{F}_{\mathrm{H},i}[l]\right);\right. \\ & \left. i = 0, 1, \ldots, D[l]-1\right\}, \end{aligned} \tag{2.9}$$

with $l = 0, 1, \ldots, L_b - 1$, in the case of 3D and 2D radar imaging, respectively; here, $\hat{C}_i\left[l\right]$, $\hat{F}_i\left[l\right]$ and $\hat{\alpha}_i\left[l\right]$ denote the estimates of the complex amplitude $C_i\left[l\right]$, of the normalised frequency $F_i\left[l\right]$ and of the frequency bin $\alpha_i\left[l\right]$, respectively, characterizing the i-th target detected *on the reference* VA and in the l-th frequency bin, whereas $\hat{F}_{\mathrm{H},i}[l]$ and $\hat{F}_{\mathrm{V},i}[l]$ represent the estimates of the normalised horizontal spatial frequency $F_{\mathrm{H},i}[l]$ and of the normalised vertical spatial frequency $F_{\mathrm{V},i}[l]$, respectively, referring to the above mentioned target.

Finally, in the second (and last) stage of the SPE, the spatial coordinates of all the detected targets are computed on the basis of the spatial information collected in the L_b sets $\{\mathcal{T}_l\}$ and an overall image of the propagation scenario is generated in the form of a *point cloud*.

2.3.2 Some considerations on target detection and cancellation in the angular domain

It is worth pointing out that the STDAEC algorithm represents the most complicated part of the processing accomplished by all the blocks appearing in Fig. 2.3. The derivation of this algorithm relies on the fact that: a) each target provides an *additive contribution* to the spectra evaluated on all the VAs; b) *periodic variations* are observed in the phase of this contribution if we move horizontally or vertically along the considered virtual array (see Fig. 2.2). In fact, if we assume that the intensity of the echo received by each VA from the i-th target detected in the l-th frequency bin does not change from antenna to antenna, the complex amplitude $C_i[p,q,l]$ observed on the (p,q) VA can be expressed as (see (1.14) and (1.15))

$$\begin{aligned} C_i[p,q,l] =& C_i[l]\exp\Big[j\frac{4\pi}{\lambda}\big[d_{\mathrm{VH}}\,(p-p_R)\cos(\phi_i[l]) \\ & \cdot\sin(\theta_i[l]) + d_{\mathrm{VV}}\,(q-q_R)\sin(\phi_i[l])\big]\Big]; \end{aligned} \tag{2.10}$$

here, $\lambda = c/f_0$ is the wavelength associated with the start frequency (see (1.3)), (p_R, q_R) is the couple of integers identifying the selected *reference* VA, $\theta_i[l]$, $\phi_i[l]$ and $R_i[l]$ are the azimuth, the elevation and the range, respectively, characterizing the considered target, and $C_i[l]$ is its complex amplitude observed on the reference antenna. If (2.10) holds, the rate of the phase variations observed in the complex amplitudes $\{C_i[p,q,l]\}$ for a given l is proportional to (see (1.21) and (1.22))

$$F_{\mathrm{H},i}[l] \triangleq \frac{2d_{\mathrm{VH}}}{\lambda}\cos(\phi_i[l])\sin(\theta_i[l]) \tag{2.11}$$

and

$$F_{\mathrm{V},i}[l] \triangleq \frac{2d_{\mathrm{VV}}}{\lambda}\sin(\phi_i[l]), \tag{2.12}$$

if we move along an HULA and a *vertical* ULA (VULA), respectively. In fact, the quantity $F_{\mathrm{H},i}[l]$ ($F_{\mathrm{V},i}[l]$) represents the normalised horizontal (vertical) *spatial frequency* characterizing the i-th target detected in the l-th frequency bin; if both these frequencies are known, the elevation and the azimuth of this target can be evaluated as

$$\phi_i[l] = \arcsin\left(\frac{\lambda}{2d_{\mathrm{VV}}}F_{\mathrm{V},i}[l]\right) \tag{2.13}$$

and

$$\theta_i[l] = \arcsin\left(\frac{\lambda}{2d_{\mathrm{VH}}\cos(\phi_i[l])}F_{\mathrm{H},i}[l]\right), \tag{2.14}$$

respectively.

Moreover, in the derivation of the STDAEC algorithm, the following two techniques have been exploited:

Serial cancellation of targets – This technique is conceptually similar to the cancellation strategy exploited by the STDREC algorithm and allows us to detect multiple targets in the same frequency bin and, in particular, to identify targets having similar spatial coordinates. It is important to keep in mind that the frequencies associated

with distinct targets detected in the same frequency bin do not necessarily belongs to that bin; in fact, they can belong to adjacent bins, so that the tails (not the peak) of their spectra are really observed in the considered frequency bin. This problem originates from the fact that, generally speaking, *the contribution of a point target to the spectrum computed on each* VA *is not a line*, unless the associated normalised frequency is exactly a multiple of the fundamental frequency

$$F_{\mathrm{FFT}} = 1/N_0, \tag{2.15}$$

consequently, such a contribution is spread over multiple adjacent frequency bins (i.e., spectral leakage is observed)

Spatial folding – As already stated above, the frequency associated with a target detected in the l-th frequency bin does not necessarily fall exactly in that bin. The technique dubbed *spatial folding* has been devised to: a) evaluate a more accurate estimate of the frequency associated with a target detected in a given bin; b) discriminate real targets from ghost targets. Spatial folding is based on the following idea. Once the horizontal and the vertical spatial frequencies associated with a target detected in a given frequency bin have been estimated (see (2.11) and (2.12)), the spectra computed on multiple VAs can be combined *in a constructive fashion* by

1) taking a *reference* VA (identified by $(p, q) = (p_R, q_R)$; see Fig. 2.2), and compensating for the phase differences, estimated for that target, between the reference VA and the other VAs of the whole array, or

2) taking a *reference* ULA and compensating for the phase differences, estimated for that target, between the reference ULA and other ULAs *parallel* to it.

In case 1), folding generates a single spectrum, dubbed *folded spectrum*, and has the beneficial effects of a) averaging out the effects of the noise that affects the VAs and b) combining, in an unconstructive fashion, the contributions of all the targets different from the one which the employed spatial frequencies refer to. For this reason, in analysing the amplitude of the folded spectrum, a well defined peak in its amplitude is expected in the l-th frequency bin or in a bin close to it. When this peak is detected, a refined estimate of the frequency (and, consequently, of the range) and the complex amplitude characterizing the target for which folding has been accomplished can be computed by identifying its position. On the contrary, if no peak is found, the detected target is actually a *ghost target*. In case 2), folding generates as many *folded spectra* as the number of antennas of the reference ULA and offers the same advantages as case 1).

In the remaining part of this chapter, when folding is employed, the following terminology is adopted:

Vertical folding – This refers to the case in which folding involves a reference HULA on which other HULAs are folded.

Overall folding – This refers to the case in which folding involves all the spectra, i.e. the overall virtual URA; a single folded spectrum is computed in this case.

Note that, in any case, folding may involve the *whole* virtual receive array or *a portion of it*. The exploitation of a subset of the available VAs is motivated by the

fact that, in practice, in computing a folded spectrum that refers to the l-th frequency bin, the *estimates* $\hat{F}_{\mathrm{H},i}[l]$ and $\hat{F}_{\mathrm{V},i}[l]$ of the frequencies $F_{\mathrm{H},i}[l]$ and $F_{\mathrm{V},i}[l]$, respectively, are employed, so that the quality of the phase compensation factors computed for the antennas that are *farther* from the reference antenna or the reference HULA may be affected by significant *estimation errors*.

All the mathematical details about vertical and overall folding can be found in the next paragraph.

2.3.3 Detailed description of the range & angle serial cancellation algorithms for a frequency modulated continuous wave radar system endowed with a two-dimensional antenna array

In the following, the RASCA-FR3 is described first; then, the (minor) modifications required to obtain the RASCA-FC3 from it are illustrated. The RASCA-FR3 processing is divided in three tasks, each associated with one of the blocks appearing in Fig. 2.3 (the i-th task is denoted Ti); a description of each task is provided below. Various details about the techniques employed in these tasks, omitted here to ease the understanding of the overall flow of the algorithm, are provided in Section 2.4.

T1 – FFT processing

The processing accomplished within this task can be summarized as follows. Given the vector $\mathbf{x}_z^{(v)}$, the N-dimensional vectors

$$\mathbf{x}_1^{(v)} \triangleq \left[x_{1,0}^{(v)}, x_{1,1}^{(v)}, \ldots, x_{1,N-1}^{(v)}\right]^T \tag{2.16}$$

and

$$\mathbf{x}_2^{(v)} \triangleq \left[x_{2,0}^{(v)}, x_{2,1}^{(v)}, \ldots, x_{2,N-1}^{(v)}\right]^T \tag{2.17}$$

are evaluated for $v = 0, 1, ..., N_{\mathrm{VR}} - 1$; here,

$$x_{m,n}^{(v)} \triangleq n^m \, x_{z,n}^{(v)} \tag{2.18}$$

with $n = 0, 1, ..., N - 1$ and $m = 1, 2$. Then, the vectors $\mathbf{x}_z^{(v)}$, $\mathbf{x}_1^{(v)}$ and $\mathbf{x}_2^{(v)}$ undergo *zero padding* (ZP) for any v; this produces the N_0-dimensional vectors

$$\mathbf{x}_{0,\mathrm{ZP}}^{(v)} = \left[(\mathbf{x}_z^{(v)})^T \; \mathbf{0}_{(M-1)N}^T\right]^T, \tag{2.19}$$

$$\mathbf{x}_{1,\mathrm{ZP}}^{(v)} = \left[(\mathbf{x}_1^{(v)})^T \; \mathbf{0}_{(M-1)N}^T\right]^T \tag{2.20}$$

and

$$\mathbf{x}_{2,\mathrm{ZP}}^{(v)} = \left[(\mathbf{x}_2^{(v)})^T \; \mathbf{0}_{(M-1)N}^T\right]^T, \tag{2.21}$$

respectively; here, M is a positive integer (dubbed *oversampling factor*), $\mathbf{0}_D$ is a D-dimensional (column) null vector and

$$N_0 \triangleq M \, N. \tag{2.22}$$

Finally, the N_0-dimensional vectors

$$\begin{aligned}\mathbf{X}_m^{(v)} =& \left[X_{m,0}^{(v)}, X_{m,1}^{(v)}, ..., X_{m,N_0-1}^{(v)}\right]^T = \mathbf{X}_m\left[p,q\right] \\ =& \left[X_{m,0}\left[p,q\right], X_{m,1}\left[p,q\right], ..., X_{m,N_0-1}\left[p,q\right]\right]^T \\ \triangleq& \mathrm{DFT}_{N_0}\left[\mathbf{x}_{m,\mathrm{ZP}}^{(v)}\right],\end{aligned} \tag{2.23}$$

with $m = 0,\ 1,\ 2$, are computed for any v (i.e., for any p and q) by executing a N_0 order FFT for each of them; here, $\mathrm{DFT}_{N_0}\left[\mathbf{x}\right]$ denotes, up to a scale factor, the N_0 order *discrete Fourier transform* (DFT) of the N_0-dimensional vector $\mathbf{x}$. More specifically, we assume that

$$X_{m,k}^{(v)} \triangleq \frac{1}{N}\sum_{n=0}^{N-1} n^m\, x_{z,n}^{(v)} \exp\left(-j\frac{2\pi nk}{N_0}\right), \tag{2.24}$$

with $k = 0,\ 1,\ ...,\ N_0 - 1$ and $m = 0,\ 1,\ 2$.

T2 – RPE

The processing accomplished within this task consists of the three consecutive steps listed below (the i-th step is denoted T2-Si in the following); each step is associated with one of the blocks included in the RPE, as shown in Fig. 2.3.

T2-S1) VA *selection* – In this step, the set $\bar{\mathscr{S}}_{\mathrm{FFT}}$ (2.4) is built. This requires generating the set $\mathscr{S}_A$ (2.5), i.e. a set of N_A integers that identifies the selected VAs. In our computer simulations, the elements of $\mathscr{S}_A$ have been generated by *randomly* extracting N_A distinct integers from the set $\{0, 1, ..., N_{\mathrm{VR}} - 1\}$.

T2-S2) *Target detection and range estimation* – The processing carried out within this step is executed by the STDREC algorithm; this operates on an antenna-by-antenna basis. The STDREC processing for the v_k-th VA (with $k = 0,\ 1,\ ...,\ N_A - 1$) can be summarized as follows. A simple initialization is accomplished first by setting

$$\mathbf{X}_m^{(v_k)}[0] \triangleq \mathbf{X}_m^{(v_k)}, \tag{2.25}$$

with $m = 0,\ 1,\ 2$, and the iteration index i to 0. Then, the STDREC iterations are started; in the i-th iteration, the three steps described below are accomplished to detect a new target and cancel its contribution to the triad $(\mathbf{X}_0^{(v_k)}[i],\ \mathbf{X}_1^{(v_k)}[i],\ \mathbf{X}_2^{(v_k)}[i])$ (the p-th step of each is denoted STDREC-Sp in the following).

STDREC-S1) *Detection of a new target and estimation of its parameters* – The triad $(\mathbf{X}_0^{(v_k)}[i],\ \mathbf{X}_1^{(v_k)}[i],\ \mathbf{X}_2^{(v_k)}[i])$ is processed to detect a new (i.e., the i-th) target, and to estimate the normalised frequency $F_i^{(v_k)}$ and the complex amplitude $C_i^{(v_k)}$ associated with it. Note that, generally speaking, the normalised frequency $F_i^{(v_k)}$ is not a multiple of the fundamental frequency F_{FFT} (2.15), that characterizes the FFT processing executed in **T1**; for this reason, it can be expressed as

$$F_i^{(v_k)} = F_{c,i}^{(v_k)} + \delta_i^{(v_k)}\ F_{\mathrm{DFT}}, \tag{2.26}$$

where $F_{c,i}^{(v_k)}$ represents a *coarse estimate* of $F_i^{(v_k)}$ and $\delta_i^{(v_k)}$ is a real parameter called *residual*. This step consists in executing an algorithm, dubbed *single frequency esti-*

mator[3] (SFE) and whose detailed description is provided in Paragraph 2.4.1. In short, the SFE computes the estimates $\hat{C}_i^{(v_k)}$, $\hat{F}_{c,i}^{(v_k)}$, $\hat{\delta}_i^{(v_k)}$,

$$\hat{\alpha}_i^{(v_k)} = \left\lfloor \hat{F}_{c,i}^{(v_k)} / F_{\mathrm{DFT}} \right\rfloor \tag{2.27}$$

and

$$\hat{F}_i^{(v_k)} = \hat{F}_{c,i}^{(v_k)} + \hat{\delta}_i^{(v_k)} \; F_{\mathrm{DFT}} \tag{2.28}$$

of the parameters $C_i^{(v_k)}$, $F_{c,i}^{(v_k)}$, $\delta_i^{(v_k)}$, $\alpha_i^{(v_k)}$ and $F_i^{(v_k)}$, respectively, on the basis of the triad $(\mathbf{X}_0^{(v_k)}[i], \mathbf{X}_1^{(v_k)}[i], \mathbf{X}_2^{(v_k)}[i])$; here, $\hat{\alpha}_i^{(v_k)}$ represents the index of the frequency bin in which the i-th target is detected on the v_k-th antenna. Note that the parameter $\hat{F}_i^{(v_k)}$, even if useless in the construction of the set $\mathscr{S}_{v_k}$ (2.6), is exploited in the next step.

STDREC-S2) *Cancellation of the new target* – The contribution $(\mathbf{C}_{X_0}^{(v_k)}[i], \mathbf{C}_{X_1}^{(v_k)}[i], \mathbf{C}_{X_2}^{(v_k)}[i])$, given by the i-th (i.e., by the last) target detected on the v_k-th VA, to the triad $(\mathbf{X}_0^{(v_k)}[i], \mathbf{X}_1^{(v_k)}[i], \mathbf{X}_2^{(v_k)}[i])$ is computed on the basis of (2.91)–(2.93) (see Paragraph 2.4.3) and cancelled from the triad itself. Cancellation consists in the computation of the new *residual* triad

$$\begin{aligned} \mathbf{X}_m^{(v_k)}[i+1] =& \left[X_{m,0}^{(v_k)}[i+1], ..., X_{m,N_0-1}^{(v_k)}[i+1] \right]^T \\ \triangleq& \mathbf{X}_m^{(v_k)}[i] - \mathbf{C}_{X_m}^{(v_k)}[i], \end{aligned} \tag{2.29}$$

with $m = 0,\ 1,\ 2$.

STDREC-S3) *Computation of the residual energy in the frequency domain* – The energy

$$E_{i+1}^{(v_k)} \triangleq \left\| \mathbf{X}_0^{(v_k)}[i+1] \right\|^2 = \sum_{p=0}^{N_0-1} \left| X_{0,p}^{(v_k)}[i+1] \right|^2 \tag{2.30}$$

characterizing the residual spectrum vector $\mathbf{X}_0^{(v_k)}[i+1]$ (2.29) is computed and compared with the positive threshold T_{STDREC} (which may depend on range, i.e. on the detected frequency). If this energy is below the threshold, the STDREC algorithm stops and $L_k = i$ relevant targets are detected on the v_k-th VA; otherwise, the recursion index i is increased by one and a new recursion is started by going back to **STDREC-S1**.

T2-S3) *Fusion of range information* – This step aims at merging the information provided by the N_A sets $\{\mathscr{S}_{v_k}\}$ (2.6) evaluated in the previous step. Its output is represented by the set $\mathscr{S}_{\mathrm{RPE}}$ (2.7), whose elements (i.e., the L_b couples $\{(\hat{\alpha}_l, E_{b,l})\}$) are evaluated as follows. If we define the set

$$\mathscr{A}_b^{(v_k)} \triangleq \{\hat{\alpha}_i^{(v_k)}; i = 0, 1, \ldots, L_k - 1\}, \tag{2.31}$$

identifying all the bins in which *at least one target* has been detected on the v_k-th VA (with $k = 0,\ 1,\ ...,\ N_A - 1$), the set

$$\mathscr{A}_b \triangleq \{\hat{\alpha}_l; l = 0, 1, \ldots, L_b - 1\} \tag{2.32}$$

[3]Note that our general description of the SFE includes the computation of three DFTs, that, in this case, are already evaluated in **T1**.

is generated by putting together all the distinct integers that appear at least once in the N_A sets $\{\mathcal{A}_b^{(v_k)}; k = 0, 1, ..., N_A - 1\}$. Then, the average energy $E_{b,l}$ associated with the $\hat{\alpha}_l$-th bin (with $l = 0, 1, \ldots, L_b - 1$) is computed as

$$E_{b,l} = \frac{1}{N_{b,l}} \sum_{k=0}^{N_A-1} \sum_{i=0}^{L_k-1} \left| \hat{C}_i^{(v_k)} \right|^2 \delta \left[\hat{\alpha}_i^{(v_k)} - \hat{\alpha}_l \right], \tag{2.33}$$

where

$$N_{b,l} = \sum_{k=0}^{N_A-1} \sum_{i=0}^{L_k-1} \delta \left[\hat{\alpha}_i^{(v_k)} - \hat{\alpha}_l \right] \tag{2.34}$$

represents the overall number of antennas that contribute to this energy (here, $\delta[z] = 1$ if $z = 0$ and $\delta[z] = 0$ if $z \neq 0$)

T3 – SPE

The processing accomplished within this task consists of the two consecutive steps listed below (the i-th step is denoted T3-Si in the following); each step is associated with one of the blocks contained in the SPE represented in Fig. 2.3.

T3-S1) *Bin analysis* – Within this step, L_b STDAEC algorithms are run *in parallel*, one for each of the L_b ranges (i.e., frequency bins) appearing in the TRP. A schematic description of l-th STDAEC algorithm is provided below (with $l = 0, 1, ..., L_b - 1$). This algorithm consists of three steps (its r-th step is denoted STDAEC-Sr in the following) and is initialised by

1. Setting its iteration index i to 0.

2. Setting

$$\mathbf{X}^{(0)}[l] \triangleq \mathbf{X}[l], \tag{2.35}$$

where

$$\mathbf{X}[l] \triangleq [X_{0,\hat{\alpha}_l}[p, q]], \tag{2.36}$$

is a $N_{\mathrm{VH}} \times N_{\mathrm{VV}}$ matrix collecting the spectral information available on the whole virtual receive array and referring to the $\hat{\alpha}_l$-th frequency bin only.

Then, the STDAEC algorithm starts executing its iterations. Within its i-th iteration, it accomplishes the three steps described below.

STDAEC-S1) *Detection of a new target and estimation of its angular parameters* – In this step, the $N_{\mathrm{VH}} \times N_{\mathrm{VV}}$ matrix

$$\mathbf{X}^{(i)}[l] \triangleq \left[X_l^{(i)}[p, q] \right], \tag{2.37}$$

is processed to detect the strongest target contributing to it, and to compute the estimates $\hat{\theta}_i[l]$, $\hat{\phi}_i[l]$ and $\hat{C}_i[l]$ of $\theta_i[l]$, $\phi_i[l]$) and $C_i[l]$, respectively (note that this target represents the i-th one detected in the considered frequency bin, since $(i - 1)$ targets have been detected in the previous recursions). This result is achieved by executing a novel iterative detection and estimation algorithm called *single target detection and*

angular estimation (STDAE), whose description is provided after illustrating the overall structure of the RASCA-FR3 to ease reading.

STDAEC-S2) *Target cancellation* – The contribution $\mathbf{C}_{X_0}^{(i)}[l]$, given by the i-th target detected in the l-th frequency bin, to the vector $\mathbf{X}^{(i)}[l]$ (2.37) is computed on the basis of (2.105)–(2.106) (see Paragraph 2.4.3) and is cancelled. Cancellation consists in the computation of the new *residual* vector

$$\mathbf{X}^{(i+1)}[l] \triangleq \mathbf{X}^{(i)}[l] - \mathbf{C}_{X_0}^{(i)}[l]\,. \tag{2.38}$$

STDAEC-S3) *Residual energy test* – The energy

$$E^{(i+1)}[l] \triangleq \left\|\mathbf{X}^{(i+1)}[l]\right\|^2 = \sum_{p=0}^{N_{\mathrm{VH}}-1}\sum_{q=0}^{N_{\mathrm{VV}}-1}\left|X_l^{(i+1)}[p,q]\right|^2 \tag{2.39}$$

of the residual spectrum vector $\mathbf{X}^{(i+1)}[l]$ (2.38) is compared with the positive threshold T_{STDAEC} (which may dependend on angular coordinates). If this energy is below the threshold, the STDAEC algorithm stops; otherwise, the recursion index i is increased by one and a new iteration is started by going back to **STDAEC-S1**. If $D[l]$ iterations are accomplished by the STDAEC algorithm operating on the $\hat{\alpha}_l$-th frequency bin, no more than $D[l]$ distinct targets are identified in that bin ($D[l]$ targets are found if none of them is deemed to be a ghost target). All the targets information acquired from the $\hat{\alpha}_l$-th frequency bin are collected in the set $\mathcal{T}_l$ (2.8).

T3-S2) *Evaluation of the target spatial coordinates and generation of the overall image* – In this step, the estimates of the range, of the elevation and of the azimuth of the i-th target detected in the $\hat{\alpha}_l$-th frequency bin are computed as (see (1.18), (2.13) and (2.14))

$$\hat{R}_i[l] = \frac{c}{2\mu}\hat{f}_i[l]\,, \tag{2.40}$$

$$\hat{\phi}_i[l] = \arcsin\left(\frac{\lambda}{2d_{\mathrm{VV}}}\hat{F}_{\mathrm{V},i}[l]\right) \tag{2.41}$$

and

$$\hat{\theta}_i[l] = \arcsin\left(\frac{\lambda}{2d_{\mathrm{VH}}\cos(\hat{\phi}_i[l])}\hat{F}_{\mathrm{H},i}[l]\right), \tag{2.42}$$

respectively; here, $\hat{f}_i[l] = F_i[l]\,f_s$ (see (1.13)). Finally, these information are fused to generate the overall set $\mathscr{I}_t$ (2.1), describing the generated radar image; in general, this image is a cloud of $\hat{L}$ points. The set $\mathscr{I}_t$ (2.1) results from the union of all the sets $\{\mathscr{I}_t^{(l)}\}$, where

$$\mathscr{I}_t^{(l)} \triangleq \Big\{\big(\hat{R}_i[l], \hat{\theta}_i[l], \hat{\phi}_i[l], \big|\hat{C}_i[l]\big|\big); i = 0, 1, \ldots, D[l]-1\Big\}, \tag{2.43}$$

with $l = 0, 1, \ldots, L_b - 1$.

This concludes our description of the RASCA-FR3.

Let us focus now on the most complicated part of the STDAEC algorithm, i.e. on the STDAE algorithm. This algorithm makes use of the so called *spatial folding* (see the previous paragraph). The exploitation of this procedure in the STDAE algorithm requires:

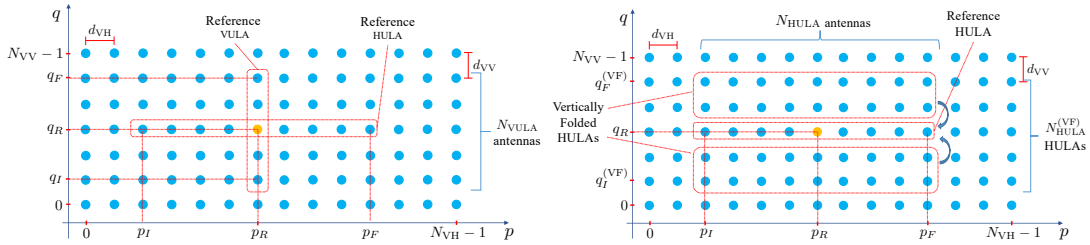


Figure 2.4: *Example of reference VULA and reference HULA including the reference antenna (a) and representation of a set of vertically folded HULAs (b).*

1. Selecting a *reference* VULA, that consists of N_{VULA} adjacent and vertically aligned VAs (with $N_{\mathrm{VULA}} \leq N_{\mathrm{VV}}$), within the virtual array; in the following, we assume, without any loss of generality, that the reference VULA includes the reference antenna and, consequently, is identified by $p = p_R$ and $q = q_I$, $q_I + 1$, ..., q_F (with $q_I \leq q_R \leq q_F$), so that $N_{\mathrm{VULA}} = q_F - q_I + 1$ (see Fig. 2.4a).

2. Selecting a *reference* HULA, that consists of N_{HULA} adjacent and horizontally aligned VAs; in the following, we assume, without any loss of generality, that the reference HULA is the horizontal ULA containing the reference antenna and, consequently, is identified by $p = p_I$, $p_I + 1$, ..., p_F (with $p_I \leq p_R \leq p_F$) and $q = q_R$, so that $N_{\mathrm{HULA}} = p_F - p_I + 1$ (see Fig. 2.4a).

3. Selecting a set of HULAs, different from the reference HULA and having the same size of it (i.e., the same number of VAs); in the following, we assume, without any loss of generality, that these HULAs, called *vertically folded* HULAs, correspond to $q = q_I^{(\mathrm{VF})}$, $q_I^{(\mathrm{VF})} + 1$, ..., $q_R - 1$, $q_R + 1$, ..., $q_F^{(\mathrm{VF})}$, with $q_I^{(\mathrm{VF})} < q_R < q_F^{(\mathrm{VF})}$, as illustrated in Fig. 2.4b; note that the overall number of involved HULAs is $N_{\mathrm{HULA}}^{(\mathrm{VF})} = q_F^{(\mathrm{VF})} - q_I^{(\mathrm{VF})} + 1$.

The STDAE algorithm consists in the four steps described below (its r-th step is denoted STDAE-Sr in the following).

STDAE-S1) FFT *processing on the reference* VULA *and vertical frequency estimation* – The portion of the initial spectral information referring to the reference VULA is extracted from the matrix $\mathbf{X}^{(i)}[l]$ and stored in the N_{VULA}-dimensional vector

$$\begin{aligned}\mathbf{S}_{\mathrm{VULA},0}^{(i)}[l] &= \left[S_{0,0}^{(i)}[l], S_{0,1}^{(i)}[l], ..., S_{0,N_{\mathrm{VULA}}-1}^{(i)}[l]\right]^T \\ &\triangleq \left[X_l^{(i)}[p_R, q_I], ..., X_l^{(i)}[p_R, q_F]\right]^T,\end{aligned} \tag{2.44}$$

that is processed by the *complex single frequency estimator*[4] (CSFE). This algorithm detects the i-th (strongest target) appearing in $\hat{\alpha}_l$-th frequency bin and computes the

[4] A detailed description of this estimator is provided in Paragraph 2.4.2. Note that this algorithm represents the complex counterpart of the SFE, in the sense that the former is fed by a complex sequence, whereas the latter by a real one.

estimates $\hat{C}_{\mathrm{V},i}[l]$ and $\hat{F}_{\mathrm{V},i}[l]$ of the parameters $C_i[l]$ and $F_{\mathrm{V},i}[l]$ (see (2.12)), respectively. Note that the quantity $\hat{C}_{\mathrm{V},i}[l]$ is not exploited in the following since, it represents a *preliminary estimate* of $C_i[l]$.

It is worth pointing out that the execution of the CSFE entails:

a) The evaluation of the N_{VULA}-dimensional vector

$$\mathbf{S}^{(i)}_{\mathrm{VULA},k}[l] \triangleq \left[S^{(i)}_{k,0}[l], S^{(i)}_{k,1}[l], ..., S^{(i)}_{k,N_{\mathrm{VULA}}-1}[l]\right]^T, \tag{2.45}$$

with $k = 1$ and 2; here,

$$S^{(i)}_{k,p}[l] \triangleq p^k \; X^{(i)}_{\mathrm{VULA},p}[l] = p^k \; X^{(i)}_{l}[p_R, q_I + p] \tag{2.46}$$

with $p = 0, 1, ..., N_{\mathrm{VULA}} - 1$.

b) The computation of an $\bar{N}_0$ order FFT of the vector $\bar{\mathbf{S}}^{(i)}_{\mathrm{VULA},k}[l]$, that represents a zero padded version of the vector $\mathbf{S}^{(i)}_{\mathrm{VULA},k}[l]$ (with $k = 0, 1$ and 2); here,

$$\bar{N}_0 \triangleq \bar{M} \cdot N_{\mathrm{VULA}} \tag{2.47}$$

and $\bar{M}$ represents the adopted oversampling factor. This produces the vector

$$\begin{aligned}\mathbf{s}^{(i)}_{\mathrm{VULA},k}[l] =& \left[s^{(i)}_{k,0}[l], s^{(i)}_{k,1}[l], ..., s^{(i)}_{k,\bar{N}_0-1}[l]\right]^T \\ \triangleq& \mathrm{DFT}_{\bar{N}_0}\left[\bar{\mathbf{S}}^{(i)}_{\mathrm{VULA},k}[l]\right],\end{aligned} \tag{2.48}$$

with $k = 0$, 1 and 2. Note that the m-th element of the vector $\mathbf{s}^{(i)}_{\mathrm{VULA},k}[l]$ can be expressed as

$$s^{(i)}_{k,m}[l] \triangleq \frac{1}{N_{\mathrm{VULA}}} \sum_{p=0}^{N_{\mathrm{VULA}}-1} S^{(i)}_{k,p}[l] \exp\left(-j\frac{2\pi pm}{\bar{N}_0}\right), \tag{2.49}$$

with $m = 0, 1, ..., \bar{N}_0 - 1$.

STDAE-S2) *Vertical folding* – The estimate $\hat{F}_{\mathrm{V},i}[l]$ of the normalised vertical frequency $F_{\mathrm{V},i}[l]$ (2.12) is employed to compensate for the phase difference between each of the HULAs selected for vertical folding and the reference HULA (i.e., for the phase differences along the *vertical* direction), so that the spectral information associated with all these HULAs can be combined (i.e., summed) in a constructive fashion. To this aim, the *phase rotation factor*

$$R^{(\mathrm{VF})}_i[l,q] \triangleq \left[\exp\left(-j\Delta\psi^{(\mathrm{VF})}_i[l]\right)\right]^{q-q_R}, \tag{2.50}$$

with

$$\Delta\psi^{(\mathrm{VF})}_i[l] \triangleq 2\pi\hat{F}_{\mathrm{V},i}[l], \tag{2.51}$$

is computed for the q-th HULA, with $q = q^{(\mathrm{VF})}_I, q^{(\mathrm{VF})}_I + 1, ..., q_R - 1, q_R + 1, ..., q^{(\mathrm{VF})}_F$. Then, vertical folding is accomplished by computing the N_{HULA}-dimensional vector

$$\mathbf{X}^{(\mathrm{VF})}_i[l] = \mathbf{X}^{(i)}[l, q_R] + \sum_{\substack{q=q^{(\mathrm{VF})}_I \\ q\neq q_R}}^{q^{(\mathrm{VF})}_F} \mathbf{X}^{(i)}[l,q] \; R^{(\mathrm{VF})}_i[l,q], \tag{2.52}$$

that collects the values taken on by the N_{HULA} vertically folded spectra referring to the $\hat{\alpha}_l$-th frequency bin; here,

$$\mathbf{X}^{(i)}[l,q] \triangleq \left[X_l^{(i)}[p_I,q], X_l^{(i)}[p_I+1,q], ..., X_l^{(i)}[p_F,q]\right]^T, \tag{2.53}$$

is a N_{HULA}-dimensional row vector extracted from the q-th row of the matrix $\mathbf{X}^{(i)}[l]$ (2.37).

STDAE-S3) FFT *processing and horizontal frequency estimation* – The processing accomplished in this step is very similar to that carried out in **STDAE-S1**. In fact, the only difference is represented by the fact that the N_{VULA}-dimensional vector $\mathbf{S}_{\mathrm{VULA},0}^{(i)}[l]$ (2.44) is replaced by the N_{HULA}-dimensional vector $\mathbf{X}_i^{(\mathrm{VF})}[l]$ (2.52) generated in the previous step. Therefore, in this case, the CSFE algorithm is exploited to compute the estimate $\hat{F}_{\mathrm{H},i}[l]$ of the horizontal frequency $F_{\mathrm{H},i}[l]$ (2.11) and a new estimate, denoted $\hat{C}_{\mathrm{H},i}[l]$, of the complex amplitude $C_i[l]$ associated with the i-th target. Note that: a) in general, an order different from $\bar{N}_0$ (2.47) can be selected for the three DFTs computed by the CSFE algorithm in this step; b) the quantity $\hat{C}_{\mathrm{H},i}[l]$ is not exploited in the following since, it represents a preliminary estimate of $C_i[l]$; c) the estimates $\hat{C}_{\mathrm{V},i}[l]$ and $\hat{C}_{\mathrm{H},i}[l]$ can be significantly different if multiple targets having similar horizontal frequencies or similar vertical frequencies contribute to the considered frequency bin.

STDAE-S4) *Overall folding and frequency/amplitude estimation* – In this step, the angular information i.e., the frequencies $\hat{F}_{\mathrm{V},i}[l]$ and $\hat{F}_{\mathrm{H},i}[l]$ computed in **STDAE-S2** and **STDAE-S3**, respectively, are exploited to accomplish overall folding[5]; this step involves the whole spectrum computed on the selected VAs. If the whole receive antenna array is exploited, overall folding consists in computing the N_0-dimensional vector

$$\mathbf{X}_{0,\mathrm{OF}}[l] \triangleq \sum_{p=0}^{N_{\mathrm{VH}}-1}\sum_{q=0}^{N_{\mathrm{VV}}-1} \mathbf{X}_0[p,q]\ R_i^{(\mathrm{HV})}[l,p,q], \tag{2.54}$$

where

$$R_i^{(\mathrm{HV})}[l,p,q] \triangleq R_i^{(\mathrm{VF})}[l,q]\ R_i^{(\mathrm{HF})}[l,p] \tag{2.55}$$

is a phase rotation factor, $R_i^{(\mathrm{VF})}[l,q]$ is expressed by (2.50),

$$R_i^{(\mathrm{HF})}[l,p] \triangleq \left[\exp\left(-j\Delta\psi_i^{(\mathrm{HF})}[l]\right)\right]^{p-p_R} \tag{2.56}$$

and

$$\Delta\psi_i^{(\mathrm{HF})}[l] \triangleq 2\pi\hat{F}_{\mathrm{H},i}[l]; \tag{2.57}$$

note that $R_i^{(\mathrm{HV})}[l,p,q]=1$ if $p=p_R$ and $q=q_R$. Given $\mathbf{X}_{0,\mathrm{OF}}[l]$ (2.54), the sequence of the absolute values of its elements is analysed to verify the presence of a peak in the $\hat{\alpha}_l$-th frequency bin or in a bin close to it. To this aim, after evaluating

$$\hat{\alpha}_{\mathrm{OF}} \triangleq \arg\max_{\tilde{a}\in\{0,1,...,N_0-1\}} |\mathbf{X}_{0,\mathrm{OF}}[\tilde{a}]|, \tag{2.58}$$

[5]As already mentioned above, a portion of the whole virtual array can be exploited to mitigate the impact of the estimation errors affecting these spatial frequencies.

the quantity $d_{\hat{\alpha}}[l] \triangleq |\hat{\alpha}_{\rm OF} - \hat{\alpha}_l|$ is compared with the positive threshold $T_{\rm OF}$. If $d_{\hat{\alpha}}[l]$ exceeds $T_{\rm OF}$, the presence of a ghost target is detected; otherwise, the N_0-dimensional vector

$$\mathbf{X}_{m,{\rm OF}}\,[l] \triangleq \sum_{p=0}^{N_{\rm VH}-1} \sum_{q=0}^{N_{\rm VV}-1} \mathbf{X}_m\,[p,q]\;\; R_i^{({\rm HV})}\,[l,p,q]\,, \tag{2.59}$$

is computed for $m = 1$ and 2, and the CSFE algorithm[6] is run to estimate, on the basis of the triad $(\mathbf{X}_{0,{\rm OF}}[l], \mathbf{X}_{1,{\rm OF}}[l], \mathbf{X}_{2,{\rm OF}}[l])$, the *final estimates* $\hat{F}_i[l]$ and $\hat{C}_i[l]$ of the parameters $F_i[l]$ and $C_i[l]$, respectively; these parameters characterize the i-th target detected in the $\hat{\alpha}_l$-th frequency bin. Note that the integer part (see (2.27))

$$\hat{\alpha}_i\,[l] \triangleq \left\lfloor \hat{F}_i[l]/F_{\rm DFT} \right\rfloor \tag{2.60}$$

of $\hat{F}_i[l]$ does not necessarily coincide with $\hat{\alpha}_l$ but, if it differs, it is certainly close to $\hat{\alpha}_l$. If $\hat{\alpha}_i[l]$ is different from $\hat{\alpha}_l$ and appears in one of the couples forming the set $\mathscr{S}_{\rm RPE}$ (2.7), it is discarded, because the corresponding frequency bin is already being analysed by one of the other STDAEC algorithms. Otherwise, the new couple

$$(\hat{\alpha}_i\,[l]\,, E_{b,L_b})\,, \tag{2.61}$$

where $E_{b,L_b} \triangleq |\hat{C}_i[l]|^2$, is added to the set $\mathscr{S}_{\rm RPE}$ and the number of its elements (i.e., L_b) is increased by one. This means that an additional STDAEC algorithm is run on the (new) $\hat{\alpha}_i[l]$-th bin.

This concludes our description of the STDAE algorithm and, consequently, of the RASCA-FR3, whose overall structure is summarised in Algorithm 1.

The RASCA-FC3 can be easily obtained from RASCA-FR3 by: a) replacing $C_i^{(v_k)}$ and $\hat{C}_i^{(v_k)}$ with $A_i^{(v_k)}$ and $\hat{A}_i^{(v_k)}$, respectively (see **STDREC-S1**); b) replacing the SFE with the CSFE in **STDREC-S1**; c) computing the vectors of the triad $(\mathbf{C}_{X_0}^{(v_k)}[i]$, $\mathbf{C}_{X_1}^{(v_k)}[i]$, $\mathbf{C}_{X_2}^{(v_k)}[i])$ on the basis of (2.101)–(2.103) (see Paragraph 2.4.3) in **STDREC-S2**; d) replacing $C_i[l]$ and $\hat{C}_i[l]$ with $\hat{A}_i[l]$ and $A_i[l]$, respectively (see **STDAEC-S1**); e) replacing $\hat{C}_{{\rm V},i}\,[l]$ and $\hat{C}_{{\rm H},i}[l]$ with $\hat{A}_{{\rm V},i}\,[l]$ and $\hat{A}_{{\rm H},i}\,[l]$, respectively (see **STDAE-S1** and **STDAE-S3**).

Additional comments

The structure of the RASCA-FR3 (RASCA-FC3) deserves a number of comments, that are listed below for the different tasks and the steps they consist of.

T1 – In this task, each of the vectors $\{\mathbf{X}_0^{(v)}, \mathbf{X}_1^{(v)}, \mathbf{X}_2^{(v)}\}$ is computed by executing a N_0 order FFT. The vector $\mathbf{X}_0^{(v)}$ collects N_0 equally spaced samples of the spectrum of the sequence $\{x_{z,n}^{(v)}\}$ acquired on the v-th VA (see (2.18), (2.23) and (1.17)). The vectors $\mathbf{X}_1^{(v)}$ and $\mathbf{X}_2^{(v)}$, instead, consist of, up to a scale factor, N_0 equally spaced samples of the *first* and the *second derivatives*, respectively, of the same spectrum.

[6]Note that our general description of the CSFE includes the computation of three order N_0 DFTs, that, in this case, are already available, being represented by $\{\mathbf{X}_{m,{\rm OF}}\,[l];\, m = 0,\, 1,\, 2\}$.

Algorithm 1: *Range & angle serial cancellation algorithm* for an FMCW radar system (real case)

T1 – FFT Processing:
Compute the vectors $\mathbf{x}_{0,\mathrm{ZP}}^{(v)}$, $\mathbf{x}_{1,\mathrm{ZP}}^{(v)}$ and $\mathbf{x}_{2,\mathrm{ZP}}^{(v)}$ according to (2.19)–(2.21); then, compute the triad $\{\mathbf{X}_0^{(v)}, \mathbf{X}_1^{(v)}, \mathbf{X}_2^{(v)}\}$ according to (2.23).

T2 – RPE:
S1) Extract N_A VAs from all the available VAs; then, build the set $\bar{\mathscr{S}}_{\mathrm{FFT}}$ (2.4).
for $k = 0$ *to* $N_A - 1$ **do**
  S2) Set $\mathbf{X}_m^{(v_k)}[0] \triangleq \mathbf{X}_m^{(v_k)}$ for $m = 0, 1, 2$ (see (2.25)); then, set the iteration index $i = 0$ and compute the initial energy $E_0^{(v_k)}$ according to (2.30).
  while $E_i^{(v_k)} > T_{\mathrm{STDREC}}$ **do**
    STDREC-S1) Compute the couple $(\hat{C}_i^{(v_k)}, \hat{F}_i^{(v_k)})$ running the SFE algorithm on the triad $(\mathbf{X}_0^{(v_k)}[i], \mathbf{X}_1^{(v_k)}[i], \mathbf{X}_2^{(v_k)}[i])$.
    STDREC-S2) Compute the vectors $(\mathbf{C}_{X_0}^{(v_k)}[i], \mathbf{C}_{X_1}^{(v_k)}[i], \mathbf{C}_{X_2}^{(v_k)}[i])$ according to (2.91)–(2.93); then, compute the new residual triad $(\mathbf{X}_0^{(v_k)}[i+1], \mathbf{X}_1^{(v_k)}[i+1], \mathbf{X}_2^{(v_k)}[i+1])$ according to (2.29).
    STDREC-S3) Compute the residual energy $E_{i+1}^{(v_k)}$ according to (2.30).
  end
end
S3) Build the set $\mathscr{S}_{\mathrm{RPE}}$ (2.7) (see (2.32) and (2.34)).

T3 – SPE:
S1) Set the iteration index $i = 0$ and set the initial vector $\mathbf{X}^{(0)}[l]$ according to (2.35); then, compute the initial energy $E^{(0)}[l]$ according to (2.39).
Parallel For $l = 0$ *to* $l = L_b - 1$ **do**
  while $E^{(i)}[l] > T_{\mathrm{STDAEC}}$ **do**
    STDAEC-S1) Compute the couple $(\hat{C}_{\mathrm{V},i}[l], \hat{F}_{\mathrm{V},i}[l])$ running the CSFE algorithm fed by the vector $\mathbf{S}_{\mathrm{VULA},0}^{(i)}[l]$ evaluated according to (2.44). Then, compute the phase rotation factor $R_i^{(\mathrm{VF})}[l,q]$ and the matrix $\mathbf{X}_i^{(\mathrm{VF})}[l]$ according to (2.50) and (2.52), respectively. Then, run the CSFE algorithm to compute the couple $(\hat{C}_{\mathrm{H},i}[l], \hat{F}_{\mathrm{H},i}[l])$ and compute the phase rotation factor $R_i^{(\mathrm{HF})}[l,p]$ according to (2.56). Finally, compute the vectors $\{\mathbf{X}_{m,\mathrm{OF}}[l]; m = 0, 1, 2\}$ according to (2.54) and (2.59) and run the CSFE algorithm fed by the set $\{\mathbf{X}_{m,\mathrm{OF}}[l]; m = 0, 1, 2\}$ to evaluate the couple $(\hat{C}_i[l], \hat{F}_i[l])$.
    STDAEC-S2) Compute the vector $\mathbf{C}_{X_0}^{(i)}[l]$ according to (2.105)–(2.106); then compute the new residual vector $\mathbf{X}^{(i+1)}[l]$ according to (2.38).
    STDAEC-S3) Compute the residual energy $E^{(i+1)}[l]$ according to (2.39).
  end
  S2) Compute $\hat{R}_i[l]$, $\hat{\phi}_i[l]$, $\hat{\theta}_i[l]$ according to (2.40)–(2.42).
end

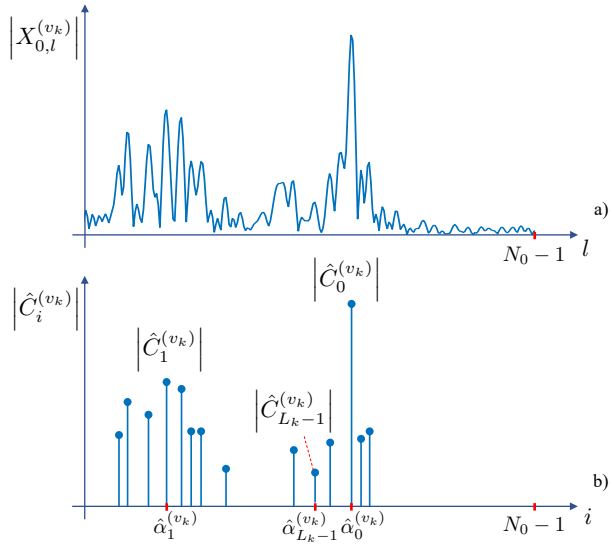


Figure 2.5: *Representation of: a) the absolute value of the elements of the vector* $\mathbf{X}_0^{(v_k)}$*; b) the corresponding discrete amplitude-range profile generated by the STDREC algorithm (the RASCA-FC3 is considered).*

T2-S1 – The exploitation of a subset of the available antennas is motivated by the need of reducing the computational effort required by **T2** as much as possible. The adoption of a *deterministic* method for the selection of N_A antennas (with $N_A < N_{\mathrm{VR}}$) is not recommended. In fact, when multiple consecutive snapshots are processed to generate independent images, randomly changing the subset of N_A antennas from snapshot to snapshot allows the considered radar system to benefit from *antenna diversity*.

T2-S2 – The STDREC algorithm deserves the following comments:

a) The availability of *accurate* estimates of the normalised frequency $F_i^{(v_k)}$ and of the complex amplitude $C_i^{(v_k)}$ ($A_i^{(v_k)}$) (see (1.10) and (1.12)) plays an important role in this step, since these parameters are exploited in the serial cancellation procedure based on (2.29). In particular, ignoring the frequency residual $\delta_i^{(v_k)}$ of the normalised frequency $F_i^{(v_k)}$ (2.26) in this procedure (i.e., assuming that $\hat{F}_i^{(v_k)} = \hat{\alpha}_i^{(v_k)}$; see (2.27)) may result in a significant *error accumulation*.

b) A threshold on the maximum computational effort required by the STDREC algorithm can be set by requiring that the recursion index i never exceeds a fixed threshold; this is equivalent to limit the overall number of targets that can be detected on each VA.

c) The STDREC algorithm generates N_A different data sets; the k-th data set consists of the triads $\{(\hat{\alpha}_i^{(v_k)}, \hat{F}_i^{(v_k)}, \hat{C}_i^{(v_k)});\ i = 0, 1, \ldots, L_k - 1\}$ ($\{(\hat{\alpha}_i^{(v_k)}, \hat{F}_i^{(v_k)}, \hat{A}_i^{(v_k)});\ i = 0, 1, \ldots, L_k - 1\}$), characterizing the L_k targets detected on the v_k-th antenna (with $k = 0, 1, ..., N_A - 1$). Note that the overall number of targets may change from antenna to antenna, especially in the presence of extended targets; this is due to the fact that the signals acquired on different VAs can exhibit significant differences in their spectral content.

d) The following important interpretation of the processing accomplished by the STDREC algorithm on the v_k-th VA can be given. The vector $\mathbf{X}_0^{(v_k)}$ can be seen as a collection of noisy spectral information referring to N_0 distinct *frequency bins* (i.e., to N_0 distinct *range bins*) and is usually *dense* in the presence of multiple extended targets, as illustrated in Fig. 2.5-a) (where the absolute value of its elements is represented). The STDREC allows to extract a *discrete frequency* (i.e., *range*) *profile* from the vector $\mathbf{X}_0^{(v_k)}$, as illustrated in Fig. 2.5-b). In various real world scenarios, this profile turns out to be *sparse*, even in the presence of a dense vector $\mathbf{X}_0^{(v_k)}$; this is beneficial, since allows to concentrate the RPE computational effort on a set of *specific ranges* (i.e., frequency bins). The *range profile* characterizing the v_k-th VA is described by the set of L_k couples $\mathscr{S}_{v_k} = \{(\hat{\alpha}_i^{(v_k)}, \hat{C}_i^{(v_k)});\ i = 0, 1, \ldots, L_k - 1\}$ ($\mathscr{S}_{v_k} = \{(\hat{\alpha}_i^{(v_k)}, \hat{A}_i^{(v_k)});\ i = 0, 1, \ldots, L_k - 1\}$), with $k = 0, 1, \ldots, N_A - 1$; the parameter $\hat{\alpha}_i^{(v_k)}$ identifies the frequency bin associated with the i-th target detected on the considered VA, whereas the absolute value of $\hat{C}_i^{(v_k)}$ ($\hat{A}_i^{(v_k)}$) represents an estimate of the *strength* of the echo associated with it.

e) The STDREC algorithm can be used for detecting multiple targets and accurately estimating their range in a monostatic radar.

f) The STDREC algorithm can be easily extended in a way that *multiple targets* are detected and estimated in parallel in each of its iterations. If we focus on its i-iteration and the v_k-th VA, this result is achieved by running multiple (say, $m_i^{(v_k)}$) instances of the SFE (CSFE) algorithm in parallel. Each of these instances is initialised with the frequency corresponding to the absolute maximum or a relative maximum detected in the sequence of the absolute values of the elements of the vector $\mathbf{X}_0^{(v_k)}[i]$ (see (2.23)). In this case, a constraint is set on the minimum spacing between the $m_i^{(v_k)}$ detected frequencies in order to minimize the interference between the instances running in parallel. Moreover, after identifying the absolute maximum in the above mentioned sequence, a threshold, proportional to such a maximum, is set on the minimum value of the acceptable relative maximum/maxima, so that unrelevant frequencies are discarded. It is also worth stressing that, if a *cluster* of $m_i^{(v_k)}$ distinct frequencies is estimated, each of the components of the triad $(\mathbf{C}_{X_0}^{(v_k)}[i], \mathbf{C}_{X_1}^{(v_k)}[i], \mathbf{C}_{X_2}^{(v_k)}[i])$ appearing in the RHS of (2.29) consists of the sum of $m_i^{(v_k)}$ terms, each associated with one of these frequencies.

g) The STDREC algorithm employed in the RASCA-FR3 (RASCA-FC3) represents an instance of the *single frequency estimation and cancellation* (*complex single frequency estimation and cancellation*) algorithm derived in [65] for the estimation of multiple overlapped *real* (*complex*) tones. For this reason, in the case of complex received signals, it can be replaced by one of the *multiple tone estimators* available in the technical literature, like the CFH algorithm [70], the algorithm developed by *Ye* and *Aboutanios* in [71, 72] and the algorithm derived by *Serbes* in [73] (the last two algorithms are denoted Alg-YA and Alg-S, respectively, in the following). In fact, all these algorithms are recursive and rely on a serial cancel-

lation procedure since, within each recursion, they detect a single tone, estimate its parameters and subtract its contribution from the residual signal emerging from the previous iteration.

h) The estimates generated by the STDREC algorithm are potentially biased if the parameters of the SFE (CSFE) executed in its first step are not properly selected (see [65]). In principle, this bias can be arbitrarily reduced by increasing the overall number of iterations and/or re-estimations accomplished by the SFE (CSFE). However, we found out that, in the case of complex received signal, a computationally efficient alternative to this approach is represented by running an additional step (i.e., STDREC-S4) after that the first three steps of the STDREC algorithm has been carried out. In this final step, the Alg-YA is run after initializing it with the estimates of the normalised frequencies and the associated complex amplitudes generated by the STDREC. The hybrid technique that results from interconnecting the STDREC algorithm with the above mentioned algorithm is dubbed *hybrid* STDREC (HSTDREC) in the following; note that this algorithm represents an instance of the hybrid CSFE proposed in [65].

T3-S1 – This step is the most complicated of the whole algorithm and deserves the following comments:

a) In principle, the horizontal and vertical spatial frequencies (see (2.11) and (2.12)) of multiple targets contributing to the $\hat{\alpha}_l$-th frequency bin can be detected by first computing a 2D DFT of the matrix $\mathbf{X}[l]$ (2.36) and, then, by looking for local maxima over the absolute values of the elements of the resulting 2D matrix; note that the matrix $\mathbf{X}[l]$ can be also zero-padded before computing its 2D FFT to improve the resulting spectral resolution. This procedure may require a significant computational effort and its accuracy is affected by the spectral leakage due to any potential strong target. In the STDAEC algorithm, instead, 2D processing is avoided by alternating vertical and horizontal 1D FFTs. Consequently, relevant spatial frequencies are estimated by searching for the peaks of 1D amplitude spectra (i.e., in the absolute values of the elements of the vectors $\mathbf{S}^{(i)}_{\mathrm{VULA},0}[l]$ and $\mathbf{X}^{(\mathrm{VF})}_i[l]$); in other words, an AM approach is adopted. Note that this approach allow us to mitigate the overall computational complexity and to detect weak targets hidden by close strong targets through successive cancellations.

b) In **STDAE-S1**, each of the three vectors $\{\mathbf{s}^{(i)}_{\mathrm{VULA},k}[l];\, k = 0,\, 1,\, 2\}$ is computed by executing a $\bar{N}_0$ order FFT (see (2.48)). Note that, on the one hand, the vector $\mathbf{s}^{(i)}_{\mathrm{VULA},0}[l]$ collects $\bar{N}_0$ equally spaced samples of the spectrum of the sequence $\{X^{(i)}_{\mathrm{VULA},p};\, p = 0,\, 1,\, ...,\, N_{\mathrm{VULA}} - 1\}$ (see (2.46)). On the other hand, the k-th vector $\mathbf{s}^{(i)}_{\mathrm{VULA},k}[l]$ (with $k = 1$ and 2) collects, up to a scale factor, N_0 equally spaced samples of the k-th order derivative of the same spectrum.

c) The processing accomplished in **STDAE-S3** is very similar to that carried out in **STDAE-S1**. In fact, the only difference is represented by the fact that the N_{VULA}-dimensional vector $\mathbf{S}^{(i)}_{\mathrm{VULA},0}[l]$ (2.44) is replaced by the N_{HULA}-dimensional

vector $\mathbf{X}_i^{(\mathrm{VF})}[l]$ (2.52) generated in **STDAE-S2**. Therefore, in this case, the CSFE is exploited to estimate the horizontal frequency $F_{\mathrm{H},i}[l]$ and, again, the complex amplitude $C_i[l]$ ($A_i[l]$) associated with the i-th target.

d) Similarly as the STDREC algorithm, the STDAEC algorithm can also be considered as an instance of the CSFEC algorithm mentioned at point g) of **T2-S2**. Therefore, in principle, it can be replaced by the CFH algorithm [70], the Alg-YA [71, 72] or the Alg-S [73]. Moreover, a further (and final) step, based the Alg-YA can be added to the STDAEC algorithm to mitigate its estimation bias.

e) As already suggested for the STDREC algorithm, the STDAEC algorithm can be employed to detect and estimate multiple angles in parallel; this requires running multiple instances of the CSFE algorithm in parallel.

Our final comments concern the use of RASCA-FR3 and RASCA-FC3 in FMCW radar systems whose virtual antenna array is not an URA; for instance, in our experimental work (see Section 2.8), a colocated MIMO FMCW radar equipped with the virtual receive array shown in Fig. 2.6 has been employed. Note that the first two processing tasks of the RASCAs are carried out on an antenna-by-antenna basis; therefore, they are not influenced by the shape of the considered virtual array. However, this shape influences the way spatial folding is accomplished in **T3**. More specifically, as far as the last point is concerned, the following considerations can be made:

1) The array structure represented in Fig. 2.6 can be treated as an URA if its gaps are zero-padded.

2) The reference VULA should be selected in a way to maximize the number of non-zero vertically aligned VAs and, consequently, the number of VAs contributing to the estimation of the elevation angle, as illustrated in Fig. 2.6.

3) The reference HULA should be selected in the middle of the antenna array; this mitigates the effects of the errors affecting the estimate of normalised vertical frequency in the vertical folding procedure (do not forget that such errors may have a significant impact on the contributions of the HULAs that are farther from the reference HULA; see (2.50)).

4) The vertical folding accomplished by the STDAE algorithm involves VULAs of different sizes. More specifically, in the i-th iteration of the STDAEC algorithm, vertical folding is accomplished by computing the N_{HULA}-dimensional vector (see (2.37) and **STDAE-S2**)

$$\mathbf{X}_i^{(\mathrm{VF})}[l] = \left[X_i^{(\mathrm{VF})}[p_I, l], X_l^{(i)}[p_I+1, l], ..., X_l^{(i)}[p_F, l]\right]^T \tag{2.62}$$

where

$$\begin{aligned} X_i^{(\mathrm{VF})}[p, l] =& \frac{1}{N_V[p]} \sum_{\substack{q=q_I[p] \\ q \neq q_R}}^{q_F[p]} X_l^{(i)}[p, q] \; R_i^{(\mathrm{VF})}[l, q] \\ &+ \frac{X_l^{(i)}[p, q_R]}{N_V[p]}, \end{aligned} \tag{2.63}$$

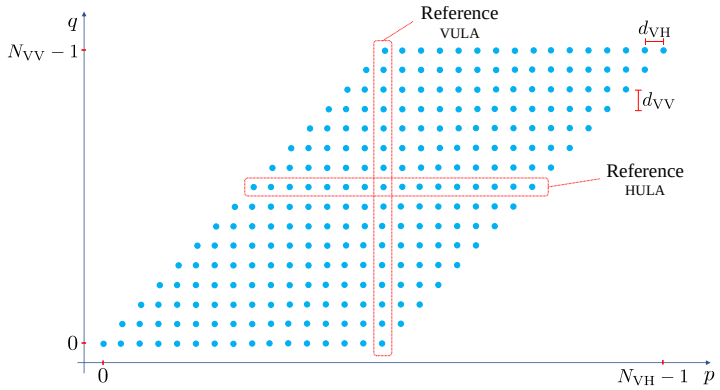


Figure 2.6: *Virtual array considered in our experimental work.*

with $p = p_I, p_I + 1, ..., p_F$; here, $R_i^{(\mathrm{VF})}[l, q]$ is expressed by (2.50), $q_I[p]$ ($q_F[p]$) is the index identifying the first (last) antenna of the p-th VULA and $N_{\mathrm{V}}[p]$ is the overall number of VAs which that VULA consists of.

2.3.4 Range & angle serial cancellation algorithms for a radar system endowed with a one-dimensional antenna array

The algorithms described in the previous paragraph can be easily adapted to the case in which the considered colocated MIMO radar system is equipped with a single ULA and, consequently, can be exploited for 2D imaging only; this leads to RASCA-FR2 and RASCA-FC2. The changes made in RASCA-FR3 and RASCA-FC3 to obtain RASCA-FR2 and RASCA-FC2, respectively, concern only the SPE and can be summarized as follows:

1. The first three steps of the STDAE in **T3-S1** are not performed; therefore, the fourth step of that algorithm is the first one to be executed. Moreover, the matrix $\mathbf{X}^{(i)}[l]$ (2.37) is replaced by the N_{VH}-dimensional vector

$$\mathbf{X}^{(i)}[l] \triangleq \left[X_l^{(i)}[p]\right], \tag{2.64}$$

   collecting the spectral information available on the whole virtual receive array and referring to the $\hat{\alpha}_l$-th frequency bin only.

2. The spatial frequency $\hat{F}_{\mathrm{V},i}[l]$ is unavailable and, therefore, it is not included in the set $\mathcal{T}_l$ (2.8); note that the elevation angle $\hat{\phi}_i[l]$ (2.41) is not estimated in this case.

2.4 Description of Various Algorithms Employed in the Proposed Embodiments

In this section, various mathematical details about the techniques employed in the RASCAs are provided.

Since the processing accomplished in **T1** of the RASCAs has been fully analysed in the previous section, in this paragraph we provide a detailed description of: a) the SFE (see **T2-S2**); b) the CSFE (see **T2-S2** and **T3-S1**); c) the target cancellation procedures employed in **T2-S3** and **T3-S1**.

2.4.1 Single frequency estimator

In this Paragraph, the SFE derived in [65] is summarized. This algorithm processes the samples of the real sequence $\{x_{r,n};\, n = 0,\, 1,\, ...,\, N-1\}$, whose n-th element is

$$\begin{aligned} x_{r,n} &= a\cos\left(2\pi n\, F\, + \psi\right) + w_{r,n} \\ &= C\exp\left(j2\pi n\, F\right) + C^{*}\exp\left(-j2\pi n\, F\right) + w_{r,n}, \end{aligned} \tag{2.65}$$

with $n = 0,\, 1,\, ...,\, N-1$, and generates an estimate of the *normalised frequency* F and of the *complex amplitude*

$$C \triangleq \frac{1}{2} a\exp\left(j\psi\right) \tag{2.66}$$

of the real tone appearing in the RHS of (2.65); here, N is the overall number of elements of the sequence $\{x_{r,n}\}$, a and ψ are the tone amplitude and phase, respectively, and $\{w_{r,n};\, n = 0,\, 1,\, ...,\, N-1\}$ is a real AWGN sequence. This algorithm is initialised by

1) Evaluating: a) the vector

$$\mathbf{X}_0 = \left[X_{0,0}, X_{0,1}^{(v)}, ..., X_{0,N_0-1}^{(v)}\right]^T \triangleq \mathrm{DFT}_{N_0}\left[\mathbf{x}_{0,\mathrm{ZP}}^{(v)}\right], \tag{2.67}$$

where the DFT order N_0 is defined by (2.22),

$$\mathbf{x}_{0,\mathrm{ZP}} \triangleq \left[\left(\mathbf{x}_0\right)^T\ \mathbf{0}_{(M-1)N}^T\right]^T, \tag{2.68}$$

M is the *oversampling factor* and

$$\mathbf{x}_0 \triangleq \left[x_{r,0}, x_{r,1}, ..., x_{r,N-1}\right]^T; \tag{2.69}$$

b) the initial coarse estimate $\hat{F}_c^{(0)}$ of F as

$$\hat{F}_c^{(0)} = \hat{\alpha}\, F_{\mathrm{DFT}}, \tag{2.70}$$

where the integer $\hat{\alpha}$ is computed as

$$\hat{\alpha} = \arg \max_{\tilde{\alpha}\in\{0,1,...,N_0/2-1\}} \left|\bar{X}_{0,\tilde{\alpha}}\right|; \tag{2.71}$$

c) the quantity

$$\hat{\rho}^{(0)} \triangleq \frac{\hat{F}_c^{(0)}}{F_{\mathrm{DFT}}} = \hat{\alpha}; \tag{2.72}$$

d) the initial estimate $\hat{C}^{(0)}$ of C as

$$\hat{C}^{(0)} = G\big(\hat{F}_c^{(0)}\big) \tag{2.73}$$

where

$$G\big(\tilde{F}\big) \triangleq \frac{\bar{X}(\tilde{F}) - \bar{X}^*(\tilde{F})g(\tilde{F})}{1 - \left|g(\tilde{F})\right|^2}, \tag{2.74}$$

$$\bar{X}(\tilde{F}) \triangleq \frac{1}{N}\sum_{n=0}^{N-1} x_{r,n} \exp\left(-j2\pi n\tilde{F}\right) \tag{2.75}$$

and

$$g(\tilde{F}) \triangleq \frac{1}{N}\sum_{n=0}^{N-1} \exp\left(-j4\pi n\tilde{F}\right); \tag{2.76}$$

e) the spectral coefficients $\bar{X}_{1,\hat{\alpha}}$ and $\bar{X}_{2,\hat{\alpha}}$, and the coefficients $\{K_p(2\hat{\alpha});\, p = 1,\, 2\}$ and $\{b(\hat{\alpha}),\, c(\hat{\alpha})\}$ on the basis of the definitions

$$\bar{X}_{k,\rho} \triangleq \frac{1}{N}\sum_{n=0}^{N-1} x_{k,n} \exp\left(-j\frac{2\pi n\rho}{N_0}\right), \tag{2.77}$$

$$K_p\left(x\right) \triangleq \frac{1}{N}\sum_{n=0}^{N-1} g_p\left[n\right] \exp\left(-j\frac{2\pi nx}{N_0}\right), \tag{2.78}$$

$$b\left(\rho\right) \triangleq -\Re\left\{\hat{C}^*\bar{X}_{2,\rho}\right\} + 2\Re\left\{\left(\hat{C}^*\right)^2 K_2\left(2\rho\right)\right\} \tag{2.79}$$

and

$$c\left(\rho\right) \triangleq \Im\left\{\hat{C}^*\bar{X}_{1,\rho}\right\} - \Im\left\{\left(\hat{C}^*\right)^2 K_1\left(2\rho\right)\right\}, \tag{2.80}$$

respectively; f) the initial estimate $\hat{\Delta}^{(0)}$ of Δ as

$$\hat{\Delta}^{(0)} = P\big(\hat{\rho}^{(0)}\big), \tag{2.81}$$

where

$$P\left(\tilde{\rho}\right) \triangleq -c\left(\tilde{\rho}\right)/b\left(\tilde{\rho}\right); \tag{2.82}$$

g) the first fine estimate $\hat{F}^{(0)}$ of F as

$$\hat{F}^{(0)} = \hat{F}_c^{(0)} + \frac{\hat{\Delta}^{(0)}}{2\pi} \tag{2.83}$$

2) Setting its iteration index i to 1.

Then, an iterative procedure is started. The i-th iteration is fed by the estimates $\hat{F}^{(i-1)}$ and $\hat{C}^{(i-1)}$ of F and C, respectively, and produces the new estimates $\hat{F}^{(i)}$ and $\hat{C}^{(i)}$ of the same quantities (with $i = 1,\, 2,\, ...,\, N_{\mathrm{SFE}}$, where N_{SFE} represents the overall number of iterations); the procedure employed for the evaluation of $\hat{F}^{(i)}$ and $\hat{C}^{(i)}$ consists of the two steps described below (the p-th step is denoted SFE-Sp).

SFE-S1) - The new estimate $\hat{\Delta}^{(i)}$ of Δ is computed as[7] (see (2.81)–(2.82))

$$\hat{\Delta}^{(i)} = P(\hat{\rho}^{(i-1)}) = -c\big(\hat{\rho}^{(i-1)}\big)/b\big(\hat{\rho}^{(i-1)}\big); \tag{2.84}$$

[7]The quantities $\{\bar{X}_{k,\rho};\, k = 1, 2\}$ required in the computation of the coefficients $b(\rho)$ and $c(\rho)$ can be also evaluated by means of the interpolation-based method illustrated in [65, Sect. III, p. 12]. In our work, *barycentric interpolation* has been always used [74]; in the following, the parameter I represents the interpolation order. These considerations hold also for the CSFE described below.

in the evaluation of the coefficients $\{b(\rho),\, c(\rho)\}$ appearing in the RHS of (2.82), $\hat{C} = \hat{C}^{(i-1)}$ and

$$\rho = \hat{\rho}^{(i-1)} \triangleq \hat{F}^{(i-1)}/F_{\mathrm{DFT}} \tag{2.85}$$

are assumed. Then,

$$\hat{F}^{(i)} = \hat{F}^{(i-1)} + \hat{\Delta}^{(i)}/(2\pi) \tag{2.86}$$

is evaluated.

SFE-S2) - The new estimate $\hat{C}^{(i)}$ of $\hat{C}$ is evaluated as $\hat{C}^{(i)} = G(\hat{F}^{(i)})$ (see (2.73)–(2.74)). Moreover, the index i is incremented by one before starting the next iteration.

At the end of the last (i.e., of the N_{SFE}-th) iteration, the fine estimates $\hat{F} = \hat{F}^{(N_{\mathrm{SFE}})}$ and $\hat{C} = \hat{C}^{(N_{\mathrm{SFE}})}$ of F and C, respectively, become available.

2.4.2 Complex single frequency estimator

All the results illustrated in the previous paragraph refer to the real sequence $\{x_{r,n}\}$, whose n-th element is expressed by (2.65). However, a similar estimation method (namely, the CSFE) has been developed for the complex counterpart, i.e. for a complex sequence $\{x_{c,n};\, n = 0,\, 1,\, ...,\, N-1\}$, whose n-th element is

$$x_{c,n} = A \exp\left(j2\pi n\, F\right) + w_{c,n}, \tag{2.87}$$

with $n = 0,\, 1,\, ...,\, N-1$. Here, A is the complex amplitude of the single tone appearing in the RHS of the last equation, $\{w_{c,n};\, n = 0,\, 1,\, ...,\, N-1\}$ is a complex AWGN sequence and all the parameters have exactly the same meaning as that illustrated for (2.65). The description of the CSFE is similar to that illustrated for the SFE in the previous paragraph, the only differences being represented by the fact that: a) the parameter C (2.66) is replaced by A; b) (2.73) is replaced by

$$\tilde{A} = \hat{A} = \bar{X}(\hat{F}), \tag{2.88}$$

where $\bar{X}(\hat{F})$ is computed according to (2.75) (in which $x_{r,n}$ is replaced by $x_{c,n}$ (2.87)); c) (2.79) and (2.80) are replaced by

$$b\,(\rho) \triangleq \Re\{\hat{A}^* \bar{X}_{2,\rho}\} \tag{2.89}$$

and

$$c\,(\rho) \triangleq -\Im\{\hat{A}^* \bar{X}_{1,\rho}\}, \tag{2.90}$$

respectively.

2.4.3 Target cancellation procedures employed in FMCW radar systems

In **T2** of the RASCA-FR2 and RASCA-FR3 (and, in particular, in **STDREC-S2**; see (2.29)), a target cancellation procedure is used in combination with the SFE. This procedure requires the evaluation of the triad $(\mathbf{C}_{X_0}^{(v)}[i],\ \mathbf{C}_{X_1}^{(v)}[i],\ \mathbf{C}_{X_2}^{(v)}[i])$, that represents the contribution given by the i-th (i.e., by the last) *point target* detected on the v-th

VA. If $\hat{F}_i^{(v)}$ and $\hat{C}_i^{(v)}$ denote the estimates of the normalised frequency and the complex amplitude, respectively, characterizing this target, the expressions

$$\mathbf{C}_{X_0}^{(v)}[i] = \hat{C}_i^{(v)}\ \bar{\mathbf{W}}_0^{(v)}[i] + \left(\hat{C}_i^{(v)}\right)^* \left(\bar{\mathbf{W}}_{0,c}^{(v)}[i]\right), \tag{2.91}$$

$$\mathbf{C}_{X_1}^{(v)}[i] = \hat{C}_i^{(v)}\ \bar{\mathbf{W}}_1^{(v)}[i] + \left(\hat{C}_i^{(v)}\right)^* \left(\bar{\mathbf{W}}_{1,c}^{(v)}[i]\right) \tag{2.92}$$

and

$$\mathbf{C}_{X_2}^{(v)}[i] = \hat{C}_i^{(v)}\ \ \bar{\mathbf{W}}_2^{(v)}[i] + \left(\hat{C}_i^{(v)}\right)^* \left(\bar{\mathbf{W}}_{2,c}^{(v)}[i]\right) \tag{2.93}$$

are employed; here, $\bar{\mathbf{W}}_k^{(v)}[i]$ denotes the N_0 order DFT of the vector

$$\begin{aligned}\bar{\mathbf{w}}_k^{(v)}[i] \triangleq \Bigg[0, 1^k \cdot \bar{w}_i^{(v)}, 2^k \cdot \left(\bar{w}_i^{(v)}\right)^2, \ldots, \\ (N-1)^k \cdot \left(\bar{w}_i^{(v)}\right)^{N-1}, 0, \ldots, 0\Bigg]^T,\end{aligned} \tag{2.94}$$

with $k = 0$, 1 and 2, $\bar{\mathbf{W}}_{k,c}^{(v)}[i]$ the N_0 order DFT of the vector $(\bar{\mathbf{w}}_k^{(v)}[i])^*$,

$$\bar{w}_i^{(v)} \triangleq \exp(j2\pi\bar{F}_i^{(v)}) \tag{2.95}$$

and

$$\bar{F}_i^{(v)} \triangleq \hat{f}_i^{(v)} T_s \tag{2.96}$$

is the normalised frequency associated with the frequency $\hat{f}_v^{(i)}$. It is important to point out that an efficient method can be used for the computation of the vectors $\bar{\mathbf{W}}_k^{(v)}[i]$ and $\bar{\mathbf{W}}_{k,c}^{(v)}[i]$ appearing in the RHS of (2.91)–(2.93) (with $k = 0$, 1 and 2); note that, for any k, these vectors represent the N_0 order DFTs of the sequences $\{n^k\ (\bar{w}_i^{(v)})^n;\ n = 0,\ 1,\ \ldots,\ N-1\}$ and $\{n^k\ ((\bar{w}_i^{(v)})^*)^n;\ n = 0,\ 1,\ \ldots,\ N-1\}$, respectively. In fact, the l-th element of the vectors $\bar{\mathbf{W}}_k^{(v)}[i]$ and $\bar{\mathbf{W}}_{k,c}^{(v)}[i]$ is given by

$$\begin{aligned}\bar{W}_k^{(v)}[i,l] =& \frac{1}{N}\sum_{n=0}^{N-1} n^k \left(\bar{w}_i^{(v)}\right)^n \exp\left(-j\frac{2\pi l}{N_0}n\right) \\ =& \frac{1}{N}\sum_{n=0}^{N-1} n^k (q[l])^n\end{aligned} \tag{2.97}$$

and

$$\begin{aligned}\bar{W}_{k,c}^{(v)}[i,l] =& \frac{1}{N}\sum_{n=0}^{N-1} n^k \left(\left(\bar{w}_i^{(v)}\right)^*\right)^n \exp\left(-j\frac{2\pi l}{N_0}n\right) \\ =& \frac{1}{N}\sum_{n=0}^{N-1} n^k (q_c[l])^n,\end{aligned} \tag{2.98}$$

respectively, where

$$q[l] \triangleq \exp\left(j2\pi\left(\bar{F}_i^{(v)} - \frac{l}{N_0}\right)\right) \tag{2.99}$$

and

$$q_c[l] \triangleq \exp\left(j2\pi\left(-\bar{F}_i^{(v)} - \frac{l}{N_0}\right)\right). \tag{2.100}$$

Therefore, the identities listed in [65, eqs. (84)-(85) and (145)] can be exploited for an efficient computation of the RHSs of (2.97) and (2.98).

A target cancellation procedure is also employed in **T2** of the RASCA-FC2 and RASCA-FC3; however, in this case, the CSFE is adopted in place of the SFE, and the vectors $\mathbf{C}_{X_0}^{(v)}[i]$, $\mathbf{C}_{X_1}^{(v)}[i]$ and $\mathbf{C}_{X_2}^{(v)}[i]$ are evaluated as

$$\mathbf{C}_{X_0}^{(v)}[i] = \hat{A}_i^{(v)}\ \bar{\mathbf{W}}_0^{(v)}[i], \tag{2.101}$$

$$\mathbf{C}_{X_1}^{(v)}[i] = \hat{A}_i^{(v)}\ \bar{\mathbf{W}}_1^{(v)}[i] \tag{2.102}$$

and

$$\mathbf{C}_{X_2}^{(v)}[i] = \hat{A}_i^{(v)}\ \bar{\mathbf{W}}_2^{(v)}[i]; \tag{2.103}$$

respectively; here, $\bar{\mathbf{W}}_k^{(v)}[i]$ denotes the N_0 order DFT of the vector

$$\bar{\mathbf{w}}_k^{(v)}[i] \triangleq \left[0, 1^k \cdot \bar{w}_i^{(v)}, 2^k \cdot \left(\bar{w}_i^{(v)}\right)^2, \ldots, (N-1)^k \cdot \left(\bar{w}_i^{(v)}\right)^{N-1}, 0, \ldots, 0\right]^T, \tag{2.104}$$

with $k = 0$, 1 and 2, and $\bar{w}_i^{(v)}$ is still expressed by (2.95). The vector $\bar{\mathbf{W}}_k^{(v)}[i]$ appearing in (2.91)–(2.93) (with $k = 0$, 1 and 2) can be efficiently computed following the same approach illustrated above for the SFE.

The CSFE is also employed in **T3-S1** and, in particular, in **STDAEC-S2** of the RASCA-FR2, RASCA-FR3, RASCA-FC2 and RASCA-FC3. In this case, the cancellation procedure requires the evaluation of the contribution

$$\mathbf{C}_{X_0}^{(i)}[l] = \left[C_{X_0}^{(i)}[p, q, l]\right] \tag{2.105}$$

given by the i-th (i.e., by the last) target detected in the l-th frequency bin to the whole array (see (2.38)). Here, we focus on the target cancellation procedure employed in the above mentioned RASCAs. In this case, if $\hat{A}_i[l]$, $\hat{F}_{\mathrm{V},i}[l]$ and $\hat{F}_{\mathrm{H},i}[l]$ denote the estimates of the complex amplitude, the normalised vertical spatial frequency and the normalised horizontal spatial frequency, respectively, characterizing the i-th target, the expression

$$C_{X_0}^{(i)}[p, q, l] = \hat{A}_i[l] \exp\left\{j2\pi\left[(p - p_R)\hat{F}_{\mathrm{H},i}[l] + (q - q_R)\hat{F}_{\mathrm{V},i}[l]\right]\right\}, \tag{2.106}$$

is employed for any VA (i.e., for any p and q).

Finally, it is important to mention that the cancellation procedure adopted in STDREC algorithm aims at removing the contribution of a single target in each of its iterations. If a *cluster* of $m_i^{(v)}$ distinct frequencies is estimated by the SFE (CSFE) in the i-th iteration of the above mentioned algorithm, each of the components of the triad ($\mathbf{C}_{X_0}^{(v)}[i]$, $\mathbf{C}_{X_1}^{(v)}[i]$, $\mathbf{C}_{X_2}^{(v)}[i]$) consists of the sum of $m_i^{(v)}$ terms and each term is evaluated on the basis of (2.91)–(2.93) ((2.101)–(2.103)).

2.5 Limitations

In this section, some technical limitations that have emerged in the implementation of our algorithms on commercial radar devices are illustrated and the solutions we have devised to mitigate their impact are described.

2.5.1 Unequal response of virtual antennas

The derivation of the RASCAs for FMCW radar systems relies on the assumption that the real (complex) sample sequence made available by the v-th VA is expressed by (1.9) ((1.11)). The adopted signal models hold if the amplitudes of the L overlapped oscillations contributing to the useful component of the received signal do not change from antenna to antenna. However, our experiments accomplished on commercial colocated radar devices have evidenced that: a) these amplitudes are not constant across the whole virtual array; b) their differences are influenced by the azimuth and the elevation of each target. We believe that all this is due to the different behavior of the multiple receive chains employed in each MIMO device and to the mismatches in the receive antenna patterns, together with electromagnetic interference patterns due to the phase coherency of the radiated waves. This problem affects the quality of the data acquired through FMCW radar systems and, consequently, the accuracy of our detection and estimation algorithms. It can be mitigated by enriching the physical array with a set of surrounding *passive antennas*; in this case, the array is artificially extended with new antennas along all its sides, so that the behavior of all its active antennas becomes more uniform.

It important to point out that, in principle, the presence of this phenomenon can be accounted for in the development of target detection and estimation algorithms by including its effects in the received signal model. For instance, (1.9) can be generalised as

$$x_{r,n}^{(v)} = \sum_{l=0}^{L-1} \alpha_v\left(\theta_l, \phi_l\right)\, a_l\, \cos\left(2\pi n F_l^{(v)} + \psi_l^{(v)}\right) + w_{r,n}^{(v)}, \tag{2.107}$$

where $\alpha_v(\theta_l, \phi_l)$ represents an attenuation factor depending on the angular coordinates of the l-th target and v is the VA index. Consequently, the complex amplitude associated with the l-th target detectable on the considered VA becomes (see (1.10))

$$C_l^{(v)}\left(\theta_l, \phi_l\right) \triangleq \frac{1}{2} a_l\, \alpha_v\left(\theta_l, \phi_l\right) \exp\left(j\, \psi_l^{(v)}\right). \tag{2.108}$$

Neglecting the presence of the factor $\alpha_v\left(\theta_l,\ \phi_l\right)$ in the development of our algorithms has the following implication: an error is introduced by the STDAEC algorithm in its cancellation procedure (see **STDAEC-S2** in Paragraph 2.4). Note, in particular, that the estimate $\hat{C}_i[l]$ of the complex amplitude characterizing to the i-th target detected in the $\hat{\alpha}_l$-th bin is computed *after* the overall spatial folding (i.e., after **STDAE-S5**); consequently, its absolute value represents a sort of *spatial average* computed over all the involved VAs. Moreover, only the phase variations of this complex gain are accounted for in the computation of the contribution $\mathbf{C}_{X_0}^{(i)}[l]$ of this target to the matrix $\mathbf{X}^{(i)}[l]$ (see (2.105)–(2.106)). Note that, if the functions $\{\alpha_v(\theta_l, \phi_l)\}$ were known for all the

VAs, their effect could be compensated for *after* evaluating the estimates $(\hat{\theta}_i, \hat{\phi}_i)$ of the angular coordinates of the i-th target; in fact, this result could be achieved by replacing the estimate $\hat{C}_i[l]$ of the complex gain $C_i[l]$ with

$$\hat{C}_i[v, l] \triangleq \hat{C}_i[l]\ \alpha_v(\hat{\theta}_i, \hat{\phi}_i). \tag{2.109}$$

in the evaluation of the term $\mathbf{C}_{X_0}^{(i)}[l]$ appearing in (2.105)–(2.106). Estimating the function $\alpha_v(\theta, \phi)$, however, is a time consuming task, since it requires a proper measurement setup and an anechoic chamber. We believe that this problem can be circumvented by: a) exploiting *deep learning techniques* [75] in the SPE; b) adopting a *data-driven approach* [76], [77]. This solution is motivated by the fact that:

a) Deep learning techniques can be employed to approximate complicated functions, that do not lend themselves to a simple parametric representation and without requiring particular expertise in data pre-processing.

b) A data-driven approach allows to train different models on the basis of data collected in a real scenario or synthetically generated data, without prior knowledge about the parametric representation of the considered problem. Note that a fundamental role is played by the adopted training procedure since it makes the involved network able to generate correct predictions on the basis of never seen data available at its input.

In practice, the adoption of the proposed approach requires modifying the STDAEC technique employed in the RASCAs (see Fig. 2.3) and, in particular, embedding a deep neural network in it. This network is employed to estimate the distorted amplitudes of all the targets detected in the l-th frequency bin (with $l = 0, 1, \ldots, L_b - 1$), so that accurate cancellation becomes possible.

The use of this solution in our radar systems is not investigated in the following, since it is out of the scope of this chapter.

2.5.2 Antenna coupling

In our description of the SFE and the CSFE (see Section 2.4), it has been implicitly assumed that the minimum frequency of the useful component contained in the observed data sequence can be arbitrarily small. Unluckily, this is not always true. For instance, in commercial colocated FMCW MIMO radar systems, a strong interference is observed in the lower portion of the spectrum evaluated on all the receive antennas. This phenomenon, known as *mutual coupling* [78], is due to the electromagnetic coupling that originates from the small distance between adjacent transmit and receive antennas [54]. Its impact can be mitigated resorting to various methods based on calibration measurements [79]. Because of mutual coupling, any target whose range is below a certain threshold cannot be detected by our algorithms in a reliable fashion.

2.6 Other target detection and estimation techniques

The detection and estimation algorithms described above have been compared, in terms of accuracy and complexity, with two different types of algorithms that, similarly as the RASCAs, are able to generate radar images in the form of point clouds. The algorithms of the first type are called FFT-*based algorithms* (FFT-BAs), since they rely

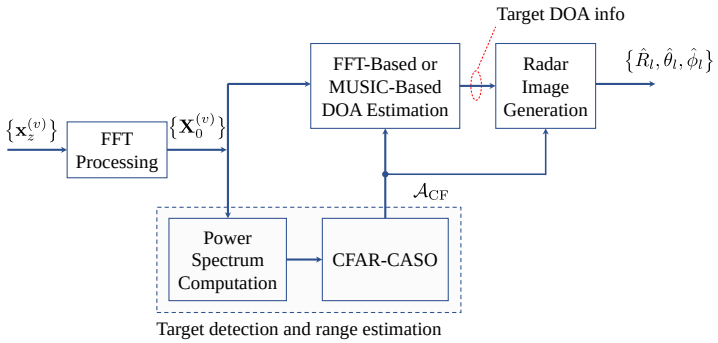


Figure 2.7: *Block diagram describing the overall structure of the FFT-BAs and the MUSIC-BAs.*

on multidimensional FFT processing for the evaluation of all the spatial coordinates of targets (i.e., their range and DOA); such algorithms have been inspired by the FFT-based algorithm proposed by Texas Instrument in [61]. The algorithms of the second type, instead, are called MUSIC-*based algorithms* (MUSIC-BAs); these make use of the same method as the first type for range estimation, but the MUSIC algorithm for DOA estimation [80]. In the remaining part of this section, a brief description is provided for both types. In our analysis, we always refer to a FMCW radar system, since, if a SFCW radar system is considered, the only change to be made in the described algorithms consists in replacing each FFT with an IFFT of the same order.

The inner structure of both types of algorithms is described by the block diagram shown in Fig. 2.7. The processing accomplished by the blocks this diagram consists of, can be summarized as follows. Each vector of the set $\{\mathbf{x}_z^{(v)}\}$, collecting N_{VR} vectors (see (1.17)), undergoes, after ZP, a N_0 order FFT; this produces a set of N_0-dimensional vectors $\{\mathbf{X}_0^{(v)}\}$ (see (2.19), (2.23) and (2.24)). Based on this set of vectors, the N_0-dimensional power spectrum

$$\mathbf{P}_0 = [P_{0,0}, P_{0,1}, ..., P_{0,N_0-1}]^T \tag{2.110}$$

is computed; here,

$$P_{0,i} \triangleq \frac{1}{N_{\mathrm{VR}}} \sum_{v=0}^{N_{\mathrm{VR}}-1} \left(X_{0,i}^{(v)}\right)^2 \tag{2.111}$$

with $i = 0, 1, ..., N_0 - 1$. The vector $\mathbf{P}_0$ (2.110) feeds the *cell-averaging smallest of - constant false alarm rate* (CFAR-CASO) algorithm developed in [81]. Based on this algorithm, a target is detected in the i-th frequency bin if

$$P_{0,i} > T_{\mathrm{CFAR}}, \tag{2.112}$$

where $i \in \{i_m, i_m + 1, ..., i_M\}$. Here,

$$T_{\mathrm{CFAR}} = K_0 \min\left(\bar{P}_l\,,\, \bar{P}_u\right) \tag{2.113}$$

represents a decision threshold, K_0 is a real parameter whose value is selected on the

basis of the required *false alarm rate*, and

$$\bar{P}_l = \frac{1}{C_s} \sum_{k=i-(G_s+C_s)}^{i-(G_s+1)} P_{0,k} \tag{2.114}$$

and

$$\bar{P}_u = \frac{1}{C_s} \sum_{k=i+G_s+1}^{i+G_s+C_s} P_{0,k} \tag{2.115}$$

represent the average of the power spectrum computed over C_s adjacent bins positioned on the left and on the right, respectively, with respect to the i-th frequency bin. Moreover, G_s and C_s are two integer parameters defining the size and the position (with respect to the i-th bin), respectively, of the set of frequency bins involved in the computation of $\bar{P}_l$ (2.114) and $\bar{P}_u$ (2.115), whereas i_m and i_M are two non negative integers such that $i_m \geq (G_s + C_s)$ and $i_M \leq N_0 - 1 - (G_s + C_s)$.

In our work, the inequality

$$P_{0,i} > P_{l,u} \tag{2.116}$$

is also required to be satisfied together with the condition (2.112), where $P_{l,u}$ represents the largest element of the set $\{P_{0,i+l};\, l = -(G_s+C_s),\, -(G_s+C_s)+1,\, -G_s-1,\, G_s+1,\, G_s+C_s\}$. This allows us to reduce the overall number of detected targets, so reducing the density of the generated point cloud.

The CFAR-CASO algorithm generates the vector

$$\mathscr{A}_{\mathrm{CF}} = [\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_{L_b-1}]^T \tag{2.117}$$

where $\hat{\alpha}_l$ represents the index of the frequency bin in which the l-th target has been detected (with $l = 0,\, 1,\, \ldots,\, L_b - 1$) and L_b is the overall number of detected targets. This vector is processed for DOA estimation. The two options (associated with the above mentioned types of algorithms) are considered for this task and are described in the remaining part of this paragraph.

FFT-*based* DOA *estimation* – Let us focus first on the case in which a virtual HULA, consisting of N_{VH} virtual elements, is employed for resolving the targets associated with a given frequency bin and estimating their azimuth. In this case, azimuth estimation consists of the following two steps:

1) The N_{VH}-dimensional column vector (see (2.64))

$$\mathbf{X}\,[l] \triangleq \left[X_{0,\hat{\alpha}_l}^{(0)}, X_{0,\hat{\alpha}_l}^{(1)}, ..., X_{0,\hat{\alpha}_l}^{(N_{\mathrm{VH}}-1)}\right]^T, \tag{2.118}$$

collecting the spectral information available on the whole array and referring to the $\hat{\alpha}_l$-th frequency bin (with $l = 0,\, 1,\, ...,\, L_b - 1$) is applied to an $\bar{N}_0$ order FFT algorithm; let $\mathbf{s}[l] = [s_0[l],\, s_1[l],\, ...,\, s_{\bar{N}_0-1}[l]]^T$ denote the $\bar{N}_0$-dimensional FFT output.

2) The dominant peaks[8] in the sequence $\{|s_k[l]|;\, k = 0, 1, ..., \bar{N}_0 - 1\}$ are identified; each peak corresponds to a distinct target. If $k_i[l]$ denotes the index of i-th peak (with

[8] It is important to distinguish peaks associated with different targets from side-lobes; in our simulations, a candidate peak is classified as a side-lobe (and, consequently, ignored) if its amplitude differs by more than 1 dB from that of a close dominant peak, as suggested in [61].

$i = 0, 1, ..., L_h[l] - 1$, where $L_h[l]$ is the overall number of targets detected in the considered frequency bin), the estimate of the azimuth of the i-th target is evaluated as

$$\hat{\theta}_i [l] = \arcsin \left(h_{\bar{N}_0} [k_i [l]] \right) \tag{2.119}$$

where

$$h_{\bar{N}_0} [x] \triangleq 2 \left(x - \bar{N}_0/2 \right) / \bar{N}_0 \, . \tag{2.120}$$

Let us now consider the case in which the URA represented in Fig. 2.2 is employed for resolving the targets associated with each frequency bin, and estimating their azimuth and elevation. The algorithm employed in this case involves the $N_{\mathrm{VH}} \times N_{\mathrm{VV}}$ matrix $\mathbf{X}[l] \triangleq [X_{0,\hat{\alpha}_l}[p,q]]$ (2.36), collecting the spectral information available on the whole array for the $\hat{\alpha}_l$-th frequency bin. This algorithm consists of the following four steps:

1) The p_R-th row of the matrix $\mathbf{X}[l]$ is processed to generate the $\bar{N}_0$-dimensional column vector $\mathbf{s}_{\mathrm{VULA},0}[l] = [s_{0,0} [l] , s_{0,1} [l] , ..., s_{0,\bar{N}_0-1} [l]]^T$ on the basis of (2.44); here, p_R represents the column index of the reference antenna in the considered URA (see Fig. 2.2).

2) The dominant peaks of the sequence $\{[|s_{0,r} [l] |; r = 0, 1, ..., \bar{N}_0 - 1\}$ are identified. If $r_i[l]$ denotes the index of i-th peak (with $i = 0, 1, ..., L_v[l] - 1$, where $L_v[l]$ is the overall number of targets detected in the considered frequency bin), the estimate of the elevation $\hat{\phi}_i[l]$ of the associated target is evaluated as

$$\hat{\phi}_i [l] = \arcsin \left(h_{\bar{N}_0} [r_i [l]] \right) . \tag{2.121}$$

3) The 2D FFT of the matrix $\mathbf{X}[l]$ is computed; this produces the $\bar{N}_0 \times \bar{N}_0$ matrix $\bar{\mathbf{S}}[l] = [\bar{S}_{k,r}[l]]$, such that

$$\bar{S}_{k,r} [l] \triangleq \frac{1}{N_{\mathrm{VR}}} \sum_{q=0}^{N_{\mathrm{VV}}-1} \sum_{p=0}^{N_{\mathrm{VH}}-1} X_{0,\hat{\alpha}_l} [p,q] \cdot \exp \left(-j \frac{2\pi}{\lambda} \psi_{r,k} \right) , \tag{2.122}$$

where

$$\psi_{r,k} \triangleq q \, h_{\bar{N}_0}[r] \, d_{\mathrm{VV}} + p \, h_{\bar{N}_0}[k] \, d_{\mathrm{VH}}. \tag{2.123}$$

4) The dominant peaks of the sequence $\{[|\bar{S}_{k,r_i[l]}[l]|; \; k = 0, 1, ..., \bar{N}_0 - 1\}$ are identified (with $i = 0, 1, ..., L_v[l] - 1$); let $L_h[i,l]$ denote their overall number. If the m-th peak is found for $k = k_{m,i}[l]$ (with $m = 0, 1, ..., L_h[i,l] - 1$), the azimuth $\hat{\theta}_{i,r_i[l]}[l]$ of the associated target is evaluated as

$$\hat{\theta}_{i,r_i[l]} [l] = \arcsin \left(\frac{h_{\bar{N}_0} [k_{m,i}[l]]}{\cos \left(\hat{\phi}_i [l] \right)} \right) , \tag{2.124}$$

where $\hat{\phi}_i [l]$ is expressed by (2.121); consequently, the angular coordinates of the i-th target detected in the $\hat{\alpha}_l$-th frequency bin are $(\hat{\theta}_{i,r_i[l]} [l], \; \hat{\phi}_i [l])$, whereas its range is computed on the basis of $\hat{\alpha}_l$.

The last step concludes our description of the FFT-BAs. Note that the overall number of detected targets is given by

$$\hat{L} = \sum_{l=0}^{L_b-1} \sum_{i=0}^{L_v[l]-1} L_h\,[i,l]\,. \tag{2.125}$$

MUSIC-*based* DOA *estimation* – Similarly as our description of the FFT-BAs, we first focus on the case in which a virtual HULA, consisting of $N_{\rm VH}$ virtual elements, is employed for resolving the targets associated with a given frequency bin and estimating their azimuth. In this case, the algorithm considered for DOA estimation consists of the following three steps:

1) The $N_{\rm VH} \times N_{\rm VH}$ autocorrelation matrix

$$\mathbf{R}_X\,[l] = \mathbf{X}\,[l]\;\mathbf{X}[l]^H \tag{2.126}$$

is computed; here, $\mathbf{X}\,[l]$ is defined by (2.118).

2) The $\bar{N}_0$-dimensional *pseudo-spectrum* $\mathbf{P}_{\rm MU}^{(l)}$ is evaluated; its k-th element is given by

$$\mathscr{P}_{\rm MU}^{(l)}[k] = \frac{1}{\mathbf{a}^H\,[k]\;\mathbf{Q}_{N_{\rm VR}}\mathbf{Q}_{N_{\rm VR}}^H\mathbf{a}\,[k]} \tag{2.127}$$

with $k = 0, 1, ..., \bar{N}_0-1$; here, $(\cdot)^H$ denotes the conjugate and transpose operator, $\mathbf{Q}_{N_{\rm VH}}$ is a matrix having size $N_{\rm VH} \times (N_{\rm VH}-1)$ and whose columns are the $(N_{\rm VH}-1)$ noise eigenvectors (associated with the $(N_{\rm VH}-1)$ smallest eigenvalues) of $\mathbf{R}_X[l]$ (2.126) and $\mathbf{a}[k]$ is a $N_{\rm VH}$-dimensional steering vector, whose n-th element $a_n[k]$ is given by[9]

$$a_n[k] = \exp\left(j\pi n\, h_{\bar{N}_0}\,[k]\right), \tag{2.128}$$

with $n = 0, 1, ..., N_{\rm VH} - 1$.

3) The dominant peaks appearing in the sequence $\{\mathscr{P}_{\rm MU}^{(l)}[k];\ k = 0,\ 1,\ ...,\ \bar{N}_0 - 1\}$, consisting of the ordered elements of $\mathbf{P}_{\rm MU}^{(l)}$, are identified; let $L_h[l]$ denote their overall number. If the i-th peak is found for $k = k_i[l]$ (with $i = 0,\ 1,\ ...,\ L_h[l]-1$), the azimuth $\hat{\theta}_i[l]$ of the associated target is evaluated on the basis of (2.119)–(2.120).

Let us consider now the case in which the uniform rectangular array shown in Fig. 2.2 is employed for resolving the targets associated with each frequency bin, and estimating their azimuth and elevation. In this case, the adopted procedure involves the $N_{\rm VH} \times N_{\rm VV}$ matrix $\mathbf{X}[l] \triangleq [X_{0,\hat{\alpha}_l}[p,q]]$ (2.36) for any $\hat{\alpha}_l$ and consists of the following four steps:

1) The pseudo-spectrum referring to the reference VULA (that consists of $N_{\rm VULA}$ virtual elements) is evaluated. In this step, we assume that the p_R-th row of $\mathbf{X}[l]$ is employed for the evaluation of the autocorrelation matrix $\mathbf{R}_X[l]$ (2.126) and that the $\bar{N}_0$-dimensional vector $\mathbf{P}_{\rm MU}^{(\rm VULA)}[l]$ is computed on the basis of (2.127)–(2.128) (note that $N_{\rm VR}$ and $\delta[k]$ are replaced by $N_{\rm VULA}$ and $\delta[r]$, respectively).

2) The dominant peaks appearing in the sequence of the ordered elements of $\mathbf{P}_{\rm MU}^{(\rm VULA)}[l]$ are identified; let $L_v[l]$ denote their overall number. If the i-th peak is

[9] Note that the following equation applies to an FMCW radar; if an SFCW radar is considered, the sign of the argument of the complex exponential appearing in its RHS must be reversed.

found for $r = r_i[l]$ (with $i = 0, 1, ..., L_v[l] - 1$), the elevation $\hat{\phi}_i[l]$ of the associated target is evaluated on the basis of (2.121).

3) The pseudo-spectrum $\mathbf{P}_{\mathrm{MU}}^{(\mathrm{HULA})}[l, i]$ associated with the i-th estimated elevation is evaluated for the whole virtual array. In this step, if we assume that the autocorrelation matrix $\mathbf{R}_X$ is computed according to (2.126) (where, however, $\mathbf{X}[l]$ is the $N_{\mathrm{VH}} \times N_{\mathrm{VV}}$ matrix defined above), the $\bar{N}_0$-dimensional vector $P_{\mathrm{MU}}^{(\mathrm{HULA})}[l, i]$ is generated on the basis of (2.127). Note that, in this case, N_{VR} is replaced by N_{HULA} and that the n-th element $a_n[k]$ of the N_{HULA}-dimensional steering vector $\mathbf{a}[k]$ is

$$a_n[k] = \exp\left(j\pi n\, h_{\bar{N}_0}[k] \cos\left(\hat{\phi}_i[l]\right)\right) \tag{2.129}$$

with $n = 0, 1, ..., N_{\mathrm{HULA}}$.

4) The dominant peaks appearing in the sequence of the ordered elements of $\mathbf{P}_{\mathrm{MU}}^{(\mathrm{HULA})}[l]$ are identified; let $L_h[i, l]$ denote their overall number. If the m-th peak is found for $k = k_{m,i}[l]$ (with $m = 0, 1, ..., L_h[i, l] - 1$), the azimuth $\hat{\theta}_{i,r_i[l]}[l]$ of the associated target is evaluated as

$$\hat{\theta}_{i,r_i[l]}[l] = \arcsin\left(h_{\bar{N}_0}[k_{m,i}[l]]\right). \tag{2.130}$$

Consequently, the angular coordinates of this target are $(\hat{\theta}_{i,r_i[l]}[l],\ \hat{\phi}_i[l])$, whereas its range is computed on the basis of its bin index $\hat{\alpha}_l$. The last step concludes our description of the MUSIC-BAs. Finally it is important to point out that:

a) The overall number of targets detected by these algorithms is still expressed by (2.125).

b) The order adopted in the computation of the pseudo-spectra (first the vertical pseudo spectrum $\mathbf{P}_{\mathrm{MU}}^{(\mathrm{VULA})}[l]$, then the horizontal pseudo-spectra $\{\mathbf{P}_{\mathrm{MU}}^{(\mathrm{HULA})}[l, i]\}$) is dictated by the fact $\mathbf{P}_{\mathrm{MU}}^{(\mathrm{HULA})}[l, i]$ depends on the elevation estimate $\hat{\phi}_i[l]$ for any i.

The performance of the FFT-BAs and the MUSIC-BAs has been assessed for FMCW radar systems working in 2D and 3D propagation scenarios. The acronyms adopted in the following for these types of algorithms are summarized in Table 2.2.

Table 2.2: *Acronyms adopted for the FFT-based and MUSIC-based algorithms.*

Alg. \ Radar	Complex FMCW	Real FMCW
FFT-BA (2D)	FFT-FC2	FFT-FR2
MUSIC-BA (2D)	MUSIC-FC2	MUSIC-FR2
FFT-BA (3D)	FFT-FC3	FFT-FR3
MUSIC-BA (3D)	MUSIC-FC3	MUSIC-FR3

2.7 Computational complexity

The computational cost of the algorithms described in Sections 2.3 and 2.6 has been carefully assessed in terms of *floating point operations* (flops) to be executed in the detection of L targets[10]. Various details about the method we adopted for the evaluation of the computational cost of each algorithm are provided for the RASCA-FC3 only.

RASCA-FC3 computational cost evaluation

Here, the computational complexity, in terms of flops, is assessed for the RASCA-FC3 developed in Section 2.3. The overall computational cost of this algorithm can be expressed as

$$\mathscr{C}_{\mathrm{FC3}} = N_{\mathrm{VR}}\,\mathscr{C}_{\mathrm{T}_1} + N_A\,K_{\mathrm{T}_2}\,\mathscr{C}_{\mathrm{T}_2} + L_b\,K_{\mathrm{T}_2}\,\mathscr{C}_{\mathrm{T}_3} + \mathscr{C}_{\mathrm{T}_{\mathrm{sc}}}, \tag{2.131}$$

where $\mathscr{C}_{\mathrm{T}_1}$ is the contribution due to the first task of the the RASCA-FC3, K_{T_2} (K_{T_3}) represents the overall number of iterations carried out by the STDREC (STDAEC) algorithm, $\mathscr{C}_{\mathrm{T}_2}$ ($\mathscr{C}_{\mathrm{T}_3}$) is the contribution due to a single iteration of the STDREC (STDAEC) executed on a single VA (on the whole virtual array for a given frequency bin) and $\mathscr{C}_{\mathrm{T}_{\mathrm{sc}}}$ is the contribution due to the computation of the spatial coordinates of the overall image. The general criteria adopted in estimating the computational costs appearing in the RHS of (2.131) are illustrated in [315] and can be summarised as follows:

- $4d-2$ flops are required to compute the inner product $\mathbf{u}_c^T\,\mathbf{v}$ of the d-dimensional complex column vector $\mathbf{u}_c$ and the d-dimensional real column vector $\mathbf{v}$;
- $6d + 2(d-1)$ flops are required to compute the inner product $\mathbf{u}_c^T\,\mathbf{v}_c$ of the d-dimensional complex vectors $\mathbf{u}_c$ and $\mathbf{v}_c$;
- d flops are required to find the largest element of d-dimensional real vector $\mathbf{v}$;
- $4d^2+14d-8$ flops are required to compute an interpolation based on the elements of the d-dimensional complex vector $\mathbf{v}$.
- $8d\log_2(d)$ flops are required to compute the FFT of the d-dimensional complex vector $\mathbf{v}$.

The expressions of the computational costs associated with each of the three tasks of the RASCA-FC3 are illustrated below.

$\mathbf{T}_1$ - The cost $\mathscr{C}_{\mathrm{T}_1}$ can be expressed as

$$\mathscr{C}_{\mathrm{T}_1} \triangleq \mathscr{C}_{\mathbf{x}_{k,\mathrm{ZP}}} + \mathscr{C}_{\mathbf{X}_k}, \tag{2.132}$$

where: a) $\mathscr{C}_{\mathbf{x}_{\mathrm{ZP}}} = 4\,N$ is the contribution due to the computation of the vectors $\{\mathbf{x}_{k,\mathrm{ZP}}^{(v)};\, k=0,1,2\}$ (see (2.19)–(2.21)); b) $\mathscr{C}_{\mathbf{X}} = 24N_0\log_2 N_0$ is the contribution due to the computation of the vectors $\{\mathbf{X}_k^{(v)};\, k=0,1,2\}$ (see (2.23)).

[10] In the remaining part of this section, the overall number of estimated targets ($\hat{L}$) is assumed to be equal to L, for simplicity.

$\mathbf{T}_2$ - The computational cost of this task is mainly due to its main algorithm, i.e., to the STDREC algorithm. The cost $\mathscr{C}_{\mathrm{T}_2}$ can be expressed as

$$\mathscr{C}_{\mathrm{T}_2} \triangleq \mathscr{C}_{\mathrm{CSFE}} + \mathscr{C}_{\mathbf{C}_{X_k}} + \mathscr{C}_E, \tag{2.133}$$

where: a) $\mathscr{C}_{\mathrm{CSFE}} = 4N_{\mathrm{CSFE}}I^2$ is the cost originating from the CSFE[11] employed in **STDREC-S1**; b) $\mathscr{C}_{\mathbf{C}_{X_k}} = 18N_0$ is the contribution due to the computation of the vectors $(\mathbf{C}_{X_0}^{(v_k)}[i], \mathbf{C}_{X_1}^{(v_k)}[i], \mathbf{C}_{X_2}^{(v_k)}[i])$ (see (2.101)–(2.103)); c) $\mathscr{C}_E = 8N_0 - 2$ is the contribution due to the computation of the residual energy (see (2.30)).

$\mathbf{T}_3$ - The cost $\mathscr{C}_{\mathrm{T}_3}$ can be expressed as

$$\mathscr{C}_{\mathrm{T}_3} \triangleq \mathscr{C}_{\mathrm{CSFE_V}} + \mathscr{C}_{\mathbf{X}^{(\mathrm{VF})}} + \mathscr{C}_{\mathrm{CSFE_H}} + \mathscr{C}_{\mathbf{X}_{\mathrm{OF}}} + \mathscr{C}_{\mathrm{CSFE_{OF}}} + \mathscr{C}_E, \tag{2.134}$$

where: a) $\mathscr{C}_{\mathrm{CSFE_V}} = 8\bar{N}_0 \log_2(\bar{N}_0) + 4N_{\mathrm{CSFE}}I^2$ is the cost originating from the CSFE employed in **STDAE-S1**; b) $\mathscr{C}_{\mathbf{X}^{(\mathrm{VF})}} = 6N_{\mathrm{VV}}N_{\mathrm{VH}} + 2N_{\mathrm{VV}}$ is the contribution due to the computation of the vertically folded spectrum $\mathbf{X}_i^{(\mathrm{VF})}[l]$ (2.52) in **STDAE-S2**; c) $\mathscr{C}_{\mathrm{CSFE_H}} = 8\bar{N}_0 \log_2(\bar{N}_0) + 4N_{\mathrm{CSFE}}I^2$ is the cost originating from the CSFE employed in **STDAE-S3**; d) $\mathscr{C}_{\mathbf{X}_{\mathrm{OF}}} = 6N_{\mathrm{VV}}N_{\mathrm{VH}} + 18N_{\mathrm{VV}}N_{\mathrm{VH}}N_0$ is the contribution due to the computation of the overall folded spectrum $\{\mathbf{X}_{m,\mathrm{OF}}[l]; m = 0, 1, 2\}$ (see (2.54) and (2.59)); e) $\mathscr{C}_{\mathrm{CSFE_{OF}}} = 4N_{\mathrm{CSFE}}I^2$ is the cost due to the CSFE[12] in **STDAE-S4**; f) $\mathscr{C}_E = 6N_{\mathrm{VV}}N_{\mathrm{VH}}$ is the contribution due to the computation of the residual energy in **STDAEC-S3** (see (2.39)). Finally, the cost $\mathscr{C}_{\mathrm{T_{sc}}} = 5L$ is required to generate the overall point cloud. Based on the results illustrated above, (2.131) can be rewritten as

$$\begin{aligned}\mathscr{C}_{\mathrm{FC3}} =& N_{\mathrm{VR}}(4N + 24N_0 \log_2(N_0)) \\ &+ N_{\mathrm{T}_2}(4N_{\mathrm{CSFE}}I^2 + 26N_0 - 2) \\ &+ N_{\mathrm{T}_3}(12N_{\mathrm{CSFE}}I^2 + 2N_{\mathrm{VV}} + 18N_{\mathrm{VH}}N_{\mathrm{VV}} \\ &+ 18N_{\mathrm{VH}}N_{\mathrm{VV}}N_0 + 16\bar{N}_0 \log_2(\bar{N}_0) + 5L,\end{aligned} \tag{2.135}$$

where $N_{\mathrm{T}_2} \triangleq N_A K_{\mathrm{T}_2}$ and $N_{\mathrm{T}_3} \triangleq L_b K_{\mathrm{T}_2}$.

Our analysis leads to the conclusion that the overall cost of RASCA-FC3 and RASCA-FC2 is approximately of order $\mathcal{O}(M_{\mathrm{R-FC3}})$ and $\mathcal{O}(M_{\mathrm{R-FC2}})$, respectively, where (see (2.135))

$$\begin{aligned}M_{\mathrm{R-FC3}} =& 24N_{\mathrm{VR}} N_0 \log_2(N_0) + 26N_A K_{\mathrm{T}_2} N_0 \\ &+ L_b K_{\mathrm{T}_3} \left(18N_{\mathrm{VH}} N_{\mathrm{VV}} N_0 + 16\bar{N}_0 \log_2\left(\bar{N}_0\right)\right)\end{aligned} \tag{2.136}$$

and

$$\begin{aligned}M_{\mathrm{R-FC2}} =& 24N_{\mathrm{VH}} N_0 \log_2(N_0) + 26N_A K_{\mathrm{T}_2} N_0 \\ &+ L_b K_{\mathrm{T}_3} \left(18N_{\mathrm{VH}} N_0 + 8\bar{N}_0 \log_2(\bar{N}_0)\right);\end{aligned} \tag{2.137}$$

here, K_{T_2} (K_{T_3}) represents the overall number of iterations carried out by the STDREC (STDAEC) algorithm.

[11]Note that, in this case, the cost of the CSFE does not account for the evaluation of three DFTs, since these have been already evaluated in **T1**.

[12]Note that, in this case, the cost of the CSFE does not account for the evaluation of three DFTs, since these are made available by overall folding.

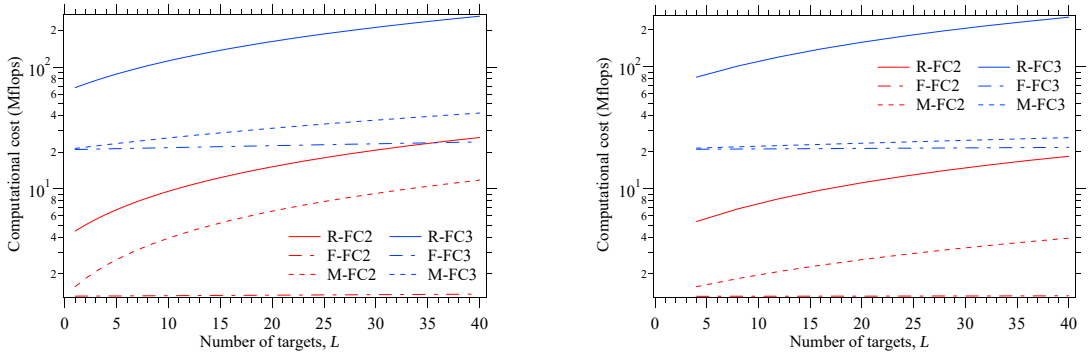


Figure 2.8: *Computational cost versus overall number of targets for: a) the first scenario; b) the second scenario. The RASCA-FC3, RASCA-FC2, FFT-FC3, FFT-FC2, MUSIC-FC3 and MUSIC-FC2 are considered.*

MUSIC and FFT computational cost

In evaluating the cost of the FFT-based and MUSIC-based algorithms described in Section 2.6, we have assumed that: a) the cost due to the computation of the eigenvalue decomposition of a $d \times d$ matrix is $\mathcal{O}(d^3)$; b) the computational effort required to find the dominant peaks in a sequence is negligible. Based on these assumptions, it can be shown that the computational complexity of the FFT-FC3, FFT-FC2, MUSIC-FC3 and MUSIC-FC2 algorithms are approximately of order $\mathcal{O}(M_{\mathrm{F-FC3}})$, $\mathcal{O}(M_{\mathrm{F-FC2}})$, $\mathcal{O}(M_{\mathrm{M-FC3}})$ and $\mathcal{O}(M_{\mathrm{M-FC2}})$, respectively, where

$$\begin{aligned} M_{\mathrm{F-FC3}} =& 8N_{\mathrm{VR}}N_0\log_2(N_0) \\ &+ 8L_b\left(\bar{N}_0^2\log_2(\bar{N}_0^2) + \bar{N}_0\log_2(\bar{N}_0)\right), \end{aligned} \tag{2.138}$$

$$M_{\mathrm{F-FC2}} = 8\big(N_{\mathrm{VH}}N_0\log_2(N_0) + L_b\bar{N}_0\log_2(\bar{N}_0)\big), \tag{2.139}$$

$$\begin{aligned} M_{\mathrm{M-FC3}} =& 8N_{\mathrm{VR}}N_0\log_2(N_0) + L_b\bar{N}_0(N_{\mathrm{VV}}^3 + N_{\mathrm{VH}}^3) \\ &+ 16L_b\bar{N}_0(N_{\mathrm{VV}}^2 + N_{\mathrm{VH}}^2) \end{aligned} \tag{2.140}$$

and

$$M_{\mathrm{M-FC2}} = 8N_{\mathrm{VH}}N_0\log_2(N_0) + L_b\bar{N}_0(N_{\mathrm{VH}}^3 + 16N_{\mathrm{VH}}^2). \tag{2.141}$$

Computational cost comparison

It is important to keep in mind that a comparison among the computational costs listed above does not fully account for the gap that can be observed in the execution speed of the corresponding algorithms. In fact, in practice, a portion of the computation time is absorbed by the procedure employed to find the dominant peaks of real sequences in both the FFT-BAs and the MUSIC-BAs. Moreover, the vector $\mathscr{A}_{\mathrm{CF}}$ (2.117), collecting the indices of the frequency bins in which *at least* one target has been detected, may include *ghost* targets; as evidenced by our computer simulations, the impact of this phenomenon on the overall computation time may not be negligible. Despite this,

some interesting insights on how the complexity is influenced by the overall number of targets can be obtained by comparing the computational costs (2.136), (2.138) and (2.140) ((2.137), (2.139) and (2.141)) in two specific scenarios. The first scenario we take into consideration refers to the case in which the mutual distance between the targets is *above* the range resolution of the employed radar system, so that $K_{\mathrm{T}_2} = L$, $K_{\mathrm{T}_3} = 1$ and $L_b = L$ can be assumed in the RHS of (2.136)–(2.141). In our second scenario, instead, the targets form clusters, each of which consists of four targets having the *same* range, but different angular coordinates; for this reason, $K_{\mathrm{T}_2} = L/4$, $K_{\mathrm{T}_3} = 4$ and $L_b = L/4$ can be assumed in the RHS of (2.136)–(2.141). Moreover, the following parameters have been chosen for both scenarios: a) $N_{\mathrm{VR}} = 256$; b) $N_0 = 1024$; c) $N_A = 10$ d) $N_{\mathrm{VV}} = 16$; e) $N_{\mathrm{VH}} = 16$; f) $\bar{N}_0 = 32$. The dependence of the complexity M_{alg} on L is represented in Fig. 2.8a (Fig. 2.8b) for the first (second) scenario; here, alg denotes the algorithm which this complexity refers to. From these figures it is easily inferred that:

a) The RASCAs require the largest computational effort in both the considered scenarios for any value of L; for instance, $M_{\mathrm{RASCA-FC3}}$ is approximately 4.1 (4.6) times greater than $M_{\mathrm{MUSIC-FC3}}$ in the first (second) scenario for $L = 8$.

b) The ratio between $M_{\mathrm{RASCA-FC2}}$ and $M_{\mathrm{MUSIC-FC2}}$ is approximately 2.4 for any value of L in the first scenario, but it increases with L in the second scenario; for instance, $M_{\mathrm{RASCA-FC2}}$ is 3.9 (4.6) times greater than $M_{\mathrm{MUSIC-FC2}}$ for $L = 12$ ($L = 36$).

c) The computational cost estimated for the MUSIC-BAs in the first scenario is larger than that referring to the second scenario for $L \geq 4$; for instance, if $L = 20$, the value of $M_{\mathrm{MUSIC-FC3}}$ ($M_{\mathrm{MUSIC-FC2}}$) evaluated in first scenario is 1.33 (2.5) times larger than that found in the second scenario .

Finally, it is important to stress that:

a) in real world scenarios, the overall number of targets detected by the FFT-BAs and the MUSIC-BAs may be greater than the true number of targets, since some targets are detected multiple times; this may have a significant impact on the overall computational effort required by these algorithms.

b) The computational complexity of the RASCA-FR3 (RASCA-FR2), the FFT-FR3 (FFT-FR2), and the MUSIC-FR3 (MUSIC-FR2) is of the same order as the RASCA-FC3 (RASCA-FC2), the FFT-FC3 (FFT-FC2) and the MUSIC-FC3 (MUSIC-FC2), respectively.

2.8 Numerical results

In this section, the accuracy of the RASCAs is assessed on the basis of both synthetically generated and experimental data, and is compared with that provided by various FFT-BAs and MUSIC-BAs.

2.8.1 Numerical results based on synthetically generated measurements

In this paragraph, the accuracy achieved by the RASCA-FC3, the FFT-FC3 and the MUSIC-FC3 in the generation of 3D radar images is assessed. The performance of these algorithms has been evaluated in a colocated MIMO FMCW radar system providing both the in-phase and quadrature components of all its received signals, and equipped with an URA consisting of $N_T = 16$ TX and $N_R = 16$ RX antennas; therefore, the available virtual array is made of $16 \cdot 16 = 256$ VAs with inter-antenna spacings $d_{\mathrm{VV}} = d_{\mathrm{VH}} = \lambda/4$. The other relevant parameters of the considered radar system are: a) chirp slope $\mu = 4 \cdot 10^{13}$ Hz/s; b) bandwidth $B = 2.5$ GHz; c) central frequency $f_0 = 77$ GHz; d) sampling frequency $f_s = 8$ MHz; e) number of samples per chirp $N = 512$. Note that, in principle, the available antenna array allows us to achieve the range resolution

$$\Delta R = \frac{c}{2B} \cong 6 \text{ cm}, \tag{2.142}$$

the azimuthal resolution

$$\Delta\theta = \frac{\lambda}{2d_{\mathrm{VH}}(N_{\mathrm{HULA}} - 1)} \cong 7.45^\circ \tag{2.143}$$

and the elevation resolution

$$\Delta\phi = \frac{\lambda}{2d_{\mathrm{VV}}(N_{\mathrm{VULA}} - 1)} \cong 7.45^\circ. \tag{2.144}$$

The considered radar system is assumed to operate in the presence of $L = 10$ targets, whose echoes have unit amplitude. The range, the azimuth and the elevation of each target are sequentially generated at the beginning of each run. Moreover, the range R_k, the azimuth θ_k and the elevation ϕ_k of the k-th target (with $k = 1, 2, ..., 10$) have been randomly evaluated in a way that: a) they belong to the intervals [1, 10] m, $[-\pi/3, \pi/3]$ rad and $[-\pi/3, \pi/3]$ rad, respectively; b) the minimum spacing between the k-th target and the previously generated $(k-1)$ targets is not smaller than ΔR (2.142), $\Delta\theta$ (2.143) and $\Delta\phi$ (2.144) in the range, azimuth and elevation dimensions, respectively (scenario **S1**) or is not smaller than ΔR (2.142) in the range domain, but can be arbitrarily small in the azimuth and elevation dimensions (this scenario is denoted **S2**). In our computer simulations, the following values have been selected for the parameters of the RASCA-FC3:

a) FFT Processing: $M = 2$;

b) RPE: $N_A = 10$, $N_{\mathrm{CSFE}} = 10$, $I = 7$ and $T_{\mathrm{STDREC}} = 0.001 \cdot E_0^{(v_k)}$ (see (2.30));

c) SPE: $\bar{M} = 2$, $N_{\mathrm{CSFE}} = 10$, $I = 7$ and $T_{\mathrm{STDAEC}} = 0.001 \cdot E^{(0)}[l]$ (see (2.39)), $T_{\mathrm{OF}} = 0$.

In addition, the following values have been selected for the parameters of the FFT-FC3 and the MUSIC-FC3: $C_s = 3$, $G_s = 2$ and $K_0 = 1.5$. The SNR $\triangleq 1/\sigma^2$ has been assumed to be equal to 10 dB and the following performance indices have been evaluated to assess estimation accuracy: a) The *detection rate* (R_{D}) defined as the

Table 2.3: Root mean square error $\bar{\varepsilon}_X$, peak error $\hat{\varepsilon}_X$ *and* detection rate R_D *evaluated in the two simulation scenarios defined in Paragraph 2.8.1. Target range, azimuth and elevation are taken into consideration.*

		$\bar{\varepsilon}_X$			$\hat{\varepsilon}_X$			
Scenario	Algorithm	R (m)	θ (°)	ϕ (°)	R (m)	θ (°)	ϕ (°)	R_{D} (%)
S1	RASCA-FC3	0.01	1.56	0.79	0.02	9.86	2.64	100
	FFT-FC3	0.02	2.07	1.05	0.04	14.34	3.06	71
	MUSIC-FC3	0.02	1.74	0.83	0.04	34.03	13.48	70
S2	RASCA-FC3	0.01	1.54	0.79	0.02	8.01	2.31	100
	FFT-FC3	0.02	2.05	1.05	0.04	14.78	2.95	70
	MUSIC-FC3	0.02	1.53	0.81	0.04	7.56	2.03	0.72

percentage of simulation runs in which the considered algorithm detects all the targets; b) the *root mean square error* (RMSE)

$$\bar{\varepsilon}_X \triangleq \sqrt{N_m^{-1} \sum_{k=0}^{N_m-1} \left[X_k - \hat{X}_k\right]^2}; \tag{2.145}$$

c) the *peak error*

$$\hat{\varepsilon}_X \triangleq \max_k \left|X_k - \hat{X}_k\right|; \tag{2.146}$$

here, X_i and $\hat{X}_i$ represent the exact value of a parameter X and its corresponding estimate, whereas N_m represents the overall number of synthetically generated values of X; note that, if all the targets are detected by the considered algorithm in each run,

$$N_m \triangleq N_r\, L \tag{2.147}$$

where N_r is the overall number of simulation runs. In our work, the performance of the above mentioned algorithms has been assessed by: a) evaluating the detection rate for both the considered scenarios; b) ignoring the failure events (i.e., the events in which not all the targets have been detected) in the evaluation of all the RMSEs. The three performance indices defined above have been assessed on the basis of the estimates generated by executing $N_r = 500$ runs; the resulting values are summarised in Table 2.3. From these results it is easily inferred that:

a) The RASCA-FC3 achieves the lowest RMSEs in range, azimuth and elevation (range, elevation) estimation in the first (second) scenario; for instance, the RMSE $\bar{\varepsilon}_\theta$ characterising the RASCA-FC3 is about 1.3 (1.3) times smaller than the corresponding RMSE obtained for the FFT-FC3 in the first (second) scenario.

b) The RASCA-FC3 exhibits the lowest peak errors in range, azimuth and elevation (range) in the first (second) scenario; for instance, its peak error $\hat{\varepsilon}_R$ is 2 times smaller than the corresponding RMSE obtained for the FFT-FC3 and the MUSIC-FC3 in both scenarios.

c) All the considered algorithms achieve an excellent accuracy in both scenarios, since the RMSEs evaluated for range, azimuth and elevation are smaller than the corresponding resolutions given above.

d) The FFT-FC3 and the MUSIC-FC3 are outperformed by the RASCA-FC3 in terms of detection rate; in fact, the value of this parameter is about 70 % for the first two algorithms, but is equal to 100 % for the RASCA-FC3, since the last algorithm has been able to detect all the targets in every simulation run in both scenarios.

2.8.2 Numerical results based on experimental measurements

In this paragraph, we first describe the radar devices employed in our measurement campaigns and the adopted experimental setup. Then, we analyse: 1) the accuracy achieved by our RPE (and, in particular, by the STDREC algorithm) in range and phase estimation on multiple antennas of the same array in the presence of a single target and of multiple targets; 2) the accuracy of the 2D (3D) images generated by RASCA-FR2 (RASCA-FR3) and RASCA-FC2 (RASCA-FC3) in the presence of multiple targets.

Employed radar devices and adopted experimental setup

A measurement campaign has been accomplished to acquire a data set through two FMCW MIMO radars, all operating in the E-band. The first FMCW device, dubbed **TI FMCW** radar in the following, is the TIDEP-01012 *Cascade mmWave* radar (see Fig. 2.9-a)). It is manufactured by *Texas Instrument Inc.* [82], classified as a *long range radar* (LLR) and provides both the in-phase and quadrature components of received signals (i.e., *complex measurements*). Its main parameters are: a) chirp slope $\mu = 4 \cdot 10^{13}$ Hz/s; b) bandwidth $B_1 = 2.5$ GHz; c) central frequency $f_c = 77$ GHz; d) sampling frequency $f_s = 8$ MHz; e) number of samples per chirp $N = 512$. Moreover, it is endowed with a planar array made of $N_T = 12$ TX and $N_R = 16$ RX antennas (each consisting of an array of four patch elements), as shown in Fig. 2.9-a). The corresponding virtual array consists of $12 \cdot 16 = 192$ VAs; however, only 134 of them are available, since the remaining 58 VAs overlap with the other elements of the virtual array. As shown in Fig. 2.9-b) (where each VA is represented by a small blue circle), the virtual array has the following characteristics:

1. the non-overlapped VAs form an *horizontal* ULA (HULA_1), consisting of $N_{\text{HULA}_1} = 86$ VAs and three smaller HULAs, each made of 16 equally-spaced VAs;
2. the inter-antenna spacing of all the HULAs is $d_{\text{VH}} = \lambda/4$;
3. the vertical spacing of the three smaller HULAs is not uniform, since $d_{\text{VV}_1} = \lambda/4$, $d_{\text{VV}_2} = \lambda$ and $d_{\text{VV}_3} = 3\lambda/2$ (see Fig. 2.9-b)).

This virtual antenna array allows us to achieve a range, azimuth and elevation resolution equal to $\Delta R_1 = 5.8$ cm (see (2.142)), $\Delta\theta_1 = 1.35°$ (see (2.143)), and $\Delta\phi_1 = 16.4°$ (see (2.144)), respectively; note that the elevation resolution is coarser than the azimuth one since $N_{\text{VV}} = 7$ equally aligned antennas ($d_{\text{VV}} = d_{\text{VV}_1} = \lambda/4$) are assumed

along the vertical direction (this is equivalent to considering an elevation aperture $D_y = 3\lambda$ along the vertical direction; see [54]). In our work, on the one hand, a central portion of the first HULA (contained inside the red rectangle appearing in Fig. 2.9-b)), consisting of $N_{\mathrm{VV}} = 16$ antennas, has been exploited for 2D imaging, in order to guarantee a fair comparison with the other two radar devices. On the other hand, the whole array have been employed for 3D imaging.

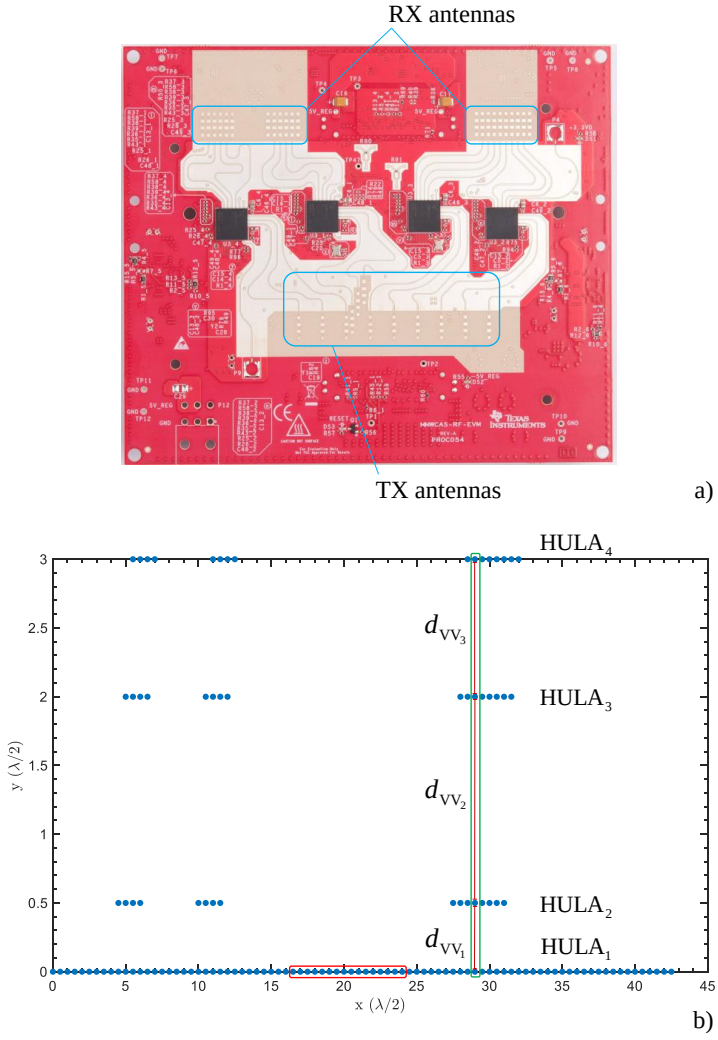


Figure 2.9: *Representation of: a) the TI FMCW radar (the physical TX/RX antennas are contained inside the three rounded rectangles); b) the corresponding virtual array (the lower rounded rectangle contains the portion of $HULA_1$ employed for 2D imaging, whereas the green one the vertical array chosen as a reference for 3D imaging).*

The second FMCW device, dubbed **Inras FMCW** radar in the following (see Fig. 2.10-a)), is a modular system manufactured by *Inras GmbH* [83] and consisting of: a) the so called *Radar Log* board; b) an RF front-end including multiple TX/RX antennas and *monolithic microwave integrated circuits* (MMIC) operating at 77 GHz. This system is classified as a LLR and its main parameters are: a) chirp slope $\mu =$

$9.7656 \cdot 10^{12}$ Hz/s; b) bandwidth $B_2 = 2.5$ GHz; c) central frequency $f_c = 77$ GHz; d) sampling frequency $f_s = 8$ MHz; e) number of samples per chirp $N = 2048$. Unlike the TI FMCW radar, this device provides only the in-phase component of the RF received signals and, consequently, *real measurements*. Moreover, it is endowed with a custom designed planar array made of $N_T = 16$ TX antennas and $N_R = 16$ RX antennas, each consisting of an array of six patch elements, as shown in Fig. 2.10-a). The resulting virtual array, consisting of $N_{\mathrm{VR}} = 16 \cdot 16 = 256$ VAs is shown in Fig. 2.10-b). As it can be inferred from the last figure, the virtual array has the following characteristics:

1. It consists of 16 HULAs, each of which is made of 16 antennas with inter-antenna spacing $d_{\mathrm{VH}} = \lambda/4$.

2. The vertical distance between each couple of its adjacent HULAs is $d_{\mathrm{VV}} = \lambda/2$; this entailes the unambiguous elevation range $[-45°, 45°]$.

3. Its shape is not rectangular (the horizontal shift of adjacent HULAs is equal to $\lambda/4$).

This virtual array allows us to achieve the same range resolution as the TI FMCW radar, and azimuth and elevation resolutions equal to $\Delta\theta_2 = 7.6°$ and $\Delta\phi_2 = 3.8°$, respectively (see (2.143)–(2.144)). In our work, the HULA contained inside the red rectangle appearing in Fig. 2.10-b) (the whole array) has been exploited for 2D (3D) imaging.

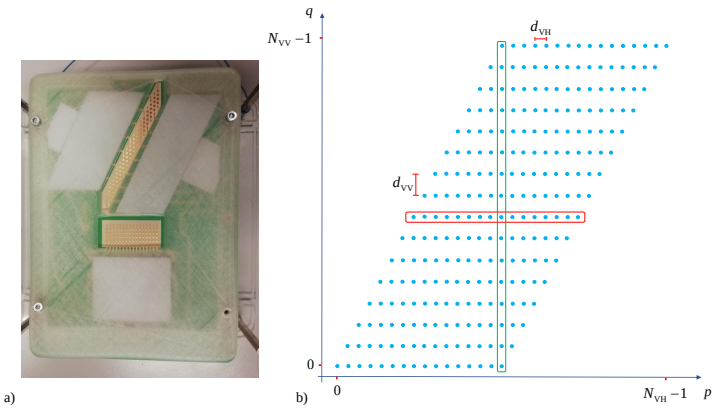


Figure 2.10: *Representation of: a) the physical array of the Inras FMCW radar; b) the corresponding virtual array (the red rounded rectangle contains the HULA employed for 2D imaging, whereas the green one the vertical array chosen as a reference for 3D imaging).*

Our measurement campaigns have been conducted in a large empty room (whose width, depth and height are 10 m, 8 m and 2.5 m, respectively). Each of the employed radar devices has been mounted on an horizontal wooden bar together with a pico-flexx camera manufactured by *PMD Technologies Inc.* [84] and has been lifted by a tripod at an height of roughly 1.60 m from ground, as shown in Fig. 2.11. The employed camera is based on a near-infrared *vertical cavity surface emitting laser*, and is able to provide a depth map or, equivalently, a 3D point-cloud of a small region of the

observed environment (its maximum depth is equal to 4 m, whereas its *field of view* is $62° \times 45°$).

In each measurement campaign, the experiments have been repeated for all the radar devices exactly in the same conditions.

It is important to point out that:

a) in all the radar systems, the target ranges have been estimated with respect to the central virtual channel of the employed ULA;

b) the exact target positions have been acquired with respect to the centre of the pico-flexx camera;

c) the data processing has been accomplished in the MATLAB environment;

d) all our detection and estimation algorithms have been run on a desktop computer equipped with a single i7 processor.

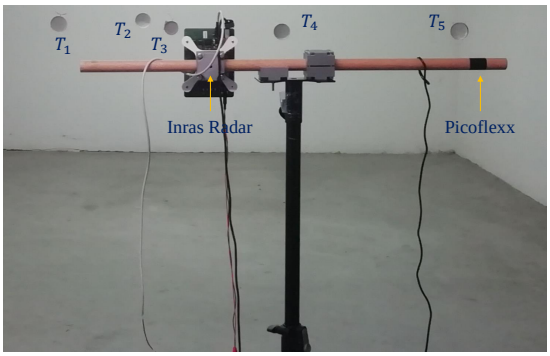


Figure 2.11: *Experimental set-up developed for our measurement campaigns. The radar device (the Inras FMCW radar in this case) and a reference sensor (pico-flexx) are mounted on a wooden bar. A group of metal targets, placed at the same height as our sensors, is also visible.*

Range and amplitude estimation

In this paragraph, the accuracy of the STDREC algorithm employed by the RPE is analysed for two specific static scenarios. The first scenario is characterized by a single detectable target (a small metal disk[13] having a diameter equal to 5.5 cm) placed in ten different positions. The target range R and azimuth θ have been selected in the interval [1.0, 3.0] m, with a step of 0.5 m and $[-40°, 40°]$. The range and azimuth of the considered targets are listed in Table 2.4 for all the employed radar devices (the data referring to the i-th position are collected in the column identified by T_i, with $i = 1$, 2, ..., 10). The second scenario, instead, is characterized by the presence of an overall number of targets ranging from 1 to 9 (so that $1 \leq L \leq 9$). The targets are represented by small coins with a diameter of 2 cm; the range and azimuth characterizing their exact positions are listed in Table 2.5 (the data referring to the i-th target are collected

[13]Each target is hung from the ceiling: a nylon thread has been used for suspending it.

in the column identified by T_i, with $i = 1, 2, ..., 9$). Each target has been sequentially added in our scenario; this has allowed us to assess how the performance of the STDREC algorithm is influenced by the value of the parameter L in the presence of *closely spaced targets*.

Prior knowledge of L has been assumed during the processing; moreover, the following values have been selected for the parameters of the STDREC algorithm[14]:

TI FMCW radar - $N = 512$, $M = 4$, $N_0 = N \cdot M = 2048$, $N_{\text{CSFE}} = 5$ and $I = 7$.

Inras FMCW radar - $N = 2048$, $M = 1$, $N_0 = N \cdot M = 2048$, $N_{\text{SFE}} = 5$ and and $I = 7$.

Note that: a) the value of the oversampling factor (M) has been selected in way to guarantee approximately the same value of N_0 in all cases, i.e. roughly the same resolution in the spectral analysis of radar signals; b) the values of the parameters N_{SFE} and N_{CSFE} are all equal and large enough so that accurate range estimation is achieved by the STDREC algorithm.

Table 2.4: *Exact positions (range and azimuth) of the considered target, estimated ranges and RMSEs evaluated for the phase fitting over the considered sixteen virtual channels of each device (first experimental scenario). All our radar devices are taken into consideration.*

Method		T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Exact (TI FMCW)	R (m)	1.2	1.25	1.5	1.6	2.16	2.23	2.85	2.55	2.96	3.25
	θ (°)	-25	40	-10	30	-18	24	-14	21	-18	25
STDREC	$\hat{R}$ (m)	1.166	1.219	1.523	1.643	2.184	2.278	2.885	2.582	2.932	3.267
	$\bar{\varepsilon}_\psi$ (rad)	0.8	0.597	0.592	0.5	0.521	0.524	0.576	0.574	0.513	0.506
Exact (Inras FMCW)	R (m)	1.35	1.3	1.73	1.72	2.2	2.1	2.71	2.67	3.2	3.2
	θ (°)	-25	35	-13	30	-5	15	-10	26	-9	20
STDREC	$\hat{R}$ (m)	1.38	1.32	1.74	1.74	2.25	2.14	2.76	2.7	3.24	3.24
	$\bar{\varepsilon}_\psi$ (rad)	0.18	0.18	0.2	0.22	0.67	0.79	0.16	0.21	0.16	0.3
	θ (°)	-35	30	-20	20	-12	15	-14	15	- 18	25
STDREC	$\hat{R}$ (m)	1.242	1.06	1.517	1.413	2.025	2.06	2.78	2.409	2.905	3.041
	$\bar{\varepsilon}_\psi$ (rad)	0.25	0.35	0.67	0.345	0.422	0.45	0.686	0.919	0.379	0.54

The accuracy of range estimates has been assessed by evaluating the RMSE $\bar{\varepsilon}_R$ and the peak error $\hat{\varepsilon}_R$, expressed by (2.145)–(2.146) with $X = R$, $X_i = R_i$ and $\hat{X}_i$. Since the RCS of the considered targets was unknown, our analysis of the complex gains available over the 16 channels of the considered virtual ULA and associated with the same target has concerned only their (unwrapped) phase. The phases $\{\psi^{(v)}; v = 1, 2, ..., 16\}$ estimated by the STDREC algorithm over the considered reference HULA (consisting of 16 VAs; see the red rounded rectangles appearing in Figs. 2.9-b), 2.10-b)) and associated with a target placed at approximately[15] the same azimuth

[14]Note that in this case the stopping criterion based on eq. (2.30) has not be employed, since the overall number of targets is known.

[15]The exact range of this target can be found in the T_7 (T_3) column for the TI FMCW and the

angle with respect to the centre of the radars is shown in Fig. 2.12. Since the distance d_{VH} between adjacent virtual channels is constant, the (unwrapped) estimated phases exhibit a linear dependence on the index of the virtual channel, as illustrated in Section 1.3 (see, in particular, (1.15) and (1.16)). Moreover, if a linear fitting is drawn for these data, it should be expected that the slope of the resulting straight line is proportional to $\sin(\theta)$ (see (1.21) with $\phi = 0$); this is confirmed by the results shown in Fig. 2.12 for the FMCW radar system. To assess the quality of the estimated phases, their RMSE $\bar{\varepsilon}_{\psi}$ has been evaluated in all the scenarios; in doing so, the linear fitting of the 16 phases $\{\psi^{(v)}\}$ has been taken as a reference with respect to which the error of each of them has been computed.

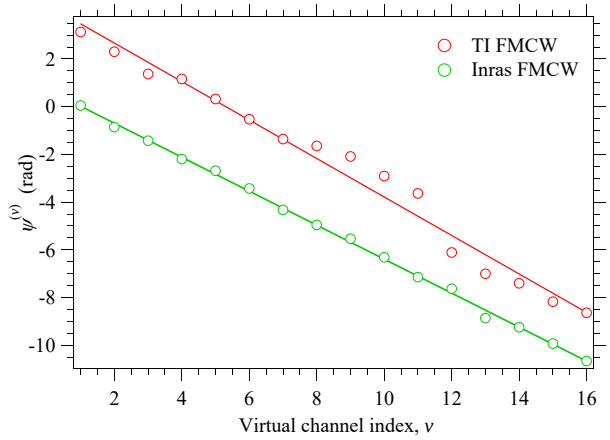


Figure 2.12: *Unwrapped phase of the complex gain versus the index of the virtual channel of the reference HULA; a single target is assumed (first experimental scenario). The phase estimates generated by the STDREC algorithm in our three radar systems are identified by red, blue and green circles, whereas straight lines represent their linear fitting.*

Inras FMCW radar in Table 2.4.

Table 2.5: *Exact ranges of the nine coins characterizing our second experimental scenario. The range estimates evaluated by the Alg-YA, the Alg-S and the STDREC, the HSTDREC and the CFH algorithm are also provided.*

Method	T_1 (m)	T_2 (m)	T_3 (m)	T_4 (m)	T_5 (m)	T_6 (m)	T_7 (m)	T_8 (m)	T_9 (m)
Exact	1.860	1.900	1.980	2.110	2.190	2.220	2.370	2.410	2.460
STDREC (TI FMCW)	1.900	1.966	2.015	2.113	2.158	2.242	2.377	2.441	2.516
HSTDREC (TI FMCW)	1.906	1.971	2.016	2.118	2.158	2.238	2.378	2.444	2.514
STDREC (Inras FMCW)	1.920	1.980	2.040	2.100	2.220	2.280	2.460	2.520	2.580
CFH (TI FMCW)	1.947	2.017	2.077	2.161	2.241	2.338	2.374	2.435	2.518
Alg-YA (TI FMCW)	2.022	2.054	2.161	2.257	2.294	2.339	2.416	2.447	2.514
Alg-S (TI FMCW)	2.020	2.142	2.142	2.142	2.153	2.236	2.383	2.433	2.522

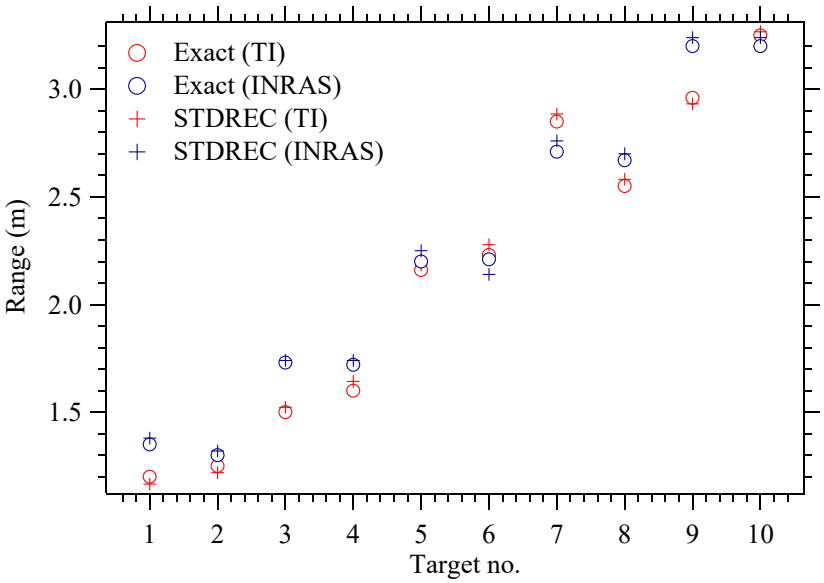


Figure 2.13: *Representation of the ranges estimated by the STDREC algorithm (first experimental scenario). All our radar devices are considered.*

The estimate of the target range generated by the STDREC algorithm for each of the $N_m = 10$ distinct positions considered in the first scenario are listed in Table 2.4; in the same table, the value of $\bar{\varepsilon}_\psi$ computed for each position is also given. The target ranges and their estimates listed in Table 2.4 are also represented in Fig. 2.13. The errors $\bar{\varepsilon}_R$ and $\hat{\varepsilon}_R$, the mean of $\bar{\varepsilon}_\psi$ (denoted $\bar{\varepsilon}_{m,\psi}$ and generated by taking the average of the N_m values available for $\bar{\varepsilon}_\psi$) and the average *computation time* (CT) evaluated on the basis of these results are listed in Table 2.6.

Table 2.6: Root mean square error $\bar{\varepsilon}_R$, peak error $\hat{\varepsilon}_R$, mean error $\bar{\varepsilon}_{m,\psi}$ *and CT evaluated for the STDREC algorithm in our first experimental scenario.*

Method	$\bar{\varepsilon}_R$ (m)	$\hat{\varepsilon}_R$ (m)	$\bar{\varepsilon}_{m,\psi}$ (rad)	CT (msec)
TI FMCW	0.033	0.048	0.570	0.3
INRAS FMCW	0.035	0.050	0.30	0.4

The results referring to the first scenario lead us to the following conclusions:

1. In all the considered cases, the STDREC is able to accurately estimate the range and the phase characterizing each target.

2. All the values of $\bar{\varepsilon}_R$ and $\hat{\varepsilon}_R$ are comparable, reasonably low and in the order of the resolution of our devices.

3. The range estimates evaluated for the two devices are closer to the correspoding exact values when the target range is above 3 m; this can be related to the fact that those are LRRs, which aperture size is big compared to the distance of the target when it is too close.

4. The Inras FMCW radar achieves the lowest $\bar{\varepsilon}_{m,\psi}$.

5. The CTs are always in the order of few milliseconds.

Let us focus on the second scenario. In this case, our range estimates have been generated by: a) the STDREC algorithm for all the radar devices; b) the HSTDREC algorithm for the TI FMCW radar; c) the Alg-YA, the Alg-S and the CFH algorithm for the TI FMCW. The obtained results are listed in Table 2.5. The errors $\bar{\varepsilon}_R$ and $\hat{\varepsilon}_R$, and the CT obtained in this case are listed in Table 2.7. From these results it can be inferred that:

1. In the case of the TI FMCW radar, all the considered algorithms achieve comparable accuracy. However, the STDREC and the HSTDREC algorithms, unlike all the other algorithms, achieve the lowest RMSE and peak error.

2. The HSTDREC algorithm is not more accurate than the STDREC algorithm; moreover, these algorithms are characterized by similar CTs.

3. The estimated RMSEs and peak errors are in the order of the resolution of our radar devices, but a little bit higher in the Inras FMCW radar systems. This is mainly due to the poorer estimates evaluated for the targets T_8 and T_9, since, in our specific experiment, the energy received from these targets has been found to be lower than that coming from the others. This problem is not so evident in the case of the TI FMCW and SFCW radars, whose RMSEs and peak errors are very low.

Finally, we would like to stress that the accuracy of STDREC and HSTDREC algorithms can be related to the accuracy of the estimation and cancellation procedure they accomplish. This is exemplified by Fig. 2.14, where the *initial amplitude spectrum* of the signal received on the central virtual channel of the TI FMCW radar in the second scenario and *its* (weak) *residual*, resulting from the cancellation of the spectral contributions due to the detected targets, are shown. Here, the range and amplitude of the targets estimated by the STDREC (HSTDREC) are also represented by red circles (green crosses).

Table 2.7: Root mean square error $\bar{\varepsilon}_R$, peak error $\hat{\varepsilon}_R$ *and* computation time *(CT) evaluated in our second experimental scenario.*

Method	$\bar{\varepsilon}_R$ (m)	$\hat{\varepsilon}_R$ (m)	CT (msec)
STDREC (TI FMCW)	0.03	0.07	20
HSTDREC (TI FMCW)	0.03	0.07	20
STDREC (Inras FMCW)	0.07	0.10	40
CFH (TI FMCW)	0.07	0.12	15
Alg-YA (TI FMCW)	0.11	0.18	40
Alg-S (TI FMCW)	0.08	0.24	45

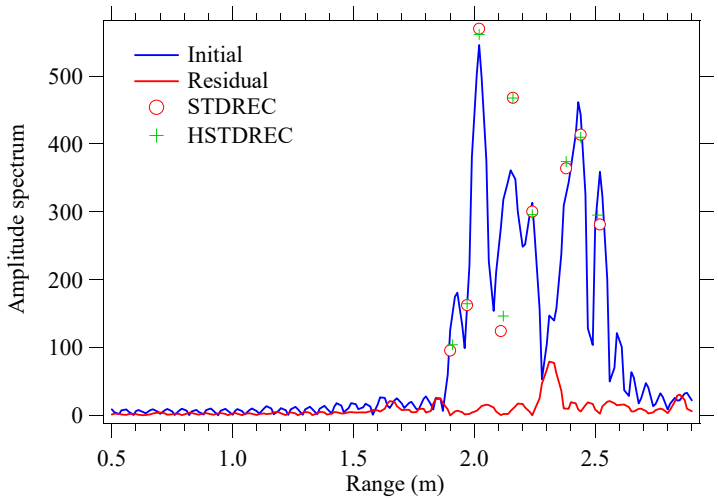


Figure 2.14: *Representation of the initial amplitude spectrum of the signal observed on the central virtual channel (blue line) and of the final* residual *amplitude spectrum generated by the STDREC algorithm (red line). The TI FMCW radar operating in our second experimental scenario is considered; moreover, the target positions estimated by STDREC (HSTDREC) are represented by red circles (green crosses).*

Two-dimensional and three-dimensional imaging

In this paragraph, the accuracy of the 2D and 3D images generated by the RASCAs is assessed. Two different groups of experiments have been carried out. The first (second) group of experiments has allowed us to assess the performance achieved by the above mentioned algorithm in 2D (3D) imaging. In both cases, the measurements have been acquired in the presence of an increasing number of targets for all our radar devices. In the first group of experiments, the following choices have been made:

1. The targets have been placed at the same height. Their range and azimuth belong to the intervals [2.2, 2.7] m and $[-15\,^{\circ}, 30^{\circ}]$, respectively (see Table 2.8).

2. The measurements have been acquired through a virtual ULA, consisting of 16 VAs, in all the considered radar systems.

As far as the second group of experiments is concerned, the following choices have been made:

1. The range, azimuth and elevation of the targets have been selected in the intervals [1.9, 2.8] m, $[-30^{\circ}, 35^{\circ}]$ and $[-10^{\circ}, 10^{\circ}]$, respectively (see Table 2.11).

2. The measurements have been acquired through the whole virtual array of each of our radar devices.

The following values have been selected for the parameters of the RASCAs: a) $N_A = 16$ ($N_A = 10$) in the RPE employed in 2D (3D) imaging; b) $N_{\mathrm{CSFE}} = N_{\mathrm{SFE}} = 5$ in both the STDREC and the STDAEC algorithms; c) the threshold $T_{\mathrm{OF}} = 0$ has been selected in the STDAE-S4 algorithm; d) the values of the parameters N_0 and M are equal to those employed for the STDREC in the previous paragraph; e) the oversampling factor is $\bar{M} = 16$ ($\bar{M} = 7$) for Inras FMCW (for the TI FMCW), so that the FFT orders are $\bar{N}_0 = 256$ ($\bar{N}_0 = 602$). Moreover, the following values have been selected for the parameters[16] of the FFT-BAs and the MUSIC-BAs: $C_s = 6$, $G_s = 6$ and $K_0 = 2$. Prior knowledge of L has been assumed and the threshold T_{STDAEC} has been selected in the range $[0.01, 0.9] \cdot E^{(0)}[l]$ (2.39) (the value of this threshold has been adjusted on the basis of the SNR characterizing the received signal and the overall number of detectable targets).

The estimates of range and azimuth generated by the RASCAs on the basis of the measurements acquired in our first group of experiments are listed in Table 2.8, whereas the values of RMSE, peak error and CT computed by averaging the RMSEs, peak errors and CTs evaluated in each single experiment are listed in Table 2.9. In the last table, the values of RMSE, peak error and CT for the employed FFT-BAs and MUSIC-BAs are also provided. These results lead to the following conclusions:

1. All the range and azimuth errors are comparable with the resolution of our devices.

[16] Our simulations have evidenced that small changes in the value of these parameters do not significantly influence the detection probability and the estimation accuracy of the considered algorithms.

2. The RASCAs always outperform the other algorithms and require a lower CT.

3. The highest range peak errors and RMSEs are found in the case of the Inras FMCW radar; the TI FMCW radar, instead, achieves the lowest range and azimuth errors. This is mainly due to the differences in the SNR available at the receive side of distinct radar devices is different.

The good accuracy achieved by the RASCAs is also evidenced by Fig. 2.15, where a *range-azimuth map* [53], generated though standard 2D FFT processing of the measurements acquired through the Inras FMWC radar, is represented as a contour plot[17]; in the same figure, the exact position of the five targets employed in our first group of experiments and their estimates evaluated by all the considered algorithms are shown.

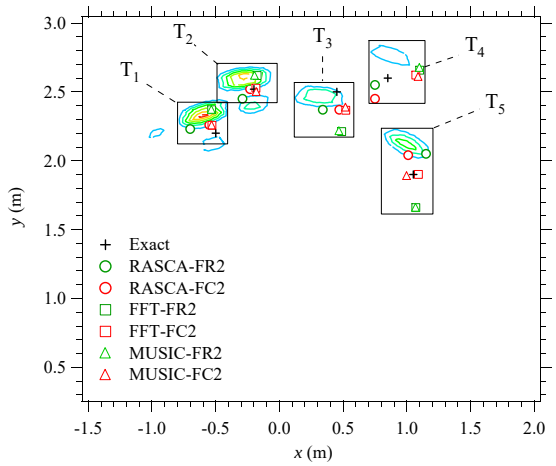


Figure 2.15: *Representation of the range-azimuth map (in $x-y$ coordinates) computed on the basis of the measurements acquired through the Inras FMCW radar in the presence of five targets. The exact position of each target and its estimate obtained through the RASCAs (shown in Table 2.8) and the other algorithms are also shown. The rectangles allow to delimit the region in which the position of each target and its estimates are located.*

[17] Note that $x-y$ coordinates are employed in this case, in place of range and azimuth; the position of the radar system corresponds to the origin of our reference system.

Table 2.8: *Exact range and azimuth of the five targets considered in our first group of experiments and corresponding estimates generated by the RASCAs.*

Exp.	Method	Params.	T_1	T_2	T_3	T_4	T_5
	Exact	R (m)	2.26	2.51	2.44	2.68	2.21
		θ (°)	-12.7	-4.5	10.6	18.0	28.3
1)	RASCA-FC2	R (m)		2.52			
		θ (°)		-5.1			
	RASCA-FR2	R (m)		2.6			
		θ (°)		-3.9			
		θ (°)	-11.8				
2)	RASCA-FC2	R (m)		2.53	2.40		
		θ (°)		-5.2	8.1		
	RASCA-FR2	R (m)		2.6	2.55		
		θ (°)		-4.3	11.3		
		θ (°)	-12.8		6.3		
3)	RASCA-FC2	R (m)	2.27	2.53	2.41		
		θ (°)	-14.0	-5.2	9.8		
	RASCA-FR2	R (m)		2.63	2.55		2.33
		θ (°)		-4.3	11.3		28.2
		θ (°)	-13.2		7.2		25.5
4)	RASCA-FC2	R (m)	2.32	2.53	2.42		2.26
		θ (°)	-13.7	-5.2	11.2		29.4
	RASCA-FR2	R (m)	2.30	2.64	2.42		2.22
		θ (°)	-12.5	-4.1	11.9		31.2
		θ (°)	-13.2		6.7	17.5	25.5
5)	RASCA-FC2	R (m)	2.31	2.53	2.42	2.56	2.27
		θ (°)	-12.4	-5.2	11.2	17.0	26.3
	RASCA-FR2	R (m)	2.34	2.47	2.39	2.66	2.36
		θ (°)	-17.3	-6.7	8.1	16.4	29.5
		θ (°)	-12.5	-7.0	4.0	17.1	27.0

Table 2.9: Root mean square error $\bar{\varepsilon}_X$, peak error $\hat{\varepsilon}_X$, *and* computation time *(CT) evaluated on the basis of our first group of measurements. Target range and azimuth are taken into consideration.*

Method	$\bar{\varepsilon}$		$\hat{\varepsilon}$		CT
	R (m)	θ (°)	R (m)	θ (°)	(sec)
RASCA-FC2	0.04	1.2	0.05	1.5	0.3
FFT-FC2	0.06	1.6	0.11	2.8	0.4
MUSIC-FC2	0.05	1.4	0.09	2.4	0.5
RASCA-FR2	0.09	1.1	0.12	1.8	0.4
FFT-FR2	0.13	1.8	0.16	2.65	0.45
MUSIC-FR2	0.13	1.6	0.16	2.49	0.45

Let us consider now on the results obtained for our second group of experiments. The estimates of range, azimuth and elevation generated by the RASCAs are listed in Table 2.11, whereas the values of RMSE, peak error and CT evaluated on the basis of this table are listed in Table 2.10. In the last table, the errors characterizing the FFT-based and MUSIC-based algorithm for 3D imaging are also provided. From these results it can be inferred that:

1. The RMSEs and the peak errors evaluated for target range, azimuth and elevation are reasonably low and comparable with those obtained in the case of 2D imaging. Moreover, these errors are smaller than the ones characterizing the FFT-based and MUSIC-based algorithms.

2. The range errors evaluated for all the radar devices are comparable; however, the lowest peak error and RMSE are achieved in the case of the TI FMCW radar.

3. The azimuth and elevation estimates computed on the basis of the measurements acquired through the TI and Inras FMCW radars are reasonably good.

4. In general, the CT of RASCAs is higher than that required by the proposed FFT-based or MUSIC-based methods. This is mainly due to the recursive cancellation procedure that is not employed by the other two methods. However, we believe that this cancellation procedure plays a fundamental role in the detection of weak targets and allows to achieve a good estimation accuracy.

The good accuracy and resolution provided by the RASCAs are highlighted by Fig. 2.16, where the exact positions of the five targets employed in our second group of experiments and their estimates produced by all the considered algorithms are shown; note that, unlike FFT-based and MUSIC-based algorithms, the RASCAs achieve good accuracy even in the presence of closely spaced targets, like T_4 and T_5.

Table 2.10: Root mean square error $\bar{\varepsilon}_X$, peak error $\hat{\varepsilon}_X$ *and* computation time *(CT) evaluated on the basis of our second group of measurements. Target range, azimuth and elevation are taken into consideration.*

Method	$\bar{\varepsilon}$			$\hat{\varepsilon}$			CT
	R (m)	θ (°)	ϕ (°)	R (m)	θ (°)	ϕ (°)	(sec)
RASCA-FC3	0.05	2.8	1.9	0.07	3.9	2.5	2.0
FFT-FC3	0.07	3.0	1.5	0.08	4.2	2.1	1.5
MUSIC-FC3	0.07	3.9	3.2	0.08	6.4	3.8	1.6
RASCA-FR3	0.06	2.0	2.3	0.08	3.0	3.0	2.6
FFT-FR3	0.1	1.0	2.1	0.17	1.5	3.2	1.1
MUSIC-FR3	0.1	1.0	2.2	0.15	1.5	3.3	1.3

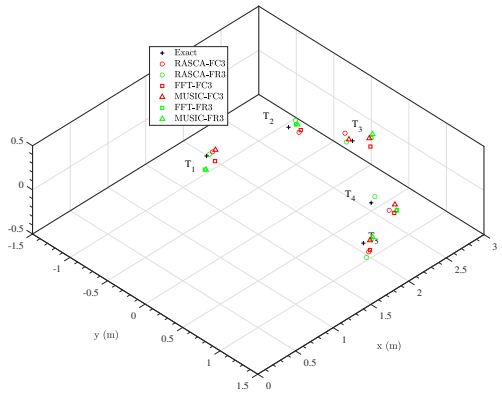


Figure 2.16: *Representation of a 3D scenario characterized by five targets. The exact position of each target and the corresponding estimates evaluated by all the considered algorithms are shown (all our radar systems are considered).*

2.9 Conclusions

In this chapter, four novel algorithms, dubbed *range & angle serial cancellation algorithms* (RASCAs), have been developed for the detection and the estimation of multiple targets in colocated MIMO FMCW radar systems. All these algorithms can be seen as instances of a general approach to target detection and estimation, and exploit new methods for the estimation of multiple overlapped real and complex tones. As evidenced by our computer simulations run on both synthetically generated data and measurements acquired through commercial devices, the devised algorithms are able to generate accurate 2D and 3D radar images in the presence of multiple closely spaced targets and outperform other algorithms based on the computation of multiple FFTs or on the MUSIC for DOA estimation.

Table 2.11: *Exact range, azimuth and elevation of the five targets considered in our second group of experiments and corresponding estimates generated by the RASCAs.*

Exp.	Method	Params.	T_1	T_2	T_3	T_4	T_5
	Exact	R (m)	1.94	2.34	2.75	2.49	2.08
		θ (°)	-27.8	-9.9	0	14.0	35.2
		ϕ (°)	-6.0	2.0	-2.1	-7.0	-2.0
1)	RASCA-FC3	R (m)	1.89				
		θ (°)	-25.8				
		ϕ (°)	-4.0				
	RASCA-FR3	R (m)	2.04				
		θ (°)	-29.3				
		ϕ (°)	-7.5				
		θ (°)	-20.9				
		ϕ (°)	-3.5				
2)	RASCA-FC3	R (m)	1.89				2.12
		θ (°)	-25.2				37.8
		ϕ (°)	-3.4				-2.3
	RASCA-FR3	R (m)	2.0				2.1
		θ (°)	-27.0				34.2
		ϕ (°)	-8.0				-2.5
		θ (°)	-21.0				28.5
		ϕ (°)	-3.5				-1.5
3)	RASCA-FC3	R (m)	1.89	2.43			2.12
		θ (°)	-25.2	-8.6			37.8
		ϕ (°)	-3.4	1.1			-3.4
	RASCA-FR3	R (m)	2.0	2.45			2.1
		θ (°)	-29	-10			31.5
		ϕ (°)	-4.0	5.0			-3.0
		θ (°)	-21.0	-9.5			29.2
		ϕ (°)	-4.0	5.0			-3.0
4)	RASCA-FC3	R (m)	1.89	2.44		2.48	2.12
		θ (°)	-25.2	-8.6		20.1	37.3
		ϕ (°)	-3.2	-0.6		-6.9	-2.9
	RASCA-FR3	R (m)	2.0	2.4		2.45	2.1
		θ (°)	-29.5	-12.0		18.3	33.5
		ϕ (°)	-4.0	5.0		5.0	-3.0
		θ (°)	-21	-9		15	29
		ϕ (°)	-3.5	0		-7	-2.5
5)	RASCA-FC3	R (m)	1.89	2.44	2.83	2.48	2.12
		θ (°)	-25.2	-8.6	3.2	20.1	37.3
		ϕ (°)	-3.2	-0.6	-2.87	-6.9	-2.9
	RASCA-FR3	R (m)	1.98	2.4	2.65	2.4	2.13
		θ (°)	-27.6	-15.0	0.4	18.3	37.3
		ϕ (°)	-5.8	5	-1.1	-3	-6.2
		θ (°)	-21.0	-9.0	2.1	14.6	29.0
		ϕ (°)	-3.5	0	-2.5	-8.2	-1.0

3. Machine Learning and Deep Learning Techniques for colocated MIMO Radar systems Techniques

3.1 Introduction

As already pointed out in the previous chapters, the full exploitation of the potentialities offered by modern colocated MIMO radar devices requires the use of proper detection and estimation methods. In the last two decades, significant advances have been made in the development of *deterministic* methods accomplishing these tasks. These are mainly based on a *maximum likelihood* approach [21, 85] or on *sub-space methods*, like the *MUltiple SIgnal Classification* (MUSIC) technique [31]. Moreover, they are *model-based*, since they require the full knowledge of the employed radar device and rely on a parametric description of the propagation environment; note that, in such a description, targets are usually represented as *points* reflecting electromagnetic energy. An overview of deterministic methods is provided by refs. [53, 58], whereas some interesting applications of them can be found in refs. [64, 86, 87]; relevant examples of these applications include the detection and the estimation of the position of cars or pedestrians in a street [50], as well as the analysis of human vital signs [88]. In many cases, these methods allow to achieve good estimation accuracy at the price of an acceptable computational effort. Unluckily, in a number of recent applications, MIMO radars operate in extremely complex, highly dynamic and time varying scenarios, affected by multipath propagation, clutter and interference, and in the presence of *extended* targets. In such conditions, deterministic algorithms may fail, since they are unable to achieve acceptable estimation accuracy and are prone to generate ghost targets [89]. When this occurs, *machine learning* (ML) and *deep learning* (DL) techniques represent an appealing alternative or the only viable technical solution. A relevant example of this class of techniques is represented by *neural networks* (NNs) [76,90]. These networks can automatically learn specific data patterns and extract useful information directly from raw data, even in the presence of strong interference. In fact, they can be trained to recognise interference and remove it, so making the recovery of useful signal components possible. Unfortunately, the application of NNs and related methods to MIMO radars is challenging, because, on the one hand, the problems tackled in this field are often substantially different from those to which such methods have been applied for a number of years (e.g., processing of RGB images in computer vision); on the other hand, the large radar dataset required for the proper training of a

NN may be unavailable. Another critical issue emerging from the exploitation of ML and DL methods in real world applications is represented by the fact that a trained machine is, by and large, a black box mapping inputs to outputs; for this reason, generally speaking, it cannot be inferred why a given output has been produced on the basis of its input data. This explains why, in various radar applications, a model-based approach could be preferred. Despite this relevant limitation, it is widely accepted that the use of ML and DL methods in colocated MIMO radars will allow to solve a number of real world problems. For instance, recent work has evidenced that they can be successfully exploited in the classification of human activities and gestures, in the detection of human falls [91] and in the classification of dynamic targets in dense and dynamic urban scenarios [92]. This aims at providing an overview of the ML and DL methods employed in all the above mentioned applications, analysing their pros and cons, discussing the main lessons that have been learnt from their use and illustrating some trends in this research area. As far as we know, in the technical literature, the few manuscripts offering related contributions refer to specific applications, namely human-motion recognition [91] and assisted living [88]. The scope of this work, instead, is offering a wider perspective on this research area. Furthermore, our description of learning methods is interspersed with various numerical examples on synthetically generated dataset and an entire section is devoted to the analysis of various numerical results generated the measurements acquired through a commercial colocated MIMO radar.

The remaining part of this chapter is organized as follows: the most relevant ML and DL methods currently being investigated for their use in colocated MIMO radars are described in Sections 3.2 and 3.3, respectively; a brief comparison among such methods is illustrated in Section 3.4. An overview of the specific applications of these techniques to colocated MIMO radars is illustrated in Section 3.5, where we focus on human-motion and human-gesture classification, healthcare monitoring, and target detection and localization in automotive scenarios. Some trends emerging in the current research activities about the application of DL techniques to colocated MIMO radars are illustrated in Section 3.6. Various ML and DL methods are compared, in terms of accuracy and computational effort, in Section 3.7, where their use in human activity classification, and in the detection and position estimation of a moving target is illustrated. Finally, some conclusions are offered in Section 3.8.

3.2 Machine Learning Techniques for Colocated MIMO Radar Systems

In this section, after illustrating the main differences between a deterministic approach and a ML-based approach to target detection and estimation, the most important ML techniques applied in the field of colocated MIMO radar systems are described. Our introduction to these techniques is based on a specific case study, involving a FMCW radar system (see Paragraph 1.3).

3.2.1 A case study

The most relevant conceptual differences between a deterministic approach and a ML approach to detection and estimation problems in MIMO radar systems can be understood by analysing the detection of a single point target, and the estimation of its range R and its azimuth ϕ in a 2D propagation scenario. In this case, we assume that an FMCW radar system equipped with an ULA, consisting of three antenna elements, is employed (see Fig. 3.1-a). This array is made of a central TX antenna and a cou-

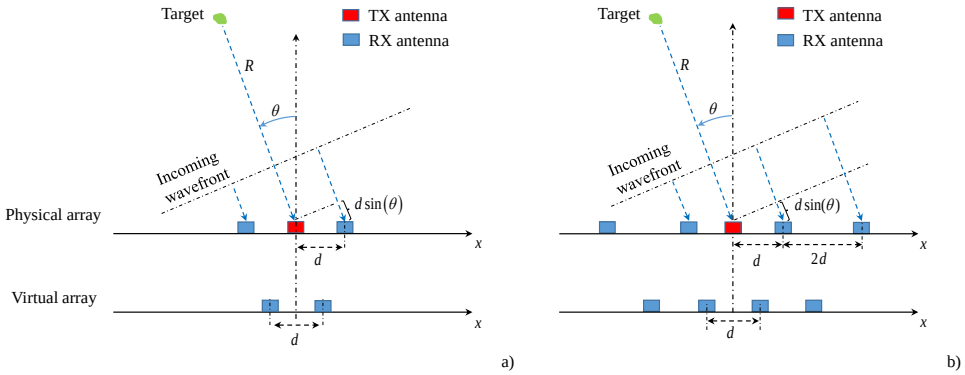


Figure 3.1: *Physical geometry and virtual array of a colocated FMCW MIMO radar equipped with an ULA composed by a single TX antenna and: a) two RX elements; b) four RX elements.*

ple of antipodal RX antennas (these are identified by a red box and two blue boxes, respectively, in the considered figure), so that $N_T = 1$ and $N_R = 2$; consequently, a *virtual* array, consisting of $N_V = 2 \cdot 1 = 2$ virtual elements, is available. The abscissa x_v of the v-th *virtual antenna element* associated with the TX antenna and the v-th RX antenna is computed as (see eq. (1.6))

$$x_v = \frac{x_t + x_{r,v}}{2} \tag{3.1}$$

with $v = 0$ and 1; here, $x_t = 0$ and $x_{r,0} = -d$ ($x_{r,1} = d$) are the abscissas of the TX and of the first (second) RX antenna, respectively (note that the origin of our reference system coincides with the center of the array). If the target is in far field[1], the wavefront of the electromagnetic echo originating from it is a straight line and is orthogonal to the line connecting the target with the center of the array. In these conditions, the n-th time-domain sample acquired on the v-th virtual antenna in a single snapshot can be expressed as (see eq. (1.9))

$$\begin{aligned} r_{v,n} &= a_v \cos\left(2\pi n f_v T_s + \psi_v\right) + w_{v,n}, & (3.2)\\ &= A_v \exp\left(j\, 2\pi n F_v\right) + A_v^* \exp\left(-j\, 2\pi n F_v\right) + w_{v,n}, & (3.3) \end{aligned}$$

for $n = 0, 1, \ldots, N-1$, where (see eqs. (1.13), (1.14) and (1.15))

$$F_v \triangleq f_v T_s, \tag{3.4}$$

[1] A rigorous definition of this condition can be found in ref. [93, Par. 2.2.4, pp. 34-36]

$$f_v = \frac{2\mu}{c}\left[R + x_v \sin(\phi)\right], \tag{3.5}$$

$$x_v = (-1)^{(v+1)}\,\frac{d}{2}, \tag{3.6}$$

d is inter-antenna spacing of the considered ULA,

$$A_v \triangleq \frac{1}{2} a_v \exp(j\psi_v) \tag{3.7}$$

is a complex parameter depending on the target reflectivity a_v and (see eqs. (1.15) and (1.16))

$$\psi_v \triangleq \angle A_v \cong \frac{4\pi}{\lambda}\left[R + x_v \sin(\phi)\right] \tag{3.8}$$

is the phase observed on the considered antenna (the wavelength λ is defined by eq. (1.20)). It is important to point out that:

a) Relevant information about the target *azimuth* are provided by the frequency difference

$$\Delta f_{0,1} \triangleq f_1 - f_0 \tag{3.9}$$

or by the phase variation

$$\Delta\psi_{0,1} \triangleq \angle A_1\, A_0^*, \tag{3.10}$$

where the quantity $\angle X$ represents the phase of the complex number X (it belongs to interval $[-\pi, \pi)$). In fact, on the one hand, from eqs. (3.5)-(3.6) it is easily inferred that (see the definition (3.9))

$$\Delta f_{0,1} = 2\frac{\mu d}{c}\sin(\phi); \tag{3.11}$$

on the other hand, based on eqs. (3.6)-(3.8), it is easy to show that (see the definition (3.10))

$$\Delta\psi_{0,1} = \psi_1 - \psi_0 = 4\pi\frac{d}{\lambda}\sin(\phi), \tag{3.12}$$

provided that the inequality

$$4\pi\frac{d}{\lambda}\left|\sin(\phi)\right| \leq \pi \tag{3.13}$$

holds for any ϕ. The last condition is met for any $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2})$ if

$$d \leq \lambda/4. \tag{3.14}$$

b) If the received signal is noiseless, the frequency f_v is known and N is large, the complex amplitude A_v can be easily estimated as[2]

$$\hat{A}_v \cong \frac{1}{N}\bar{X}_v(f_v), \tag{3.15}$$

[2]This result can be easily proved by substituting eq. (3.3) in the *right-hand side* (RHS) of the definition (3.16).

where

$$\bar{X}_v(f) \triangleq \sum_{n=0}^{N-1} r_{v,n} \exp(-j2\pi n f T_s) \tag{3.16}$$

represents the Fourier transform of the sequence $\{r_{v,n};\, n = 0,\, 1,\, ...,\, N-1\}$.

c) Information about the target *range* is provided by the *average frequency* (see eq. (3.5))

$$f_m \triangleq \frac{f_0 + f_1}{2} = \frac{2\mu}{c} R \tag{3.17}$$

Therefore, the estimation of the frequency of the sinusoid contained in the noisy data sequence acquired through each virtual antenna represents a fundamental problem in target detection and estimation. It is well known that the so called *periodogram method* can be employed to solve it in an approximate way [86, 94]. This method is based on the computation of the amplitude spectrum of the zero-padded measurement sequence and on the identification of its peak.

Based on the mathematical results and the considerations illustrated above, a simple deterministic algorithm, consisting of the three steps listed below, can be easily derived for the detection of the target and the estimation of its spatial coordinates (R, ϕ).

1. DFT *processing* - In this step, the N-dimensional vector

$$\mathbf{r}_v \triangleq [r_{v,0}, r_{v,1}, ..., r_{v,N-1}]^T, \tag{3.18}$$

with $v = 0$ and 1, undergoes *zero padding* (ZP); this results in the N_0-dimensional vector

$$\mathbf{r}_v^{(ZP)} \triangleq \left[\mathbf{r}_v^T \, \mathbf{0}_P^T\right]^T \tag{3.19}$$

where $N_0 \triangleq M_r\, N$, $\mathbf{0}_P$ denotes the P-dimensional (column) null vector and M_r represents the selected *oversampling factor* adopted in *time-domain processing*. Then, the vector $\mathbf{r}_v^{(ZP)}$ (3.19) feeds a N_0-th order *discrete Fourier transform* (DFT); this produces the N_0-dimensional vector

$$\mathbf{X}_v \triangleq [X_{v,0}, X_{v,1}, ..., X_{v,N_0-1}]^T, \tag{3.20}$$

where

$$X_{v,l} = \frac{1}{N_0} \bar{X}_v(\bar{f}_l) \tag{3.21}$$

$\bar{X}_v(f)$ is defined by eq. (3.16) and

$$\bar{f}_l \triangleq \frac{l}{N_0 T_s}, \tag{3.22}$$

with $l = 0, 1, ..., N_0 - 1$. Finally, the N_0-dimensional vector

$$\mathbf{P} \triangleq [P_0, P_1, ..., P_{N_0-1}]^T, \tag{3.23}$$

where

$$P_l \triangleq \frac{M_r^2}{2} \left[|X_{0,l}|^2 + |X_{1,l}|^2\right], \tag{3.24}$$

with $l = 0, 1, ..., N_0 - 1$, is computed; note that the quantity P_l (3.24) represents a sort of *average power spectrum* evaluated at the frequency $\bar{f}_l$ (3.22).

2. *Target detection* - The problem

$$\hat{l} = \arg \max_{\tilde{l} \in \{0,1,...,N_0/2\}} P_l \tag{3.25}$$

is solved and a target is detected if the condition

$$P_{\hat{l}} > P_{th} \tag{3.26}$$

is satisfied, where P_{th} is a proper threshold. When this occurs, the next step is executed; otherwise, the algorithm stops.

3. *Estimation of target coordinates* - The estimate

$$\hat{f}_m = \frac{\hat{l}}{N_0 T_s} \tag{3.27}$$

of the frequency f_m (3.17) and the estimate

$$\hat{A}_v = M_r \, X_{v,\hat{l}} \tag{3.28}$$

of the complex amplitude A_v (3.7) (with $v = 0$ and 1) are computed. Then, the estimate (see eq. (3.17))

$$\hat{R} = \hat{f}_m \frac{c}{2\mu} \tag{3.29}$$

of the target range R and the estimate (see eq. (3.12))

$$\hat{\phi} = \arcsin\left(\frac{\lambda}{4\pi d} \Delta\hat{\psi}_{0,1}\right) \tag{3.30}$$

of the target azimuth θ are evaluated; here,

$$\Delta\hat{\psi}_{0,1} = \angle X_{1,\hat{l}} \left(X_{0,\hat{l}}\right)^* . \tag{3.31}$$

represents an estimate of $\Delta\psi_{0,1}$ (3.12) and its expression is based on eqs. (3.15), (3.21) and (3.28).

This concludes the description of the proposed detection and estimation algorithm. It is important to point out that:

a) The accuracy achievable in range estimation is influenced by the DFT order N_0 and, consequently, for a given N, by the oversampling factor M_r. Increasing the value of the parameter M_r leads to a more refined analysis of the spectrum $\bar{X}_v(f)$ (3.16) and, consequently, allows to locate the spectral peak originating from the target with better accuracy; however, this result is achieved at the price of an higher computational cost.

b) The estimate $\hat{\phi}$ (3.30) is *unambiguous* if the condition (3.14) is satisfied or if, for a given $d > \lambda/4$, the azimuth ϕ belongs to the interval $[-\phi_m, \phi_m)$, where (see eq. (3.13))

$$\phi_m \triangleq \arcsin\left(\frac{\lambda}{4\,d}\right) \tag{3.32}$$

c) Eq. (3.11) has not been exploited to compute an estimate of the target azimuth. This is due to the fact that the quality of this estimate is limited by the accuracy of frequency estimation on each antenna; such an accuracy, in turn, is intrinsically limited by the DFT order N_0.

d) If the target reflectivity observed on the two antennas is approximately the same (i.e., if $a_0 \cong a_1$), an estimate of it can be computed as (see eqs. (3.7) and (3.28))

$$\hat{a} \triangleq M_r \left[\left| X_{0,\hat{l}} \right| + \left| X_{1,\hat{l}} \right| \right]. \tag{3.33}$$

e) The estimation of the azimuth characterizing the echo from a specific target requires at least two RX antennas, since it is based on computation of the phase variation observed at a specific frequency on at least two receive antennas (see eqs. (3.30) and (3.31)).

f) The maximum number of detectable targets depends on the number of virtual elements of the whole array. It is worth noting that, unlike a phased array system, where a single waveform is transmitted, a MIMO radar system endowed with N_T different TX antennas can radiate N_T independent signals. This leads to the conclusion that the maximum number of targets that be can uniquely identified by a MIMO radar is N_T times larger than that of its counterpart employing a phased array [8], if the first system employs an ULA whose virtual elements do not overlap (like the ULAs shown in Fig. 3.1).

The estimation accuracy achieved by the considered radar system can be improved by increasing the size of its ULA, i.e. the overall number of its antennas, so that a larger number of virtual channels becomes available. For instance, if the ULA shown in Fig. 3.1-a) is replaced by the one represented in Fig. 3.1-b) (and characterized by $N_T = 1$ and $N_R = 4$), $N_V = 4$ virtual channels become available, i.e. the overall number of virtual antennas is doubled with respect to the previous case. Note that this results not only in an increase of the maximum number of detectable targets, but also in an improvement of the *angular resolution* $\Delta\phi$, defined as the minimum angular separation below which the DOAs of two distinct targets cannot be separated. More specifically, if an ULA is used and the bore-sight direction is considered, we have that (e.g., see [64, Par. 4, eq. (51)])

$$\Delta\phi = \frac{\lambda}{2\,d\,\left(N_V - 1\right)\cos(\phi)}. \tag{3.34}$$

It is also worth mentioning that the algorithm illustrated above for a couple of virtual channels can be easily extended to the case of an ULA providing N_V virtual channels. The only relevant modification concerns step 3., since the N_V-dimensional vector

$$\hat{\mathbf{A}} = \left[\hat{A}_0, \hat{A}_1, ..., \hat{A}_{N_V-1} \right]^T, \tag{3.35}$$

where $\hat{A}_v$ is still expressed by eq. (3.28) for any v, becomes available and, consequently, $(N_V - 1)$ phase variations, referring to the $(N_V - 1)$ distinct couples of adjacent virtual antennas can be evaluated. If we assume that the variations of the target reflectivity over the whole virtual array are negligible and that the SNR on each virtual antenna is high, such variations are approximately constant, being all expressed by the RHS

of eq. (3.12). This means that a *phase modulation*, characterized by the *normalised spatial frequency*

$$F = 2\frac{d}{\lambda}\sin(\phi), \tag{3.36}$$

is observed in the sequence $\{\hat{A}_v;\, v = 0,\, 1,\, ...,\, N_V - 1\}$. An estimate of the parameter F can be computed by exploiting, once again, the *periodogram method*. In practice, this requires executing the following three steps:

1. DFT *processing* - The vector $\hat{\mathbf{A}}$ is zero padded by appending to it a null vector of size $(M_A - 1)N_V$, where M_A represents the *oversampling factor* adopted in *spatial processing*; this produces the $\bar{N}_0$-dimensional vector $\hat{\mathbf{A}}_{ZP}$, where $\bar{N}_0 \triangleq M_A N_V$. The vector $\hat{\mathbf{A}}_{ZP}$ feeds a $\bar{N}_0$-th order DFT, generating the $\bar{N}_0$-dimensional vector

$$\mathbf{s} \triangleq \left[s_0, s_1, ..., s_{\bar{N}_0/2}, s_{-\bar{N}_0/2+1}, s_{-\bar{N}_0/2+2}, ..., s_{-2}, s_{-1}\right]^T. \tag{3.37}$$

2. *Azimuth estimation* - After solving the problem

$$\hat{p} = \arg \max_{\tilde{p} \in \{-\bar{N}_0/2+1, -\bar{N}_0/2+2, ..., \bar{N}_0/2\}} |s_{\tilde{p}}|, \tag{3.38}$$

the estimate (see eq. (3.30))

$$\hat{\phi} = \arcsin\left(2\frac{\hat{p}}{\bar{N}_0}\right) \tag{3.39}$$

of the target azimuth ϕ is evaluated. Note that the angular resolution provided by the DFT computed in step 1. improves as $\bar{N}_0$ increases.

The deterministic algorithm and its extension illustrated above have the following relevant properties: a) their derivation is based on a well defined mathematical model originating from our knowledge of the propagation of electromagnetic waves and of the radar system (and, in particular, of the geometry of its array and of the processing it accomplishes); b) if they fail detecting a given target, or generate inaccurate estimates of its range and/or azimuth, the causes of such events can be identified; c) being based on the DFT and other simple formulas, they are computational efficient.

An alternative to the approach to algorithm design illustrated above is offered by ML methods [76]. In fact, if such methods are employed, the inner structure of the considered radar system and the physical laws on which its operation is based can be ignored, since the required information are automatically extracted by an algorithm able to *learn the regularities characterizing the set of observed data*. Let us reconsider now the detection and estimation problem described above from this new perspective and show how a solution based on ML methods can be devised. To this aim, we take into consideration again a FMCW radar system equipped with the antenna array shown in Fig. 1.2-a) and assume that it is employed to perform a measurement campaign. In this campaign, N_t independent trials are accomplished in the presence of a single point target or in the absence of it; in each trial, the couple $[\mathbf{r}_0,\, \mathbf{r}_1]$ of noisy vectors (see eq. (3.18)) is acquired and stored in a database together with the target range and azimuth (when the target is present). In the following, $[\mathbf{r}_{q,0}, \mathbf{r}_{q,1}]$ and

$$\mathbf{t}_q \triangleq [d_q, R_q, \phi_q]^T \tag{3.40}$$

denote the value of the couple $[\mathbf{r}_0,\ \mathbf{r}_1]$ and the associated *label* acquired in the q-th trial (with $q = 0,\ 1,\ ...\ ,\ N_t - 1$); here, $d_q = -1$ (1) if the target is absent (present), and R_q and ϕ_q represent the target range and azimuth, respectively, in the same trial if $d_q = 1$ (if $d_q = -1$, the values of both R_q and ϕ_q are irrelevant). In this case, the ML approach consists in processing the dataset

$$\mathscr{D}_i \triangleq \{[\mathbf{r}_{q,0},\ \mathbf{r}_{q,1}]\ ,\ \mathbf{t}_q; q = 0, 1, ..., N_t - 1\} \tag{3.41}$$

to learn how to detect the presence of a target on the basis of a new couple $[\mathbf{r}_0,\ \mathbf{r}_1]$ and, if a target is detected, how to estimate its position. The accuracy of the algorithm resulting from the learning phase (i.e., from *training*) depends not only on the adopted ML method, but also on the size of $\mathscr{D}_i$ (i.e., on N_t). Generally speaking, the use of ML methods requires the availability of a large set of measurements, i.e. a large N_t (say, at least, a few thousands). Unluckily, any ML method extracting the required knowledge directly from $\mathscr{D}_i$ (3.41) has to process high dimensional vectors if the size N of the vectors $\mathbf{r}_0$ and $\mathbf{r}_1$ is large. Actually, the dimensionality of the given problem can be easily reduced by exploiting our prior knowledge. In fact, in developing our deterministic algorithm, we have learnt that essential information for target detection and estimation is provided by the complex couple $[X_{0,\hat{l}},\ X_{1,\hat{l}}]$ (see eqs. (3.26)-(3.31)), where $\hat{l}$ is expressed by eq. (3.25). These considerations suggest to:

a) *Pre-process* the couple $(\mathbf{r}_{q,0},\ \mathbf{r}_{q,1})$ in order to generate the vector

$$\mathbf{X}_q \triangleq \left[X_0^{(q)}, X_1^{(q)}\right]^T, \tag{3.42}$$

where $X_0^{(q)}$ and $X_1^{(q)}$ are the values taken on by the quantities $X_{0,\hat{l}}$ and $X_{1,\hat{l}}$, respectively, in the q-th trial (with $q = 0,\ 1,\ ...,\ N_t - 1$); $X_0^{(q)}$ and $X_1^{(q)}$ can be considered as highly informative data extracted from the received signal, i.e., briefly, as the *features* available in the considered problem.

b) Replace the set $\mathscr{D}_i$ (3.41) with the new set

$$\mathscr{D} \triangleq \{\mathbf{X}_q,\ \mathbf{t}_q; q = 0, 1, ..., N_t - 1\}, \tag{3.43}$$

that consists of low dimensional vectors only, and use it to train the considered ML method; when this occurs, the last set is called *training set*.

Once training is over, the ML algorithm resulting from it is able to infer the unknown value of $\mathbf{t}_q$ (3.40) for any new vector $\mathbf{X}_q$ (3.42) of noisy data (with $q > N_t - 1$); in other words, it is able to predict: a) d_q; b) R_q and ϕ_q if a target is detected. It is important to point out that any ML algorithm predicting d_q solves a *binary classification problem*, since it assigns a new observation to one of two *categories* of noisy data, one associated with the presence of a target, the other one with its absence; in other words, *the algorithm is exploited to recognise a specific pattern in the noisy observations*. If the considered ML algorithm is also able to *predict* the value of the couple (R_q, ϕ_q) (i.e., of two continuous variables), it solves a *regression problem* too. In the considered radar system, different ML algorithms can be employed to learn classification and regression

rules from the training set $\mathcal{D}$ (3.43). Moreover, all such algorithms can be considered as specific instances of the so called *supervised learning*, as shown in the following paragraph. Generally speaking, supervised learning techniques can be employed when:

1. A training set
$$\mathcal{D} \triangleq \{(\mathbf{r}_q, \mathbf{t}_q)\,;\, q = 0, 1, ..., N_t - 1\}, \tag{3.44}$$
collecting N_t, D_r-dimensional real *observations* (also called *covariates*, or *domain points*, or *explanatory variables*) $\{\mathbf{r}_q; q = 0,\ 1,\ ...,\ N_t - 1\}$, with
$$\mathbf{r}_q \triangleq [r_{q,0}, r_{q,1}, ..., r_{q,D_r-1}]^T, \tag{3.45}$$
and the associated D_t-dimensional real *labels* (also called *dependent variables* or *responses*) $\{\mathbf{t}_q;\ q = 0,\ 1,\ ...,\ N_t - 1\}$, with
$$\mathbf{t}_q \triangleq [t_{q,0}, t_{q,1}, ..., t_{q,D_t-1}]^T, \tag{3.46}$$
is available.

2. There exists some mechanism relating the variable $\mathbf{r}_q$ to the variable $\mathbf{t}_q$ for any q.

The last point is a fundamental one, since it does not make any sense to develop rules applicable to unseen examples in the absence of some assumptions about the mechanism underlying data generation; the set of these assumptions is known as the *inductive bias*.

3.2.2 Supervised learning

Supervised learning is a branch of ML frequently employed in the field of colocated MIMO radars for solving classification and regression problems. In this paragraph, we discuss some basic methods for supervised learning and analyse the approaches adopted in their derivation. Specific attention is paid to some classification algorithms that can be easily employed for target detection in radar systems.

Formulation of the supervised learning problem

Generally speaking, supervised learning concerns the identification of the conditional *probability density function* (pdf) $f(\mathbf{t}|\mathbf{r})$ (also called *predictive distribution*) minimizing the average *generalization loss*
$$L_p(\tilde{\mathbf{t}}) \triangleq \mathrm{E}_{f(\mathbf{t},\mathbf{r})} \left\{ \ell \left(\mathbf{t}, \tilde{\mathbf{t}} \left(\mathbf{r} \right) \right) \right\}; \tag{3.47}$$
here, $\mathrm{E}_{f(\mathbf{x})} \{\cdot\}$ denotes the expectation evaluated with respect to the joint $f(\mathbf{x})$, $\tilde{\mathbf{t}}(\mathbf{r})$ is a *prediction* of the label $\mathbf{t}$ computed on the basis of the observation $\mathbf{r}$ and $\ell(\cdot, \cdot)$ is a given *cost function*. If the label of each observation is *one-dimensional* (1D) and is real, the cost functions
$$\ell_2 \left(t, \hat{t} \right) = (t - \hat{t})^2 \tag{3.48}$$

and

$$\ell_0\left(t,\hat{t}\right)=\begin{cases}1 \text{ if } t=\hat{t}\\ 0 \text{ elsewhere}\end{cases} \tag{3.49}$$

are often employed for regression and binary classification, respectively. It is well known that, if the posterior pdf $f(\mathbf{t}|\mathbf{r})$ is *known*, the minimum value of the loss $L_p(\tilde{\mathbf{t}})$ (3.47) is achieved by selecting the *optimal prediction* (e.g., see [76, Par. III.C, eq. (4)])

$$\hat{t}\left(\mathbf{r}\right)=\arg\min_{\tilde{\mathbf{t}}}\mathrm{E}_{f(\mathbf{t}|\mathbf{r})}\left[\ell\left(\mathbf{t},\tilde{\mathbf{t}}\right)|\mathbf{r}\right], \tag{3.50}$$

whatever cost function is selected.

Supervised learning methods are employed when the conditioned pdf $f(\mathbf{t}|\mathbf{r})$ (or the joint pdf $f(\mathbf{t},\mathbf{r})$) is unknown, but a training set $\mathscr{D}$, collecting N_t distinct data generated on the basis of it and structured according to eq. (3.43), is available. The objective of these methods is to generate a *predictor*, denoted $\hat{t}_{\mathscr{D}}(\mathbf{r})$, exclusively based on the set $\mathscr{D}$ and whose performance, in terms of generalization loss, is as close as possible to that of the optimal predictor $\hat{t}(\mathbf{r})$ (3.50); this means that the loss evaluated for the prediction of the label associated with a new observation should be as small as possible. The derivation of the predictor $\hat{t}_{\mathscr{D}}(\mathbf{r})$ can be formulated as an optimization problem, whose form depends on the specific assumptions we make about the model that is being learnt. In fact, a *frequentist approach* or a *Bayesian approach* can be adopted, as illustrated in the following two paragraphs.

The frequentist approach to supervised learning

The frequentist approach relies on the assumption that all the points of the set $\mathscr{D}$ (3.44) are generated *independently* on the basis of the same *unknown* joint pdf $f(\mathbf{r},\mathbf{t})$, that is

$$\left(\mathbf{r}_q,\mathbf{t}_q\right)\sim f\left(\mathbf{r},\mathbf{t}\right)=f\left(\mathbf{t}|\mathbf{r}\right)\,f\left(\mathbf{r}\right), \tag{3.51}$$

with $q=0,1,...,N_t-1$. Under this assumption, two possible approaches can be adopted to derive the above mentioned predictor $\hat{t}_{\mathscr{D}}(\mathbf{r})$, namely: a) *separate learning* and *inference*; b) direct inference via *empirical risk minimization* (ERM). The first approach consists in learning an approximation, denoted $f_{\mathscr{D}}(\mathbf{t}|\mathbf{r})$, of the conditional pdf $f(\mathbf{t}|\mathbf{r})$, and in using the former pdf in place of the latter one to derive the expression of the predictor $\hat{t}_{\mathscr{D}}(\mathbf{r})$ on the basis of eq. (3.50). The second approach, instead, aims at directly learning $\hat{t}_{\mathscr{D}}(\mathbf{r})$ by solving the problem

$$\hat{t}_{\mathscr{D}}\left(\mathbf{r}\right)=\arg\min_{\tilde{\mathbf{t}}}L_{\mathscr{D}}\left(\tilde{\mathbf{t}}\left(\mathbf{r}\right)\right), \tag{3.52}$$

where

$$L_{\mathscr{D}}\left(\tilde{\mathbf{t}}\left(\mathbf{r}\right)\right)\triangleq\frac{1}{N_t}\sum_{q=0}^{N_t-1}\ell\left(\mathbf{t}_q,\tilde{\mathbf{t}}\left(\mathbf{r}_q\right)\right) \tag{3.53}$$

is the so called *empirical loss*. In both cases, the optimization of a *set of parameters* characterizing the model selected for the conditional pdf $f_{\mathscr{D}}(\mathbf{t}|\mathbf{r})$ or that chosen for the predictor $\hat{t}_{\mathscr{D}}(\mathbf{r})$ is required. However, the first approach is more flexible than the second one since, in principle, the approximate pdf $f_{\mathscr{D}}(\mathbf{t}|\mathbf{r})$ it generates can be exploited to

derive the predictor $\hat{t}_{\mathscr{D}}(\mathbf{r})$ for any cost function; on the contrary, the solution of the problem (3.52) holds for a specific cost function only. Moreover, it should be kept in mind that, if the first approach is adopted, two options are available. The first option consists in learning a *discriminative probabilistic model*, i.e. in learning directly an approximation of the posterior $f(\mathbf{t}|\mathbf{r})$. On the contrary, the second option consists in learning a *generative probabilistic model*, i.e. in learning the joint pdf $f(\mathbf{t}, \mathbf{r})$ and, then, in deriving an estimate of the posterior $f(\mathbf{t}|\mathbf{r})$ from it.

Let us see now how the general principles illustrated above can be employed to solve a specific regression problem concerning the first FMCW radar system described in the previous paragraph and equipped with the array shown in Fig. 1.2-a). In this case, we assume: a) the presence of a single point target placed at a fixed and known range R; b) the availability of the synthetically generated dataset[3] (see eq. (3.44))

$$\mathscr{D} \triangleq \{r_q,\, t_q; q = 0, 1, ..., N_t - 1\}, \tag{3.54}$$

where[4]

$$t_q \triangleq \phi_q, \tag{3.55}$$

$$r_q = \Delta\hat{\psi}_q \tag{3.56}$$

is an estimate of the phase difference

$$\Delta\psi_q \triangleq \psi_{q,1} - \psi_{q,0} \tag{3.57}$$

and $\psi_{q,0}$ ($\psi_{q,1}$) is the phase of the sinusoidal oscillation associated with the considered target and observed on the first (second) RX antenna for any q (see eqs. (3.2)-(3.8) and (3.12)). Moreover, in generating our dataset, the following choices have been made:

a) the distance d between adjacent virtual channels is equal to $\lambda/4$;

b) the target range R is equal to 3.0 m, whereas the target azimuth ϕ_q is uniformly distributed over the interval $[\phi_m,\ \phi_M] = [-60°\ 60°]$, respectively, for any q (this interval is comparable to the horizontal FOV of a real radar system);

c) the amplitude $a_v^{(q)}$ characterizing the sinusoid observed on the v-th virtual antenna is randomly selected in the interval [0.4, 1.2] V for any q (see eq. (1.9));

d) the random variable $a_v^{(q)}$ is independent of $a_u^{(p)}$ for any $u \neq v$ and/or $p \neq q$;

e) the observation r_q (3.56) is generated on the basis of eqs. (3.31) and (3.42), i.e. as $\Delta\hat{\psi}_q = \angle X_0^{(q)}\,(X_1^{(q)})^*$ for any q.

Moreover, the following choices have been made for the parameters of the radar system:

a) the generated frequency modulated waveform is characterized by $\mu = 7.8125{\cdot}10^{12}$ Hz s^{-1}, $T = 256\,\mu$s and $T_R = 64\,\mu$s;

b) the sampling period employed at the receive side is $T_s = 0.25$ μs and $N = 512$ time-domain samples are acquired from each of the two RX antennas;

c) the standard deviation of the noise affecting these samples is $\sigma_w = \sqrt{2}$ V (see eq. (1.9));

[3]This dataset and all the other synthetic datasets processed in our work have been generated by resorting to various functions available in the MATLAB and/or Python environment.

[4]Note that, in this case, $d_q = 1$ and $R_q = R$ in eq. (3.40), so that the label $\mathbf{t}_q$ turns into a scalar.

d) the oversampling factor $M_r = 4$ and the threshold $P_{th} = 0.5$ V^{2}Hz^{-1} are employed by the detection algorithm based on eqs. (3.25)-(3.26).

In this case, our objective is to derive a predictor of the the azimuth ϕ_q associated with the new observation $\Delta\hat{\psi}_q$ for any $q > N_t - 1$. To solve this problem, we adopt the discriminative probabilistic model

$$f\left(t|r_q, \mathbf{w}\right) = \mathcal{N}\left(t; \mu\left(r_q, \mathbf{w}\right), \beta^{-1}\right), \tag{3.58}$$

where

$$\mu\left(r_q, \mathbf{w}\right) \triangleq \sum_{j=0}^{M} w_j \, \varphi_j\left(r_q\right) = \mathbf{w}^T \boldsymbol{\varphi}\left(r_q\right), \tag{3.59}$$

M is the *order* of the model,

$$\mathbf{w} \triangleq \left[w_0, w_1, ..., w_M\right]^T \tag{3.60}$$

is a vector collecting $M + 1$ distinct real parameters (called *weights*),

$$\boldsymbol{\varphi}\left(r_q\right) \triangleq \left[\varphi_0(r_q), \varphi_1(r_q), ..., \varphi_M(r_q)\right]^T \tag{3.61}$$

is the so called *feature* vector, $\{\varphi_j(x);\, j = 0,\, 1,\, ...,\, M\}$ are $M + 1$ non linear functions and β^{-1} is the variance of the noise affecting the labels. In the following, we assume that

$$\varphi_j(x) = x^j \tag{3.62}$$

for $j = 0,\, 1,\, ...,\, M$; consequently, eq. (3.59) becomes

$$\mu(r_q, \mathbf{w}) \triangleq w_0 + \sum_{j=1}^{M} w_j \, r_q^j. \tag{3.63}$$

It is worth noting that:

a) Adopting the probabilistic model (3.58) with the mean $\mu(r_q, \mathbf{w})$ (3.63) is tantamount to postulating a polynomial dependence of the label ϕ_q on the corresponding observation $\Delta\hat{\psi}_q$.

b) The selected model depends on its order M and on the $(M + 2)$-dimensional parameter vector $\boldsymbol{\theta} \triangleq \left[\mathbf{w}^T, \beta\right]^T$.

c) The parameter M defines the number of degrees of freedom available in the model and, consequently, determines its *bias*.

As far as the last point is concerned, it is important to mention that, if M is too small, the resulting predictor may *underfit* the observations, since it is unable to accurately represent this dependence on their labels. On the contrary, if M is too large, the model is able to account for the observations of the training set, but it may generate inaccurate predictions; in other words, it memorizes the training set, but it is unable to generalise what has learnt to new examples. The last problem is known as *overfitting*. For instance, in the considered problem, good results are obtained if $M = 3$ is selected.

If the ERM approach is adopted to adjust the parameters of the probabilistic model (3.58) (and, in particular, the weight vector $\mathbf{w}$ (3.60)) in an optimal fashion, the obtained result depends on the selected cost function and cannot be always put in a closed

form. However, if the cost function $\ell_2(t, \hat{t})$ (3.48) is chosen and noise is neglected (i.e., β^{-1} is assumed to be very small), a closed form expression can be derived for $\hat{\mathbf{w}}$ for any M. In fact, under these assumptions, it can be proved that:

1. The optimal predictor $\hat{t}_{\mathscr{D}}(\mathbf{r})$ (3.52) becomes (e.g., see [75, Sect. 3.1.1, eq. (3.20)])

$$\hat{t}_{\mathscr{D}}(r_q) = \mu\left(r_q, \hat{\mathbf{w}}\right), \tag{3.64}$$

where

$$\hat{\mathbf{w}} = \arg\min_{\tilde{\mathbf{w}}} L_{\mathscr{D}}\left(\tilde{\mathbf{w}}\right), \tag{3.65}$$

$\tilde{\mathbf{w}}$ denotes a trial value of $\mathbf{w}$ and

$$L_{\mathscr{D}}\left(\tilde{\mathbf{w}}\right) \triangleq \frac{1}{N_t} \sum_{q=0}^{N_t-1} \left(t_q - \mu\left(r_q, \hat{\mathbf{w}}\right)\right)^2 \tag{3.66}$$

is the empirical loss (see eq. (3.53)).

2. The solution of the minimization problem appearing in the RHS of eq. (3.65) is

$$\hat{\mathbf{w}} = \left(\boldsymbol{\Phi}_{\mathscr{D}}^T \boldsymbol{\Phi}_{\mathscr{D}}\right)^{-1} \boldsymbol{\Phi}_{\mathscr{D}}^T \boldsymbol{t}_{\mathscr{D}}, \tag{3.67}$$

where

$$\boldsymbol{\Phi}_{\mathscr{D}} \triangleq \left[\boldsymbol{\varphi}(r_0), \boldsymbol{\varphi}(r_1), ..., \boldsymbol{\varphi}(r_{N_t-1})\right] \tag{3.68}$$

is a $N_t \times (M+1)$ matrix and

$$\mathbf{t}_{\mathscr{D}} \triangleq \left[t_0, t_1, ..., t_{N_t-1}\right]^T. \tag{3.69}$$

Given the weight vector $\hat{\mathbf{w}}$ (3.67), the estimate

$$\hat{\beta}^{-1} \triangleq \frac{1}{N_t} \sum_{q=0}^{N_t-1} \left(t_q - \hat{\mathbf{w}}^T \boldsymbol{\varphi}\left(r_{\boldsymbol{q}}\right)\right)^2, \tag{3.70}$$

of the noise variance β^{-1} can be easily evaluated.

Training the algorithm illustrated above consists in computing the weight vector $\hat{\mathbf{w}}$ (3.67) on the basis of the available training set $\mathscr{D}$ (3.54). Once training has been carried out, the generalization capability of the resulting algorithm can be assessed by evaluating the *empirical loss* (3.53) on the basis of a different dataset, called *test set* $\mathscr{D}_{ts}$ and collecting $\bar{N}_t$ observations generated in the same way as the ones of $\mathscr{D}$, but in an independent fashion.

In our computer simulations, the training set $\mathscr{D}$ (3.54) and the test set $\mathscr{D}_{ts}$ consist of $N_t = 200$ and $\bar{N}_t = 25$ observations, respectively; the points of these sets are represented in Figs. 3.2 and 3.3, respectively. First, the weight vector $\hat{\mathbf{w}}$ (3.67) and the estimate $\hat{\beta}^{-1}$ (3.70) of the noise variance have been computed on the basis of $\mathscr{D}$. Then, the accuracy of the resulting regression algorithm has been assessed on $\mathscr{D}_{ts}$. The predictions associated with the points of $\mathscr{D}_{ts}$ are represented in Fig. 3.3; in this figure, two (red) straight lines, generated on the basis of the linear equations

$$t = \mu(r, \hat{\mathbf{w}}) \pm \hat{\beta}^{-1/2}, \tag{3.71}$$

are also shown to highlight the meaning of the noise standard deviation $2\hat{\beta}^{-1/2}$. These results lead to the conclusion that, in the considered scenario, the developed regression method is able to predict the azimuth of a target with good accuracy. This is confirmed by the fact that the empirical loss computed over the set $\mathscr{D}_{ts}$ (i.e., the *generalization loss*) is close to the empirical loss $L_{\mathscr{D}}(\hat{\mathbf{w}})$ evaluated over the set $\mathscr{D}$ (see eq. (3.53)); in fact, the *root mean square error*[5] (RMSE) evaluated over $\mathscr{D}$ is equal to[6] 1.7°, whereas that computed over $\mathscr{D}_{ts}$ is equal to 1.3°.

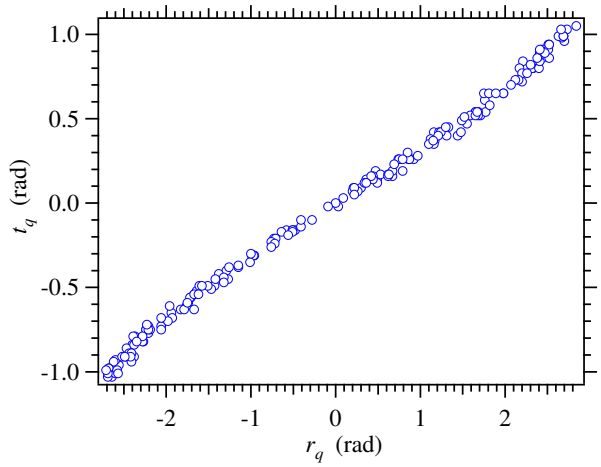


Figure 3.2: *Representation of the points of the synthetically generated training set $\mathscr{D}$ (3.54); $N_t = 200$ is assumed.*

In general, if the *discriminative probabilistic model* adopted to solve a specific regression problem is represented by the parametric pdf $f(\mathbf{t}|\mathbf{r},\boldsymbol{\theta})$, a closed form expression for the optimal value

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} L_{\mathscr{D}}\left(\tilde{\boldsymbol{\theta}}\right), \tag{3.72}$$

of the D_θ-dimensional parameter vector $\boldsymbol{\theta}$ is unavailable in most cases. When this occurs, iterative optimization techniques, like the *stochastic gradient descent* (SGD) method, can be employed to estimate $\hat{\boldsymbol{\theta}}$. The application of the SGD to the considered problem leads easily to the recursive equation

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \hat{\boldsymbol{\theta}}^{(i)} + \gamma^{(i+1)} N_S^{-1} \sum_{q\in\mathscr{S}} \nabla_{\tilde{\boldsymbol{\theta}}}\, \ell\left(\mathbf{t}_q, \hat{\mathbf{t}}\left(\mathbf{r}_q,\ \tilde{\boldsymbol{\theta}}\right)\right)|_{\tilde{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}^{(i)}}, \tag{3.73}$$

with $i = 0, 1, ..., N_E - 1$; here, $\hat{\boldsymbol{\theta}}^{(i)}$ denotes estimate of $\hat{\boldsymbol{\theta}}$ computed in the i-th recursion, $\mathscr{S}$ is a set of N_S integers randomly selected in the set $\{0, 1, ..., N_t - 1\}$ (with $N_S < N_t$), $\gamma^{(i)}$ is the *learning rate* adopted in the i-th iteration and $\nabla_{\mathbf{x}} f(\mathbf{x})$ denotes the gradient of the function $f(\mathbf{x})$. It can be proved that, if the learning rate schedule (i.e., the sequence $\{\gamma^{(i)}\}$) satisfies the so called *Robbins-Monro conditions*, the SGD converges to the optimal solution, provided that the function $L_{\mathscr{D}}(\tilde{\boldsymbol{\theta}})$ is strictly convex. The initial value $\hat{\boldsymbol{\theta}}^{(0)}$ can be either randomly selected or it can be inherited from the

[5]This parameter represents the square root of the empirical loss.

[6]The RMSE computed over $\mathscr{D}$ is given by $\hat{\beta}^{-1/2}$.

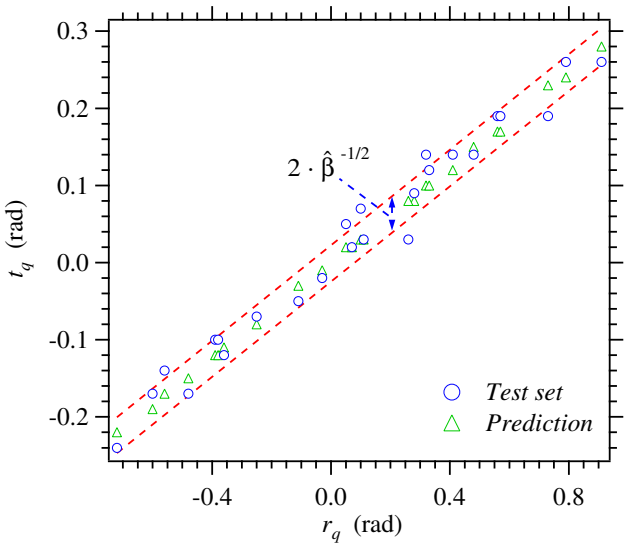


Figure 3.3: *Representation of the points of the synthetically generated test set $\mathcal{D}_{ts}$ (blue circles) and of the corresponding predictions (green triangles) evaluated on the basis of eq. (3.64); $\bar{N}_t = 25$ is assumed. Two straight lines, expressed by eq. (3.71), are also shown.*

training procedure accomplished another model; the last solution represents a specific application of the so called *transfer learning* technique (see Paragraph 3.6.1). Iterations are stopped when negligible variations are observed in the estimates generated by consecutive recursions or an upper limit set on the overall number of recursions is reached. Once the final estimate of $\hat{\boldsymbol{\theta}}$ has been computed on the basis of the available training set, the generalization capability of the resulting algorithm can be assessed by evaluating the *empirical loss* (3.53) on a given *test set* $\mathcal{D}_{ts}$.

Finally, it is worth mentioning that the selection of the parameter D_θ (i.e., of the *model complexity*) plays a fundamental role in the considered problem. In fact, if its value is too small (too large), the resulting regression method can suffer from underfitting (overfitting). The overfitting phenomenon is usually prevented by including a *regularization* term in the training of the adopted model. For instance, if the optimization problem (3.72) is considered, this result can be achieved by adopting the cost function

$$L_D\left(\tilde{\boldsymbol{\theta}}\right) + \frac{\lambda}{N_t}\left\|\tilde{\boldsymbol{\theta}}\right\|^2, \tag{3.74}$$

where λ is a real positive weight influencing the predictive capability of the resulting solution and $||\mathbf{x}||$ is the *Euclidean norm* of the vector $\mathbf{x}$.

The Bayesian approach to supervised learning

The frequentist approach illustrated in the previous paragraph leads to the identification of a specific probabilistic model through the estimation of its parameter vector $\boldsymbol{\theta}$. The Bayesian approach, instead, consists in formulating our uncertainty about the parameters of the adopted probabilistic model in statistical terms, i.e. in treating its parameter vector $\boldsymbol{\theta}$ as a *random vector*. In this paragraph, we show how the specific

regression problem analysed in the previous paragraph can be tackled from this perspective; for this reason, we assume that each observation and its label are 1D (i.e., $D_t = D_r = 1$), so that all the labels of the training set $\mathscr{D}$ (3.44) and the associated observations can be collected in the N_t-dimensional vectors $\mathbf{t}_{\mathscr{D}}$ (3.69) and

$$\mathbf{r}_{\mathscr{D}} \triangleq [r_0, r_1, ..., r_{N_t-1}]^T, \tag{3.75}$$

respectively. If the discriminative probabilistic model (3.58) introduced in the previous paragraph is exploited, a Bayesian method based on it can be developed as follows. To begin, the joint pdf

$$f\left(t, \mathbf{t}_{\mathscr{D}}, \mathbf{w} | r_q, \mathbf{r}_{\mathscr{D}}, \right) = f\left(\mathbf{t}_{\mathscr{D}}, \mathbf{w} | \mathbf{r}_{\mathscr{D}}, \alpha, \beta\right) \, f\left(t | r_q, \mathbf{w}\right) \tag{3.76}$$

is considered in place of the pdf $f(t|r_q, \mathbf{w})$ (3.58); here, β^{-1} is the variance of the noise affecting the labels,

$$f\left(\mathbf{t}_{\mathscr{D}}, \mathbf{w} | \mathbf{r}_{\mathscr{D}}, \alpha, \beta\right) = f\left(\mathbf{t}_{\mathscr{D}} | \mathbf{r}_{\mathscr{D}}, \mathbf{w}, \beta\right) \, f\left(\mathbf{w} | \alpha\right) \tag{3.77}$$

is the joint probability of the $(M+1)$-dimensional weight vector $\mathbf{w}$ (3.60) and the label vector $\mathbf{t}_{\mathscr{D}}$ (3.69) conditioned on $\mathbf{r}_{\mathscr{D}}$ (3.75), on the *hyperparameter* α and on the parameter β, and $f(\mathbf{w}|\alpha)$ is the prior pdf of $\mathbf{w}$. The Gaussian model

$$\begin{aligned} f(\mathbf{w}|\alpha) &= \mathcal{N}\left(\mathbf{w}; \mathbf{0}, \alpha^{-1}\mathbf{I}_{M+1}\right) \\ &= \left(\frac{\alpha}{2\pi}\right)^{(M+1)/2} \exp\left\{-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}\right\} \end{aligned} \tag{3.78}$$

is employed for the second pdf appearing in the RHS of eq. (3.77) (e.g., see [75, Sect. 1.2.4, p. 30, eq. (1.65)]); here, $\mathbf{I}_N$ is the $N \times N$ unit matrix and α represents the precision of the last pdf. The first pdf appearing in the RHS of eq. (3.77), instead, represents a *likelihood function* expressing how likely the response $\mathbf{t}_{\mathscr{D}}$ are, given $\mathbf{r}_{\mathscr{D}}$, $\mathbf{w}$ and β; this function can be factored as

$$f\left(\mathbf{t}_{\mathscr{D}} | \mathbf{r}_{\mathscr{D}}, \mathbf{w}, \beta\right) = \prod_{k=0}^{N_t-1} f\left(t_k | r_k, \mathbf{w}, \beta\right), \tag{3.79}$$

and, consequently, can be expressed in terms of the probabilistic model (3.58) Given the joint pdf $f(t, \mathbf{t}_{\mathscr{D}}, \mathbf{w}, |r_q, \mathbf{r}_{\mathscr{D}})$ (3.76), the *predictive distribution* $f(t|r_q, \mathbf{r}_{\mathscr{D}}, \mathbf{t}_{\mathscr{D}})$ can be evaluated as

$$f\left(t | r_q, \mathbf{r}_{\mathscr{D}}, \mathbf{t}_{\mathscr{D}}\right) = \frac{1}{f\left(\mathbf{t}_{\mathscr{D}} | \mathbf{r}_{\mathscr{D}}, \alpha, \beta\right)} \int f\left(t, \mathbf{t}_{\mathscr{D}}, \mathbf{w} | r_q, \mathbf{r}_{\mathscr{D}}\right) d\mathbf{w}, \tag{3.80}$$

where

$$\begin{aligned} f(\mathbf{t}_{\mathscr{D}} | \mathbf{r}_{\mathscr{D}}, \alpha, \beta) &= \int f\left(\mathbf{t}_{\mathscr{D}}, \mathbf{w} | \mathbf{r}_{\mathscr{D}}, \alpha, \beta\right) \, d\mathbf{w} \\ &= \int f\left(\mathbf{t}_{\mathscr{D}} | \mathbf{r}_{\mathscr{D}}, \mathbf{w}, \beta\right) f\left(\mathbf{w} | \alpha\right) d\mathbf{w} \end{aligned} \tag{3.81}$$

is a *marginal likelihood*. The expression (3.80) can be also reformulated as follows. Substituting the RHS of eq. (3.77) in that of eq. (3.76) and the resulting factorization in the RHS of eq. (3.80) yields

$$f\left(t|r_q,\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}}\right) = \int \frac{f\left(\mathbf{t}_{\mathscr{D}}|\mathbf{r}_{\mathscr{D}},\mathbf{w},\beta\right)\, f\left(\mathbf{w}|\alpha\right)}{f\left(\mathbf{t}_{\mathscr{D}}|\mathbf{r}_{\mathscr{D}},\alpha,\beta\right)} \cdot f\left(t|r_q,\mathbf{w}\right) d\mathbf{w}. \tag{3.82}$$

Then, since

$$\frac{f\left(\mathbf{t}_{\mathscr{D}}|\mathbf{r}_{\mathscr{D}},\mathbf{w},\beta\right) f\left(\mathbf{w}|\alpha\right)}{f\left(\mathbf{t}_{\mathscr{D}}|\mathbf{r}_{\mathscr{D}},\alpha,\beta\right)} = f\left(\mathbf{w}|\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}},\alpha,\beta\right), \tag{3.83}$$

eq. (3.82) can be rewritten as

$$f\left(t|r_q,\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}}\right) = \int f\left(\mathbf{w}|\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}},\alpha,\beta\right)\, f\left(t|r_q,\mathbf{w}\right)\, d\mathbf{w}. \tag{3.84}$$

The last equation shows how the predictive distribution is influenced by our uncertainty about the weight vector; such an uncertainty is expressed by the pdf $f(\mathbf{w}|\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}},\alpha,\beta)$.

Let us apply now the mathematical results derived above to the considered regression problem. If the pdf $f(\mathbf{w}|\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}},\alpha,\beta)$ is assumed to be Gaussian and, in particular,

$$f\left(\mathbf{w}|\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}},\alpha,\beta\right) = \mathcal{N}(\mathbf{w}|\mu_{\mathscr{D}},\sigma^2_{\mathscr{D}}), \tag{3.85}$$

where (e.g., see [75, Sec. 3.3, p. 153, eqs. (3.53)-(3.54)])

$$\mu_{\mathscr{D}} = \beta\, \sigma^2_{\mathscr{D}}\, \mathbf{\Phi}^T_{\mathscr{D}} \mathbf{t}_{\mathscr{D}}, \tag{3.86}$$

$$\sigma^2_{\mathscr{D}} = \left(\alpha \mathbf{I}_{M+1} + \beta \mathbf{\Phi}^T_{\mathscr{D}} \mathbf{\Phi}_{\mathscr{D}}\right)^{-1} \tag{3.87}$$

and the $N_t \times (M+1)$ matrix $\mathbf{\Phi}_{\mathscr{D}}$ is given by eq. (3.68), the expression

$$f\left(t|r_q,\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}}\right) = \mathcal{N}(t;\mu(r_q),\sigma^2(r_q)) \tag{3.88}$$

can be derived from eq. (3.84) (e.g., see [75, Sec. 1.2.4, p. 31, eq. (1.69)] for a proof of this result); here,

$$\mu\left(r_q\right) = \beta\, \boldsymbol{\varphi}\left(r_q\right)^T \mathbf{S} \sum_{k=0}^{N_t-1} \boldsymbol{\varphi}\left(r_k\right)\, t_k, \tag{3.89}$$

$$\sigma^2\left(r_q\right) = \beta^{-1} + \boldsymbol{\varphi}\left(r_q\right)^T \mathbf{S}\, \boldsymbol{\varphi}\left(r_q\right), \tag{3.90}$$

$\boldsymbol{\varphi}(r_q)$ is the $(M+1)$-dimensional vector (3.61) and

$$\mathbf{S}^{-1} \triangleq \alpha\, \mathbf{I}_{M+1} + \beta \sum_{k=0}^{N_t-1} \boldsymbol{\varphi}\left(r_k\right)\ \ \boldsymbol{\varphi}\left(r_k\right)^T \tag{3.91}$$

is an $(M+1) \times (M+1)$ matrix. It is important to point out that the variance $\sigma^2(r_q)$ (3.90) of the predictive distribution $f\left(t|r_q,\mathbf{r}_{\mathscr{D}},\mathbf{t}_{\mathscr{D}}\right)$ (3.88) (and, consequently, the accuracy of the prediction), unlike that of the Gaussian model $f(t|r_q,\mathbf{w})$ (3.58), is given by the sum of two terms; the first term originates from the noise affecting the

labels, whereas the second one from our uncertainty about the parameter vector $\mathbf{w}$. Moreover, the second term is influenced by the considered observation (i.e., it depends on r_q); in practice, smaller values of the standard deviation $\sigma(r_q)$ are usually obtained when r_q is close to the observations of the training set.

The accuracy of the new regression algorithm described above has been assessed on the test set shown in Fig. 3.3 after training it on the set illustrated in Fig. 3.2; moreover, $\alpha = 0.05$ has been selected in this case. The prediction $\mu(r_q)$ evaluated on the basis of eq. (3.89) for each observation of the test set and the corresponding standard deviation $\sigma(r_q)$ (i.e., the square roof of the RHS of eq. (3.90)) are represented in Fig. 3.4. The RMSE evaluated over the test set is equal to 1.4° and is approximately equal to the one computed over the training set. Note that this value is comparable to the ones computed for the predictor described in the previous paragraph (and based on a frequentist approach). For this reason, in this case, the Bayesian approach does not offer any advantage with respect to the frequentist one.

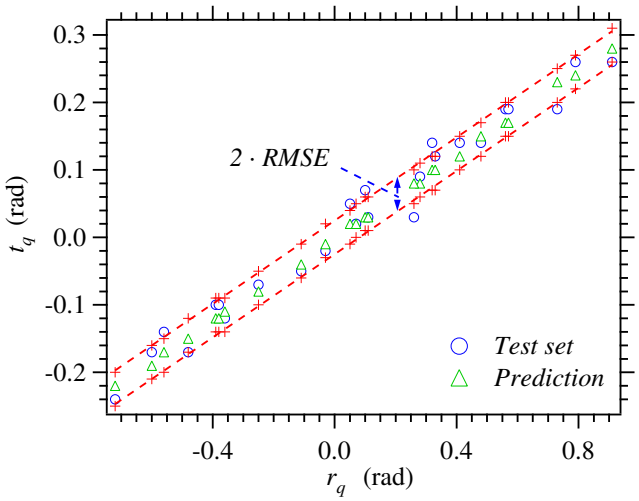


Figure 3.4: *Representation of the regression technique based on the probabilistic model (3.88). The blue circles represent the true domain points, whereas the green triangles the corresponding predictions; the red curves are generated by interpolating the points generated on the basis of the two equations* $t_q = \mu(r_q) \pm \sigma(r_q)$, *with* $r_q \in \mathcal{D}_{ts}$.

Specific methods for binary classification

In the remaining part of this section we focus on a specific supervised problem, namely *binary classification*, and develop two classification methods, based on *discriminative deterministic models*, to solve it. Moreover, we show how different classification methods can be combined to improve the overall accuracy. Note that, in general, classification methods based on *discriminative deterministic models* are able to represent the deterministic mapping between domain points and labels through specific functions, called *discriminant functions*. In the field of radar systems, these methods can be exploited for target detection.

The first method we take into consideration in this paragraph is the *support vector*

machine (SVM) technique; in the following, we limit our analysis to its *linear form*, for simplicity, and assume that the label of each observation can take on only the values ± 1 (consequently, $D_t = 1$). The SVM technique processes the training set $\mathscr{D}$ (3.44) to find the *maximum-margin hyperplane*; this divides the subset of observations for which $t_q = 1$ from that for which $t_q = -1$ in a way that the distance between itself and the nearest point from either group is maximized. In the considered case, the above mentioned hyperplane can be defined as the set of points satisfying the equation

$$y(\mathbf{r}_q, \mathbf{w}) = 0 \tag{3.92}$$

for any q, where

$$y(\mathbf{r}_q, \mathbf{w}) \triangleq \mathbf{w}^T \mathbf{r}_q + b \ , \tag{3.93}$$

$\mathbf{r}_q$ is the q-th observation of $\mathscr{D}$ (see the definition (3.45)), $\mathbf{w}$ represents a D_r-dimensional weight vector (expressed by eq. (3.60), with $M = D_r - 1$) and b is a real parameter called *bias*. The adoption of a classification strategy based on the approach illustrated above relies on the implicit assumption that, if the parameters $\mathbf{w}$ and b appearing in eq. (3.93) are properly selected, the dataset $\mathscr{D}$ (3.44) is *linearly separable in the feature space*. In fact, when this occurs, two parallel hyperplanes separating the above mentioned two subsets of observations and having their mutual distances as large as possible can be found. If the observations of the set $\mathscr{D}$ are normalised, the hyperplanes delimiting the subsets of observations associated with $t_q = 1$ and $t_q = -1$ can be represented by the equations

$$y(\mathbf{r}_q, \mathbf{w}) = 1 \tag{3.94}$$

and

$$y(\mathbf{r}_q, \mathbf{w}) = -1, \tag{3.95}$$

respectively, i.e. briefly as

$$t_q \, y(\mathbf{r}_q, \mathbf{w}) = 1. \tag{3.96}$$

The last formula expresses the *canonical representation* of the decision hyperplanes. Based on the last result, the constraint according to which each point of the set $\mathscr{D}$ (3.44) must lie on the correct side of each of the two hyperplanes (i.e., that it must fall in the correct decision region) can be formulated as

$$t_q \, y(\mathbf{r}_q, \mathbf{w}) \geq 1 \tag{3.97}$$

for any q.

A method for the optimization of the parameters b and $\mathbf{w}$ appearing in eq. (3.93) can be developed as follows. The perpendicular distance of the point $\mathbf{r}_q$ from the decision hyperplane can expressed as

$$\frac{t_q \, y(\mathbf{r}_q)}{||\mathbf{w}||} = \frac{t_q \left(\mathbf{w}^T \mathbf{r}_q + b\right)}{||\mathbf{w}||} \tag{3.98}$$

for any q; its minimum value over the set $\mathscr{D}$ is known as *margin*. The optimal choice $(\hat{\mathbf{w}}, \hat{b})$ of the parameters $(\mathbf{w}, b)$ is the one *maximizing the margin* and, consequently, can be evaluated as

$$\left(\hat{\mathbf{w}}, \hat{b}\right) = \arg \max_{\tilde{\mathbf{w}}, \tilde{b}} \left\{ \frac{1}{||\tilde{\mathbf{w}}||} \min_q \left[t_q (\tilde{\mathbf{w}}^T \mathbf{r}_q + \tilde{b}) \right] \right\} \ , \tag{3.99}$$

where $(\tilde{\mathbf{w}},\, \tilde{b})$ denotes a trial value of the couple $(\mathbf{w},\, b)$; the data points closest to the max-margin hyperplane are called *support vectors*. Unluckily, the optimization problem appearing in the RHS of eq. (3.99) does not admit a simple solution. However, since there is always at least one support vector satisfying eq. (3.96), this problem can be reformulated in a simpler form, i.e. as the maximization of $||\tilde{\mathbf{w}}||^{-1}$ or, equivalently, as

$$\hat{\mathbf{w}} = \arg\min_{\tilde{\mathbf{w}}} \frac{||\tilde{\mathbf{w}}||^2}{2} \tag{3.100}$$

under the constraint expressed by eq. (3.97); note that the parameter $\tilde{b}$ is no more visible in the last formulation, but its value is implicitly determined by the above mentioned constraint. To solve the constrained optimization problem (3.100), the *Lagrangian* function

$$\mathscr{L}\left(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\mathbf{a}}\right) \triangleq \frac{||\tilde{\mathbf{w}}||^2}{2} - \sum_{q=0}^{N_t-1} \tilde{a}_q \{t_q(\tilde{\mathbf{w}}^T \mathbf{r}_q + \tilde{b}) - 1\}. \tag{3.101}$$

is defined; this function depends not only on the parameters $\tilde{\mathbf{w}}$ and $\tilde{b}$, but also on the non negative parameters $\{\tilde{a}_q\}$, called *Lagrange multipliers* and collected in the vector $\tilde{\mathbf{a}} \triangleq [\tilde{a}_0,\, \tilde{a}_1,\, ...,\, \tilde{a}_{N_t-1}]^T$ (the q-th element of this vector is associated with the q-th constraint expressed by eq. (3.97)). Taking the partial derivatives of the function $\mathscr{L}(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\mathbf{a}})$ (3.101) with respect to $\tilde{\mathbf{w}}$ and $\tilde{b}$ and setting them to zero results in

$$\sum_{q=0}^{N_t-1} \tilde{a}_q\, t_q = 0 \tag{3.102}$$

and

$$\tilde{\mathbf{w}} = \sum_{q=0}^{N_t-1} \tilde{a}_q\, t_q\, \mathbf{r}_q, \tag{3.103}$$

respectively. Then, substituting eqs. (3.102)-(3.103) in the RHS of eq. (3.101) produces the so called *dual representation* of the margin maximization problem. Solving the last problem requires maximizing the function

$$\mathscr{L}\left(\tilde{\mathbf{a}}\right) \triangleq \sum_{q=0}^{N_t-1} \tilde{a}_q - \frac{1}{2} \sum_{q=0}^{N_t-1} \sum_{k=0}^{N_t-1} \tilde{a}_q\, \tilde{a}_k\, t_q\, t_k (\mathbf{r}_q^T\, \mathbf{r}_k) \tag{3.104}$$

with respect to the vector $\tilde{\mathbf{a}}$, under the set of constraints $\{\tilde{a}_q \geq 0 \text{ for any } q\}$ and the constraint expressed by eq. (3.102) and produces the optimal value $\hat{\mathbf{a}}$ of the vector $\tilde{\mathbf{a}}$. Given $\hat{\mathbf{a}}$, the optimal values $\hat{\mathbf{w}}$ and $\hat{b}$ of $\tilde{\mathbf{w}}$ and $\tilde{b}$, respectively, are computed as (see eq. (3.103))

$$\hat{\mathbf{w}} = \sum_{q=0}^{N_t-1} \hat{a}_q\, t_q\, \mathbf{r}_q, \tag{3.105}$$

and

$$\hat{b} = N_{\mathscr{S}_M}^{-1} \sum_{q \in \mathscr{S}_M} \left(t_q - \sum_{k \in \mathscr{S}_M} \hat{a}_k\, t_k\, \mathbf{r}_q^T\, \mathbf{r}_k \right), \tag{3.106}$$

respectively, where $\mathscr{S}_M$ and $N_{\mathscr{S}_M}$ denote the set of support vectors and its cardinality, respectively. Given $(\hat{\mathbf{w}},\ \hat{\mathbf{a}},\ \hat{b})$, the classification of a new data point $(\mathbf{r}_q, t_q)$ (with $q > N_t - 1$) is accomplished on the basis of the sign of the quantity (see eq. (3.93))

$$y(\mathbf{r}_q, \hat{\mathbf{w}}) \triangleq \hat{\mathbf{w}}^T\, \mathbf{r}_q + \hat{b}, \tag{3.107}$$

that can be also expressed as (see eq. (3.105))

$$y(\mathbf{r}_q) = \sum_{k=0}^{N_t-1} \hat{a}_k\, t_k\, \mathbf{r}_q^T\, \mathbf{r}_k + \hat{b}. \tag{3.108}$$

As already mentioned above, this classification method is derived under the assumption that the set of feature vectors $\{\mathbf{r}_q\}$ is linearly separable. When this does not occur, a specific *kernel function*, denoted $\boldsymbol{\phi}(\cdot)$, can be used to map the vector $\mathbf{r}_q$ (3.45) into the new feature vector $\boldsymbol{\phi}(\mathbf{r}_q)$ for any q (e.g., see [75, Chap. 6]). The objective is transforming the available classification space into a one characterized by linear boundaries; in principle, the dimensionality of $\boldsymbol{\phi}(\cdot)$ may be different from D_r. Well known examples of the kernels employed with the SVM method are the polynomial, the Gaussian and the Laplace kernels. It is important to note that kernel selection is very critical, since its choice can significantly influence classification accuracy.

The second method we take into consideration is the so called *K nearest-neighbour* (K-NN) technique [51], that represents an example of non-parametric approach to the classification problem. In the case of binary classification, this method can be summarised as follows. The points of the training set $\mathscr{D}$ (3.44) are partitioned into two classes, denoted $\mathscr{C}_0$ and $\mathscr{C}_1$, where

$$\mathscr{C}_k \triangleq \{(\mathbf{r}_{q_k},\, t_{q_k}); q_k = 0, 1, ..., N_k - 1\}, \tag{3.109}$$

with $k = 0$ and 1, and N_k denotes the number of points belonging to the k-th class, so that

$$\sum_{k=0}^{1} N_k = N_t. \tag{3.110}$$

Let us assume now that a new D_r-dimensional observation, denoted $\mathbf{r}_q$ (with $q > N_t - 1$) and called *query instance*, becomes available. The K-NN strategy classifies $\mathbf{r}_q$, i.e. assigns it to one of the two classes defined above, on the basis of the votes of its K *nearest neighbours* (i.e., of the K points of $\mathscr{D}$ closest to $\mathbf{r}_q$); here, K is an integer parameter, whose value is usually small and odd. The identification of the nearest neighbours unavoidably requires the computation of the distance of $\mathbf{r}_q$ from all the points of the set $\mathscr{D}$; if the Euclidean distance is employed, the distance of $\mathbf{r}_q$ from $\mathbf{r}_t \in \mathscr{D}$ is given by

$$d_q \triangleq ||\mathbf{r}_t - \mathbf{r}_q||, \tag{3.111}$$

with $t = 0,\ 1,\ ...,\ N_t - 1$. Given the set $\{d_q\}$, consisting of N_t distances, the nearest neighbours $\{\mathbf{r}_{\mathrm{nn},j};\ j = 0, 1, .., K - 1\}$ are identified by searching for the K points of $\mathscr{D}$ that satisfy the inequality

$$d_q < V_q \tag{3.112}$$

where V_q is a fixed threshold, such that all the required K points are found. Then, if K_k denotes the number of nearest neighbours belonging to $\mathscr{C}_k$ (i.e., the number of *representatives* of $\mathscr{C}_k$), $\mathbf{r}_q$ is assigned to the class having the largest number of representatives, i.e. to $\mathscr{C}_0$ ($\mathscr{C}_1$) if $K_0 > K_1$ ($K_1 > K_0$).

It is worth pointing out that the parameter K controls the degree of smoothing, i.e. the size of the regions assigned to each class. In fact, a small value of K usually results in many small regions assigned to each class, whereas a large one leads to fewer larger regions [75, Par. 2.5.2]. Moreover, if $K = 1$ is selected, a *nearest-neighbour* classifier is obtained; in this case, if the dataset is quite large, it can be shown that the error rate of a K-NN classifier is never larger than twice the minimum achievable error rate of an optimal classifier[7], i.e. of a classifier having full knowledge of the pdf of the observations [95].

Multiple classification methods can be combined to improve the overall accuracy; this idea leads to the development of the so called *ensemble classifiers* [75, Ch. 14.2]. Specific examples of these classifiers are represented by the so called *bootstrap aggregating* (also known as *bagging* [96]) and *boosting* methods [75]. The first method can be employed when M predictions, denoted $\{y^{(m)}(\mathbf{r}_q);\, m = 0, 1, ..., M-1\}$ and generated by M different classifiers (called *base classifiers*), are available; the output is computed as

$$Y_M \triangleq \frac{1}{M} \sum_{m=0}^{M-1} y^{(m)}(\mathbf{r}_q)\,, \tag{3.113}$$

i.e. as an average of all the above mentioned predictions and the predicted class is identified by the sign of this quantity; this reduces the impact of the error due to each single classifier M times. This method is really effective when the errors originating from distinct classifiers are uncorrelated; unluckily, in some cases, such errors may be significantly correlated. When this occurs, classification accuracy can be improved through boosting and, in particular, through the *adaptive boosting* method, also known as *AdaBoost* [97]. In fact, the AdaBoost technique can achieve good accuracy even if its M base classifiers do not perform well (say, their behaviour is only slightly better than random), i.e. they are *weak learners*. If a *binary classification problem* is considered, the training phase of this method evolves through M classification stages, each involving a distinct base classifier; moreover, this method is initialised assigning the same weight to all the observations, i.e. setting $\tilde{w}_q^{(0)} = 1/N_t$ for any q, where $\tilde{w}_q^{(0)}$ denotes the initial weight assigned to the q-th observation. The m-th stage (with $m = 0$, 1, ..., $M-1$) evolves through the following steps:

1) The m-th base classifier is trained to minimise the weighted error function

$$J^{(m)} \triangleq \sum_{q=0}^{N_t-1} \tilde{w}_q^{(m)} I(y^{(m)}(\mathbf{r}_q)), \tag{3.114}$$

where

$$I\left(y^{(m)}(\mathbf{r}_q)\right) \triangleq \begin{cases} 1 \text{ if } & y^{(m)}(\mathbf{r}_q) \neq t_q \\ 0 & \text{otherwise} \end{cases} \tag{3.115}$$

[7]The optimal classification strategy can be easily formulated on the basis of eq. (3.50) (see Paragraph 3.2.2).

and $\{\tilde{w}_q^{(m)}\}$ is a set of non negative weights such that

$$\sum_{q=0}^{N_t-1} \tilde{w}_q^{(m)} = 1. \tag{3.116}$$

2) The weighted measure of the *error rate*

$$\varepsilon^{(m)} \triangleq \frac{\sum_{q=0}^{N_t-1} \tilde{w}_q^{(m)} I\left(y^{(m)}(\mathbf{r}_q)\right)}{\sum_{q=0}^{N_t-1} \tilde{w}_q^{(m)}} \tag{3.117}$$

and the *weighting coefficient* (e.g., see [75, Par. 14.3, eq. (14.16)])

$$\alpha^{(m)} \triangleq \ln\left(\frac{1-\varepsilon^{(m)}}{\varepsilon^{(m)}}\right) \tag{3.118}$$

are computed.

3) The weight assigned to the q-th data point is updated on the basis of the recursive formula (e.g., see [75, Par. 14.3, eq. (14.18)])

$$\tilde{w}_q^{(m+1)} = \tilde{w}_q^{(m)} \exp\left(\alpha^{(m)} I\left(y^{(m)}(\mathbf{r}_q)\right)\right) \tag{3.119}$$

for any q.

These steps force the classifier employed in each stage to put more emphasis on those points that have been misclassified by previous classifiers. In fact, an higher error rate entails a larger increase of the weight assigned to the q-th observation (see eqs. (3.118) and (3.119)), provided that it has not been correctly classified (i.e., that $I(y^{(m)}(\mathbf{r}_q)) = 1$). The final prediction generated by the AdaBoost technique is

$$Y_{B_M}(\mathbf{r}_q) = \text{sign}\left(\sum_{m=0}^{M-1} \alpha^{(m)}\, y^{(m)}(\mathbf{r}_q)\right). \tag{3.120}$$

In assessing the accuracy of any classification method, N-fold cross validation can be used when the size of the available dataset is not so large. This consists in:

a) randomly partitioning the whole available dataset in N blocks;

b) assessing the classification accuracy on the n-th block (taken as test set) after that the considered method has been trained on the basis of the remaining $(N-1)$ blocks (with $n = 0, 1, ..., N-1$).

At the end of this procedure, N distinct accuracies are available; the final score is expressed by their average.

Let us focus now on a specific application of the SVM and K-NN techniques to an FCMW radar system equipped with the antenna array shown in Fig. 3.1-b) (and characterized by $d = \lambda/4$) and operating in the presence of *at most a single point target*. In the q-th trial, the set $\{\mathbf{r}_{0,q}, \mathbf{r}_{1,q}, \mathbf{r}_{2,q}, \mathbf{r}_{3,q}\}$, consisting of four N-dimensional noisy vectors, each associated with one of the $N_V = 4$ virtual receive channels, is available for any q (see eq. (3.18)). The dimensionality reduction technique illustrated

in Paragraph 3.2.1 is applied to this set in order to extract the 4-*dimensional* (4D) feature vector

$$\begin{aligned}\mathbf{R}_q &= \left[R_0^{(q)}, R_1^{(q)}, R_2^{(q)}, R_3^{(q)}\right]^T \\ &\triangleq \left[\left|X_0^{(q)}\right|, \left|X_1^{(q)}\right|, \left|X_2^{(q)}\right|, \left|X_3^{(q)}\right|\right]^T; \end{aligned} \tag{3.121}$$

here,

$$X_v^{(q)} = M_r \; X_{v,\hat{l}}^{(q)} \tag{3.122}$$

and $X_{v,\hat{l}}^{(q)}$ is computed on the basis of eqs. (3.21), (3.25) and (3.28), i.e. by sampling the spectrum $\bar{X}_v^{(q)}(f)$ (3.16) of the zero-padded sample sequence acquired on the v-th virtual antenna at the target frequency $\hat{f}_m$ (3.27) (with $v = 0$, 1, 2 and 3). The target detection strategy we adopt in our radar system is different from the one illustrated in Paragraph 3.2.1 and based on the computation of the average power

$$P_q \triangleq N_V^{-1} \sum_{v=0}^{N_V-1} |X_v^{(q)}|^2, \tag{3.123}$$

and on its comparison with a threshold (see eqs. (3.24) and (3.26)). This choice is motivated by the fact that, the amplitude $a_v^{(q)}$ of the sinusoid observed on the v-th virtual channel and associated with the detected point target is assumed to depend on the antenna index[8] v; the last assumption allows us to account for: a) the dependence of the target reflectivity on the direction of observation; b) the differences in the amplifications introduced by distinct receive chains of the employed MIMO radar. In fact, in the considered radar system, a target is detected if the inequality

$$\max_{v\in\{0,1,2,3\}} \left|X_v^{(q)}\right|^2 > P_{da}, \tag{3.124}$$

holds, i.e. if $|X_v^{(q)}|^2 > P_{da}$ for at least a single value of v. This strategy outperforms the one based on the average power P_q (3.123) in terms of missed detection probability; however, the price to be potentially paid for this is an increase of the false alarm probability, i.e. of the probability of erroneously detecting the presence of target.

In our experiment, the training set

$$\mathscr{D} \triangleq \{(\mathbf{R}_q,\, t_q); q = 0, 1, ..., N_t - 1\}, \tag{3.125}$$

referring $N_t = 100$ independent trials, has been synthetically generated. Half of its data are associated with the detection of a *real* target, the remaining half with the detection of a *false* target; for this reason, the vector $\mathbf{R}_q$ (3.121) is labelled by $t_q = 1$ (-1) in the presence of a *real* (*false*) target. Moreover, the following assumptions have been made in generating the q-th observation of the training set $\mathscr{D}$ and the test set $\mathscr{D}_{ts}$:

a) The amplitude $a_v^{(q)}$ of the sinusoid observed on the v-th antenna in the presence of a real target is uniformly distributed over the interval $[0, 1]$ V;

[8]For this reason, the assumption we made in writing eq. (1.9) does not hold any more.

b) The random variable $a_v^{(q)}$ is independent of $a_u^{(p)}$ for any $u \neq v$ and/or $p \neq q$.

c) The range R_q and the azimuth ϕ_q of the target (if present) are uniformly distributed over the intervals $[R_m\,,\,R_M] = [1.0\text{ m},\,5.0\text{ m}]$ and $[\phi_m\,,\,\phi_M] = [-60^\circ,\,60^\circ]$, respectively, for any q.

The values selected for most of the parameters of the considered radar system are equal to those listed in the examples of Paragraphs 3.2.2 and 3.2.2, the only differences being represented by the fact that:

a) the standard deviation of the noise affecting the received signal samples is $\sigma_w = 1.0$ V (see eq. (1.9));

b) the threshold $P_{da} = 0.3\ \text{V}^2\text{Hz}^{-1}$ is employed by the detection algorithm based on eq. (3.124).

The dataset $\mathcal{D}$ (3.125) has been employed to train the linear SVM, K-NN and Adaboost techniques; $K = 4$ and $M = 100$ has been selected for the second classifier and the third one, respectively. Moreover, the weak learner employed in the m-th step of the Adaboost technique consists in comparing one of the components of the vector $\mathbf{R}_q$ (3.121) with a threshold[9]. More specifically, the classification criterion adopted by each weak learner can be expressed as

$$R_v^{(q)} \underset{t_q=-1}{\overset{t_q=1}{\gtrless}} P_v \tag{3.126}$$

where the index v is randomly selected in the set $\{0, 1, 2, 3\}$ and $P_v \sim \mathcal{U}(\min_q(R_v^{(q)}), \max_q(R_v^{(q)}))$ is the decision threshold associated with the v-th feature $R_v^{(q)}$ acquired in the q-th trial. Note that the classification criterion (3.126) leads to partitioning the observation space into two regions, separated by an hyperplane (perpendicular to one of the reference axes).

In this case, the aim of the three classifiers is discriminating between the presence of a real target and that of a false target any time a target is detected; for this reason, they are exploited to reduce the false alarm probability. Some numerical results are shown in Figs. 3.5, 3.6 and 3.7, that refer to the SVM, to the K-NN and to the Adaboost techniques, respectively; in all these figures, the set of points[10] $\{(R_0^{(q)}, R_1^{(q)})\}$ extracted from the dataset $\mathcal{D}$ (3.125) are represented on a Cartesian plane and are identified by a green (blue) circle if associated with a false (real) target. These results deserve the following comments:

1) SVM training leads to generating the linear decision boundary shown in Fig. 3.5; in this figure, a new observation is classified as a false target, since it falls in the lower decision region.

2) The K-NN method classifies the new observation shown in Fig. 3.6 as a false target, since class $\mathcal{C}_1$ is the one having the largest number of representatives contained in the black circumference (having radius equal to $V_q = 0.07$ V and centered at the new observation).

[9] This simple classifier can be interpreted as a form of a decision tree known as *decision stump* and characterized by a single node (e.g., see [75, Ch. 14.3-14.4]

[10] Note that the observations of the dataset belong to a 4D space in this case; for this reason, all their components cannot be represented in the same figure.

3) Adaboost training leads to generating the decision boundary shown in Fig. 3.7. In the same figure, the *critical points* of the base classifiers (i.e., their misclassified points) are also shown; as it can be easily inferred from eq. (3.119), their weights of these points tend to increase with iterations. In the same figure, a new observation is classified as a false target, since it falls in the lower decision region.

In the considered scenario, our computer simulations have evidenced that the accuracy achieved by the considered classification techniques is around 90%, assuming N-fold cross validation with $N = 5$; in particular, the accuracies of SVM, K-NN and Adaboost are 91%, 89% and 93%, respectively.

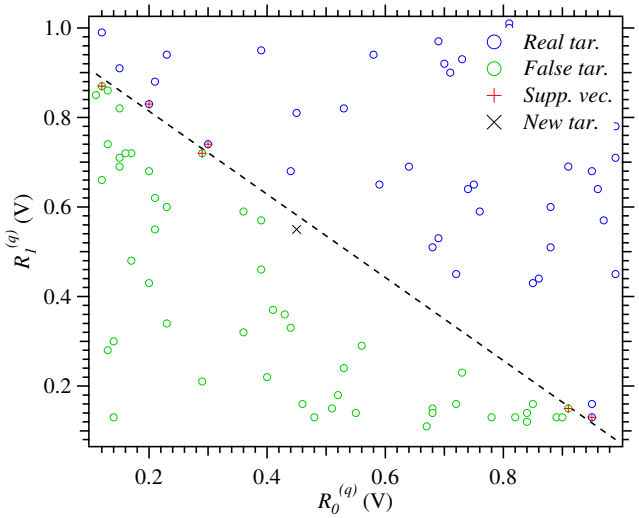


Figure 3.5: *Representation of the decision mechanism employed by a linear SVM classifier. The points of the training set corresponding to false (real) targets are identified by the green (blue) circles. The decision boundary of the SVM is represented by a dashed line, whereas the red crosses identify support vectors. A new observation, identified by a black cross, is classified as a false target, since it falls in the lower half plane delimited by the decision boundary.*

The binary classification methods illustrated above can be also exploited to develop solutions to multi-class problems; in fact, in general, any problem of this type can be represented as a sequence of binary classification problems [98]. This approach is exemplified in Paragraph 3.5.1 and, in more detail, in Paragraph 3.7.1, where its application to the classification of three human activities is illustrated. In particular, in Paragraph 3.7.1, it is shown how a specific solution to this problem can be devised by exploiting *pairwise classification* (also known as *round-robin* class binarization). If K denotes the overall number of classes, this classification method is based on a) combining $L = K(K-1)/2$ binary classifiers (called *base* learners) and b) using the so called *one-versus-one* coding scheme. In this case, each binary classifier is trained assuming one class as positive, another class as negative (the labels associated with the q-th observation are $t_q = 1$ and $t_q = -1$ for the first class and the second one, respectively), and ignoring all the other classes (the label associated with this case is $t_q = 0$). When a new observation is available, it is processed by each binary classifier,

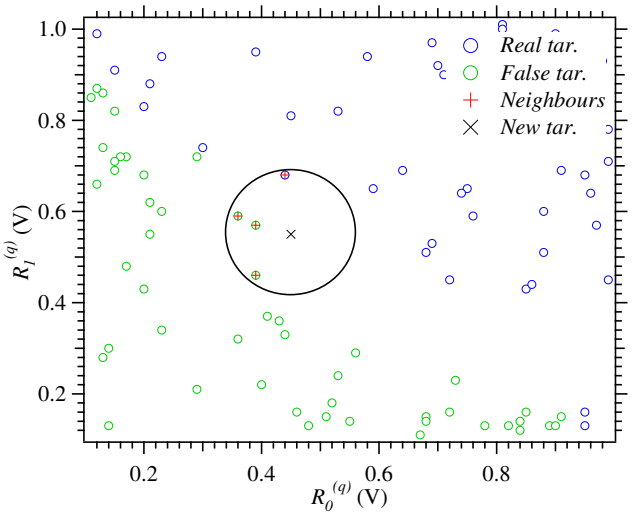


Figure 3.6: *Representation of the decision mechanism employed by a K-NN classifier (with $K = 4$). The points of the training set corresponding to false (real) targets are identified by the green (blue) circles. A new observation, identified by a black cross, is classified as a false target, since class $\mathcal{C}_1$ is the one having the largest number of representatives contained in the black circle.*

so that all its possible assignments to every class pair are taken into consideration. This procedure leads to generating a codeword of size L for each class; the l-th element of this codeword can take on the values ± 1 or 0 on the basis of the class membership established by the l-th binary learner (with $l = 0, 1, ..., L-1$). The K codewords produced by the L learners in response to the q-th observation represent the rows of the $K \times L$ *design matrix* $\mathbf{T}_q = [t_{k,l}^{(q)}]$; note that the presence of a '0' on the k-th row of the l-th column for any q means that all the observations associated with the k-th class are ignored by the l-th classifier. The class predicted for the q-th observation is the one minimizing the average of the binary losses over the L different binary learners [99]; in practice, the value of the class index for the q-th observation is computed as

$$\hat{k}_q = \arg \min_{k \in \{0,1,\ldots,K-1\}} \frac{\sum_{l=0}^{L-1} |t_{k,l}^{(q)}| \, g\left(t_{k,l}^{(q)}, y_{q,l}\right)}{\sum_{l=0}^{L-1} |t_{k,l}^{(q)}|}, \tag{3.127}$$

where $y_{q,l}$ is the *score*[11] assigned by the l-th binary learner to the considered observation and

$$g\left(t_{k,l}^{(q)}, y_{q,l}\right) \triangleq \frac{1}{2} \exp\left(-t_{k,l}^{(q)} y_{q,l}\right) \tag{3.128}$$

is the *binary loss* function.

[11] This quantity can be computed on the basis of eq. (3.108) (eq. (3.120)) if the SVM (Adaboost) method is used.

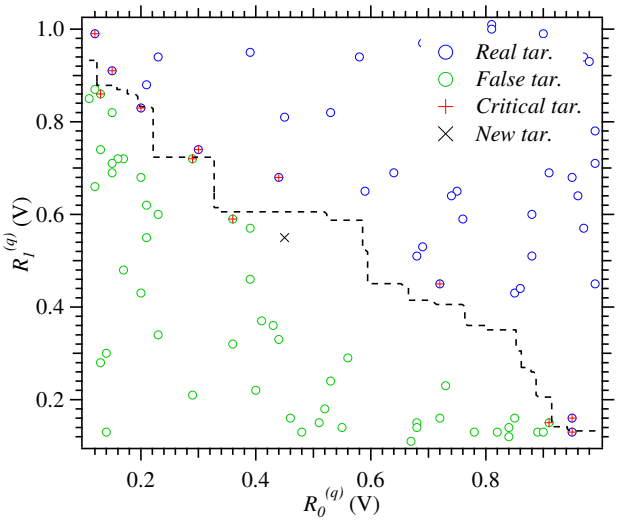


Figure 3.7: *Representation of the decision mechanism employed by the Adaboost classifier. The points of the training set corresponding to false (real) targets are identified by the green (blue) circles. The decision boundary of the Adaboost is represented by a dashed line, whereas the red crosses identify critical targets. A new observation, identified by a black cross, is classified as a false target, since it falls in the lower region delimited by the decision boundary.*

3.2.3 Unsupervised learning

Unsupervised learning is less well defined than its supervised counterpart, since it deals with learning some specific properties of the mechanism on which the generation of the considered set of observations is based. Unlike supervised methods, unsupervised learning operates over *unlabelled* datasets. In the following, we assume that:

1) Learning is based on the dataset

$$\mathcal{D} \triangleq \{\mathbf{r}_q; q = 0, 1, ..., N_t - 1\}, \tag{3.129}$$

that consists of N_t i.i.d. *unlabelled* D_r-dimensional observations;

2) All the available observations are realizations of the same random variable $\mathbf{r}$, characterized by its *unknown* pdf $f(\mathbf{r})$.

The goal of unsupervised methods is to learn some useful properties of the pdf $f(\mathbf{r})$. It is important to keep in mind that the D_r elements which the random vector $\mathbf{r}$ is made of can be highly correlated. These mutual dependencies are often modelled by introducing a new vector, denoted $\mathbf{z}$ and collecting the so called *latent* or *hidden variables*. This approach allows to model the dependencies between the elements of the observations indirectly, i.e. through the direct dependencies between such elements and the hidden vector. The relationship between the vectors $\mathbf{z}$ and $\mathbf{r}$ can be modelled in different ways. This results in various different models that can be adopted in unsupervised learning; further details about this issue can be found in ref. [100].

In the remaining part of this paragraph, we first list the typical unsupervised problems tackled in the field of MIMO radar systems. Then, we describe two specific unsupervised methods and illustrate their application to specific problems in that field.

Unsupervised problems

Unsupervised learning methods can be exploited to solve the following four relevant technical problems:

a) *Clustering* - Data clustering consists in partitioning the dataset $\mathscr{D}$ (3.129) in a number of groups such that data points in the same group are dissimilar from the data points belonging to all the other groups. In clustering problems, an hidden random variable, called *class variable*, is usually added to all the elements of the dataset; this variable describes the cluster membership for every observation of the dataset. In the last years, significant attention has been paid to the use of clustering methods in automotive radar systems, since distinct clusters can be related to different types of targets, like pedestrians, cars or obstacles. A description of two clustering methods employed in the above mentioned field is provided in Paragraph 3.5.4.

b) *Dimensionality reduction* - This aims at generating a reduced dimensionality representation of the observations. Such a representation eases the visualization and interpretation of the dataset, and the identification of specific patterns in it. A well known technique for dimensionality reduction is the *principal component analysis* (PCA); its description is provided in Paragraph 3.5.3, whereas its application to a dataset referring to a specific MIMO radar system is illustrated in Paragraph 3.2.3.

c) *Feature extraction* - This consists in deriving a vector-valued function, denoted $\mathbf{g}(\cdot)$ and such that $\mathbf{g}(\mathbf{r})$ represents a useful and lower-dimensional representation of the feature vector $\mathbf{r}$; the vector $\mathbf{g}(\mathbf{r})$ can be used as an input to a supervised learning method. A well known method for synthetizing the function $\mathbf{g}(\cdot)$ is represented by the *autoencoder*, as illustrated in Paragraph 3.3.4. A simple method for feature extraction in a MIMO radar system has been described in Paragraph 3.2.1; other techniques are illustrated in Paragraphs 3.5.1 and 3.5.4, where their use of radar in human motion characterization and in autonomous driving, respectively, is considered.

d) *Generation of new samples* - This aims at producing new samples of a random vector $\mathbf{r}$ in a way that these are approximately distributed according to its true pdf $f(\mathbf{r})$. Methods for generating new samples can be exploited to de-noise data and for interference mitigation in autonomous driving applications, as illustrated in Paragraph 3.5.4.

Specific unsupervised methods

In this paragraph we focus on two specific unsupervised methods, namely the PCA technique for dimensionality reduction [101] and the *K-means* algorithm for data clustering [102].

The PCA method is employed to project the dataset $\mathscr{D}$ (3.129) onto a new space, called *principal subspace* and having dimensionality $D_r' < D_r$; in doing so, the variance of the projected data is maximised, in order to retain the most relevant variations characterizing the original dataset. This method can be easily understood by illustrating its application to the case in which $D_r = 4$, $D_r' = 1$ and $\mathbf{r}_q = \mathbf{R}_q$, where the 4D vector $\mathbf{R}_q$ is expressed by eq. (3.121). In this case, the 4D observation $\mathbf{R}_q$ is projected onto the scalar

$$R_q^{'} \triangleq \mathbf{u}_0^T \, \mathbf{R}_q, \tag{3.130}$$

where $\mathbf{u}_0$ is a 4D unit vector [75]. If we define the data covariance matrix

$$\mathbf{\Delta} \triangleq \frac{1}{N_t} \sum_{q=0}^{N_t-1} \left(\mathbf{R}_q - \bar{\mathbf{R}}\right) \left(\mathbf{R}_q - \bar{\mathbf{R}}\right)^T, \tag{3.131}$$

where

$$\bar{\mathbf{R}} \triangleq \frac{1}{N_t} \sum_{q=0}^{N_t-1} \mathbf{R}_q, \tag{3.132}$$

is the data mean, the variance

$$\sigma_R^2 = \mathbf{u}_0^T \, \mathbf{\Delta} \, \mathbf{u}_0 \tag{3.133}$$

of the projected dataset is maximized if

$$\mathbf{u}_0^T \mathbf{\Delta} \, \mathbf{u}_0 = \lambda_0, \tag{3.134}$$

where λ_0 is the largest eigenvalue of the matrix $\mathbf{\Delta}$ (3.131) and $\mathbf{u}_0$ (that represents the *first principal component*) is the associated eigenvector.

In general, if a D'_r-dimensional projection space is considered, the principal components are represented by D'_r eigenvectors $\{\mathbf{u}_l;\ l = 0,\ 1,\ ...,\ D'_r - 1\}$ of the data covariance matrix $\mathbf{\Delta}$; these eigenvectors are associated with its D'_r largest eigenvalues $\{\lambda_l; l = 0,\ 1,\ ...,\ D'_r - 1\}$ and are chosen to be *orthonormal*. The quality of the resulting transformation can be assessed by evaluating the *distortion measure* (e.g., see [75, Par. 12.1.2, eq. (12.18)])

$$J \triangleq \sum_{l=D'_r}^{D_r-1} \lambda_l, \tag{3.135}$$

i.e., the sum of the eigenvalues associated with the eigenvectors that are orthogonal to the principal subspace; the smaller is the value taken on by this parameter, the better is the original dataset approximation.

In our experiment, the PCA method has been applied to extract a 2D dataset from the 4D dataset which Figs. 3.5-3.7 refer to (see Paragraph 3.2.2). The 2D points of the new dataset, denoted $\mathscr{D}'$, are represented in the *principal component biplot*[12] shown in Fig. 3.8. In this figure, the axes of the Cartesian plane are associated with the principal components, whereas the vector $\mathbf{w}_i$, represented by an oriented segment, allows to quantify, through its amplitude and orientation, the weight of the contribution provided by the i-th component of the original feature vectors (i.e., of the set $\{\mathbf{R}_q\}$; see eq. (3.121)) to the principal components (with $i = 0,\ 1,\ 2$ and 3). From Fig. 3.8 it is easily inferred that:

1. The weights of the contributions due to $R_2^{(q)}$ and $R_3^{(q)}$ are similar and are about half of those provided by $R_0^{(q)}$ and $R_1^{(q)}$.

2. The new 2D observations referring to real (false) targets are spread over the right (left) half plane of the Cartesian plane.

[12] A detailed description of how a bi-plot is generated can be found in ref. [101, Sect. 5.3]

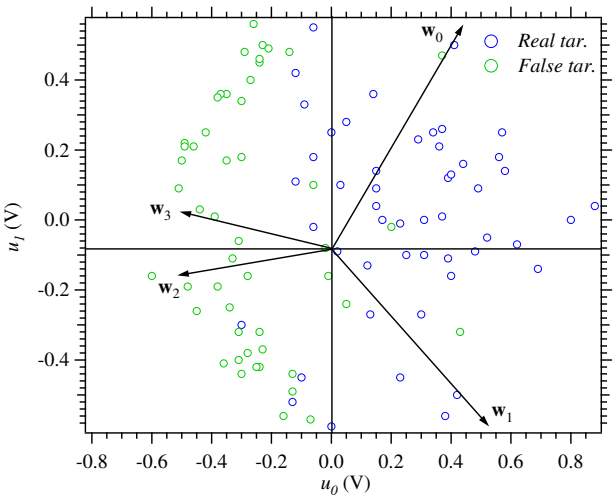


Figure 3.8: *Biplot of the dataset $\mathscr{D}'$ generated by the PCA technique. The points of the reduced dataset corresponding to false (real) targets are identified by the green (blue) circles. The four oriented segments allow to quantify the contribution provided by each of the four components of the original feature vector to the two principal components.*

The *K-means* method allows to partition the available dataset $\mathscr{D}$ into K clusters, each collecting the samples whose mutual distances are small with respect to the distances from the points outside the cluster itself. In practice, if the *center* of the k-th cluster is denoted $\boldsymbol{\mu}_k$ (with $k = 0, 1, ..., K-1$), the K-means method assigns the q-th data point $\mathbf{r}_q$ to the cluster whose center is closest to $\mathbf{r}_q$. This strategy can be formalised as the one minimizing the so-called *distortion measure*

$$V \triangleq \sum_{q=0}^{N_t-1} \sum_{k=0}^{K-1} p_{q,k} ||\mathbf{r}_q - \boldsymbol{\mu}_k||^2, \tag{3.136}$$

with respect to the variables $\{p_{q,k}\}$ and the vectors $\{\boldsymbol{\mu}_k\}$; here, $p_{q,k}$ is a binary indicator variable implementing the 1-of-K coding scheme, i.e. such that $p_{q,k} = 1$ ($p_{q,k} = 0$) if $\mathbf{r}_q$ is (is not) assigned to the k-th cluster. The problem of minimizing the function V (3.136) is solved by means of an iterative procedure, whose iterations consist of two steps. In the first step, known as *expectation*, the metric V is minimized with respect to each of the variables $\{p_{q,k}\}$, keeping the centers $\{\boldsymbol{\mu}_k\}$ fixed; on the contrary, in the second step, called *maximization*, the same metric is minimized with respect to the vectors $\{\boldsymbol{\mu}_k\}$, keeping the variables $\{p_{q,k}\}$ fixed. More specifically, in the first step, the values of the variables $\{p_{q,k}\}$ employed are computed as

$$p_{q,k} \triangleq \begin{cases} 1 \text{ if } \quad k = \arg\min_j ||\mathbf{r}_q - \boldsymbol{\mu}_j||^2 \\ 0 \text{ otherwise} \end{cases} \tag{3.137}$$

for any q (in other words, the q-th data point is assigned to the cluster whose center is

closest to it). Then, in the second step, the center of the k-th cluster is evaluated as

$$\boldsymbol{\mu}_k = \frac{\sum_{q=0}^{N_t-1} p_{q,k}\mathbf{r}_q}{\sum_{q=0}^{N_t-1} p_{q,k}}, \tag{3.138}$$

with $k = 0, 1, ..., K-1$. It is important to point out that:

1. In principle, the initial values of the cluster centers can be arbitrarily chosen. In this case, however, the algorithm may require several iterations to reach convergence. A better initialization procedure consists in choosing the initial centers in a random fashion.

2. The sum appearing in the denominator of the RHS of the eq. (3.138) gives the overall number of points assigned to the k-th cluster; consequently, the cluster center evaluated on the basis of the same equation represents the mean of all the data points $\mathbf{r}_q$ assigned to the k-th cluster.

3. Iterations are stopped when is no further change in the assignments of the data points to the K clusters or their overall number has reached a fixed threshold.

Let us analyse now an application of the K-means technique to the dataset $\mathscr{D}$ (3.129), where $N_t = 100$,

$$\mathbf{r}_q \triangleq \left[\hat{R}_q\,,\,\hat{\phi}_q\right]^T, \tag{3.139}$$

and $\hat{R}_q$ and $\hat{\phi}_q$ represent the estimates of the range and of the azimuth, respectively, of the single point target observed in the q-th trial; these estimates are generated by the algorithm illustrated in Paragraph 3.2.1 and employed in a FCMW radar system equipped with the antenna array illustrated in Fig. 3.1-b) ($d = \lambda/4$ is assumed). Moreover, in generating the q-th observation of the dataset $\mathscr{D}$ (3.129), the following assumptions have been made:

a) The amplitude $a_v^{(q)}$ of the sinusoid observed on the v-th virtual antenna is uniformly distributed over the interval $[0.3, 1.0]$ V.

b) The random variable $a_v^{(q)}$ is independent of $a_u^{(p)}$ for any $u \neq v$ and/or $p \neq q$.

c) The overall number of time-domain samples (N) acquired from each of the four RX antennas is equal to 512 and the standard deviation σ_w of the noise affecting them is equal to 1.0 V (see eq. (1.9)).

d) The oversampling factor $M_r = 4$ and the threshold $P_{th} = 0.5$ V^{2}Hz^{-1} are employed by the detection algorithm based on eqs. (3.25)-(3.26).

e) The range R_q of the target detected in the q-th trial is uniformly distributed over the interval $[R_m\,,\,R_M] = [1.0$ m, 9.0 m$]$, whereas its azimuth is randomly selected in the set of relative integers ranging from ϕ_m to ϕ_M, with $\phi_M = -\phi_m = 45°$.

f) The parameters of the employed radar system take on the same values as those selected for the example illustrated for the SVM and K-NN methods in Paragraph 3.2.2.

In this case, the K-means algorithm is employed to group the detected targets in three different clusters (consequently, $K = 3$ is selected) on the basis of their azimuth

only; the points of the first (third) cluster are characterized by $\phi_q < -15°$ ($\phi_q > 15°$), whereas those of the second one by $|\phi_q| \leq 15°$.

The observations collected in the synthetically generated dataset and their partitioning into the clusters generated by the K-means technique are shown in Fig. 3.9, where circles of different colours are used to identify targets assigned to distinct classes. From these results it is easily inferred that:

1. all the points are correctly classified on the basis of their azimuth, even if an unlabelled dataset is used;

2. each of the centroids is located in the middle of the corresponding cluster and its position is influenced by the distribution of the detected targets along the range dimension.

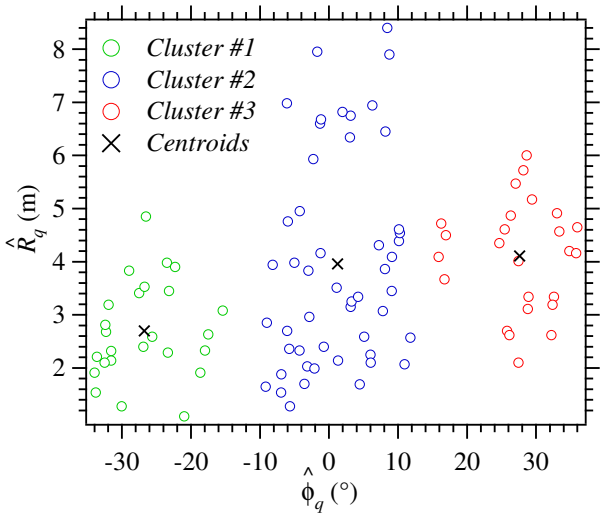


Figure 3.9: *Representation of three clusters generated by the K-means algorithm. The green (red) circles refer to the targets detected on the left (right) of the considered radar system, whereas the blue circles to the targets detected in front of it. The black crosses identify the centroids of the clusters.*

3.3 Deep Learning Techniques for Colocated MIMO Radar Systems

In this section, we first analyse some relevant differences between ML and DL techniques. Then, we introduce readers to deep neural networks by illustrating their architecture, their training and a specific application to an FMCW radar system. Finally, we briefly illustrate a few fundamental DL methods employed in the field of MIMO radar systems.

3.3.1 Relevant differences between ML and DL techniques

Machine learning techniques allow to achieve satisfying accuracy in various applications at the price of a reasonable computational complexity. Nevertheless, in pattern

recognition problems, their capability is often limited by the features selected to learn common patterns and to detect them; in fact, in these cases, devising a transformation able to extract a suitable internal representation from the observed raw data requires good expertise and engineering skills [103]. A revolutionary data-driven approach to feature extraction is offered by DL methods. Despite the significant computational complexity of these methods, in recent times their implementation has become possible thanks to the availability of low-cost powerful *graphic processing units* (GPUs), which make the exploitation of their inner parallelism possible.

Deep learning solves the problem of feature extraction by adopting a multilayer representation of raw data. This fundamental principle is exemplified by a *feedforward deep network*, also known as *multilayer perceptron* (MLP); such a network is able to represent a complicated mathematical function by composing multiple simpler functions, i.e. multiple *layers*. Generally speaking, a MLP consists of three different types of layers: an input layer, multiple hidden layers with learnable weights and an output layer. Its architecture can be represented through a *directed acyclic graph* (DAG), whose structure is exemplified in Fig. 3.10, that refers to the specific case of a fully connected MLP containing a single inner layer. The basic building block of each layer is the so called *neuron*. In general, the output $z_j^{(k)}$ generated by the j-th neuron of the k-th layer can be expressed as

$$z_j^{(k)} = h\left(a_j^{(k)}\right), \tag{3.140}$$

with $j = 1, 2, ..., M_k$ and $k = 1, 2, ..., K$; here, M_k denotes the overall number of neurons in the k-th layer, K denotes the overall number of layers, $h(\cdot)$ is a differentiable non-linear function (i.e., a sigmoid function, an hyperbolic tangent or rectifier linear unit) and the quantity

$$a_j^{(k)} \triangleq \sum_{i=1}^{M_{k-1}} w_{j,i}^{(k)} z_i^{(k-1)} + w_{j,0}^{(k)}, \tag{3.141}$$

known as *activation* function, is a linear combination of the neuron inputs $\{z_i^{(k-1)}; i = 1, 2, ..., M_{k-1}\}$ (whose *learnable weights* are the M_k parameters $\{w_{j,i}^{(k)}; i = 1, 2, ..., M_{k-1}\}$) and the *bias* $w_{j,0}^{(k)}$. The outputs of the neurons of the k-th layer are collected in the vector

$$\mathbf{z}^{(k)} \triangleq \left[z_1^{(k)}, z_2^{(k)}, ..., z_{M_k}^{(k)}\right], \tag{3.142}$$

that feeds the successive hidden layer. The input layer is fed by the D_x-dimensional input vector

$$\mathbf{x} \triangleq [x_1, x_2, ..., x_{D_x}], \tag{3.143}$$

whereas the output layer generates the D_y-dimensional output vector

$$\mathbf{y} \triangleq \left[y_1, y_2, ..., y_{D_y}\right], \tag{3.144}$$

on the basis of eqs. (3.140) and (3.141).

It is important to mention that: a) the learnable weights of the hidden layers can be interpreted as an encoded representation of the inputs; b) unlike ML methods, where a number of manually extracted features are chosen a priori, the considered neural network automatically extracts features through the use of non linear functions.

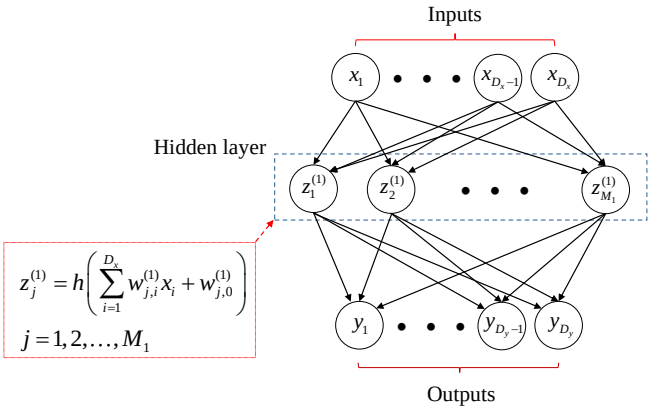


Figure 3.10: *Directed acyclic graph describing the architecture of a fully-connected neural network. Variables are represented by circles (i.e., by nodes), whereas weights by the links between nodes. A single inner layer (i.e., $K = 1$) is assumed for simplicity.*

3.3.2 Training of a deep neural network

Training a deep neural network is an art [100]. Its objective is the same as that already illustrated for ML methods, i.e. the minimization of a *loss* or an *error function* (see Paragraph 3.2.2). However, in a feedforward neural network, this result is achieved by using a local message passing scheme, according to which the internal representations of each neuron are sent, alternately, forward and backward along the graph representing the network itself (e.g., see [75, Par. 5.3]). This scheme, known as *back-propagation* [104], operates as follows (batch processing is assumed here). For each pattern of the training set, the activations of the hidden and output layers of the considered network are computed through successive applications of eqs. (3.141) and (3.140), respectively; this process is known as *forward propagation*, since it proceeds from the input to the output of the network. The back-propagation algorithm, instead, allows to compute the gradient of the selected *error function*, denoted E and corresponding to the *loss function* defined for ML methods (see eq. (3.47)), with respect to the weights appearing in each layer. The derivative of the error function E with respect to the weight $w_{j,i}^{(k)}$, referring to the i-th input of the j-th neuron in the k-th layer of the network, can be expressed as

$$\frac{\partial E}{\partial w_{j,i}^{(k)}} = \sum_q \frac{\partial E_q}{\partial w_{j,i}^{(k)}}, \tag{3.145}$$

with $j = 1, 2, ..., M_k$, $i = 1, 2, ..., M_{k-1}$ and $k = 1, 2, ..., K$; here, E_q represents the error associated with the q-th observation. Based on the chain rule, the partial derivative appearing in the RHS of eq. (3.145) can be evaluated as

$$\frac{\partial E_q}{\partial w_{j,i}^{(k)}} = \sigma_j^{(k)} z_i^{(k)}, \tag{3.146}$$

where $\sigma_j^{(k)} \triangleq \partial E_q / \partial a_j^{(k)}$, $z_i^{(k)} \triangleq \partial a_j^{(k)} / \partial w_{j,i}^{(k)}$ and $a_j^{(k)}$ is defined by eq. (3.141). Consequently, eq. (3.145) can be put in the form

$$\frac{\partial E}{\partial w_{j,i}^{(k)}} = \sum_q \sigma_j^{(k)} \, z_i^{(k)}. \tag{3.147}$$

The quantity $\sigma_j^{(k)}$ appearing in the last formula can be evaluated as follows. First, the quantity

$$\sigma_l^{(K)} \triangleq y_l - t_l \tag{3.148}$$

is computed for the l-th unit of the output layer, where t_l denotes its target. Then, the backpropagation formula

$$\sigma_j^{(k)} = h' \left(a_j^{(k)} \right) \sum_l w_{l,j}^{(k+1)} \, \sigma_l^{(k+1)}. \tag{3.149}$$

is applied for $k = K-1$, $K-2$, ..., 1 and, given k, for $j = 1$, 2, ..., M_k; here, $h'(\cdot)$ denotes the first derivative of the function $h(\cdot)$ appearing in eq. (3.140). This allows us to recursively compute all the quantities $\{\sigma_l^{(k)}\}$ on the basis of the similar quantities $\{\sigma_l^{(k+1)}\}$ made available by all the units appearing in the $(k+1)$-th layer of the network.

It is worth noting that: a) the computational complexity of the network depends on the number of neurons in each hidden layer, since this determines the number of parameters to be tuned in the network; b) overfitting may be observed in the presence of a larger number of neurons. The last problem can be mitigated by including a *regularization term* in the considered error function (a similar strategy has been also proposed for ML methods; see eq. (3.74) in Paragraph 3.2.2). An alternative to this approach is represented by the so called *early stopping* procedure, that consists in stopping network training when the error over a given validation dataset[13] is minimised.

3.3.3 A specific application

Let us focus now on a neural network having the architecture illustrated in Fig. 3.10 and analyse its possible use in an FMCW radar system equipped with the antenna array shown in Fig. 3.1-b) ($d = \lambda/4$ is assumed). In our experiment, the overall synthetically generated dataset includes $\hat{N}_t = 2500$ observations, all acquired in the presence of a single point target, whose range R_q and the azimuth ϕ_q are uniformly distributed over the intervals $[R_m, R_M] = [1 \text{ m}, 7 \text{ m}]$ and $[\phi_m, \phi_M] = [-60°, 60°]$, respectively, for any q. Moreover, the values selected for the parameters of the employed radar system are equal to those listed in the example of Paragraphs 3.2.2 and 3.2.2; the only difference is represented by the standard deviation of the noise affecting the received signal samples, that is $\sigma_w = \sqrt{2}/2$ V. The q-th observation and the associated label are[14]

$$\begin{aligned} \mathbf{r}_q &\triangleq [r_{q,0}, r_{q,1}, r_{q,2}, r_{q,3}, r_{q,4}]^T \\ &= \left[\hat{\psi}_{0,q}, \hat{\psi}_{1,q}, \hat{\psi}_{2,q}, \hat{\psi}_{3,q}, \hat{f}_q \right]^T \end{aligned} \tag{3.150}$$

[13] The *validation dataset* is a set of data on which the performance of the considered network is evaluated during its training.

[14] *Unwrapped* phases are employed in this case, since they ease network training

and

$$\mathbf{t}_q \triangleq [t_{q,0}, t_{q,1}]^T = [R_q, \phi_q]^T , \tag{3.151}$$

respectively; here, $\hat{\psi}_{v,q} = \angle \hat{A}_{v,q}$ (with $v = 0$, 1, 2 and 3) and $\hat{A}_{v,q}$ is the complex amplitude measured on the v-th virtual element at the frequency $\hat{f}_q$ (3.27) (see eqs. (3.28) and (3.35)).

The aim of the neural network is predicting the position of the target (i.e., its azimuth and range) on the basis of a new observation. In this case, the network has 5 inputs two outputs, since $x_j = r_{q,j}$ (with $j = 0$, 1, ..., 4) and $y_k = t_{q,k}$ (with $k = 0$, 1). Moreover, a single hidden layer consisting of $M_1 = 10$ neurons is used; each of these neurons is connected to all the available inputs and employs the hyperbolic tangent transfer function

$$h(x) \triangleq \frac{\exp(2x) - 1}{\exp(2x) + 1} \tag{3.152}$$

in the evaluation of its output on the basis of eqs. (3.140)-(3.141). The predictions of the target range and azimuth are computed by the output layer, that contains two neurons.

The *scaled conjugate gradient* method [105] has been employed to train the network described above. The size of the training set $\mathscr{D}$ is $N_t = 2225$, since 85% of the overall dataset has been exploited for network training; the remaining part $\mathscr{D}_{ts}$ of the dataset, whose size is $\bar{N}_t = 375$, has been used as a test set. Our simulation results have evidenced that the adopted network is able to accurately predict the position of a new target; in fact, the RMSEs evaluated for the range and the azimuth on the set $\mathscr{D}_{ts}$ are approximately equal to 4 cm and to 0.2°, respectively. Finally, it is worth noting that:

a) The use of the network described above does not require a specific expertise.

b) Unlike the regression methods illustrated in Paragraph 3.2.2, the employed network is able to predict both the azimuth and the range of a single point target; however, a by far larger dataset is used for its training.

In general, the main drawback of DL methods is represented by the size of the dataset, which is usually much larger than that needed by ML techniques; this results in a significant increase in the computational effort of the required training.

3.3.4 Specific methods

In this paragraph, we focus on specific deep learning methods, namely autoencoders, convolutional neural networks, convolutional autoencoders, recurrent neural networks and generative adversarial networks. Each method is briefly described and some considerations on its use in the field of MIMO radars are made.

Autoencoders

An *autoencoder* (AE) is a neural network that, similarly as the PCA technique, is able to perform *dimensionality reduction* by learning an efficient representation of its input data in an unsupervised fashion. Since the goal of an AE is to approximate the identity function without learning it exactly, its D_y-dimensional output vector (3.144) can be expressed as

$$\mathbf{y} = \mathbf{h}_w (\mathbf{x}) \approx \mathbf{x}, \tag{3.153}$$

where $\mathbf{h}_w(\cdot)$ represents the transformation performed by the network on its D_x-dimensional input vector $\mathbf{x}$.

The architecture of an *under-complete* AE based on a symmetric encoding-decoding structure is illustrated in Fig. 3.11 [106]. If we consider the encoder side, the number of units contained in each hidden layer[15] decreases as we move from the network input to the output of that side; this is due to the fact that the network tries to learn a compressed version of the input data. On the other hand, the decoder has the goal to reconstruct, as faithfully as possible, the data vector $\mathbf{x}$ available at the AE input, starting from its compressed representation. For this reason, the dimensionality of input layer of the decoder side is lower than that of its output layer. If this network is trained to minimise a reconstruction error, it is able to learn the most important attributes of the input data and how to best reconstruct the original input from an encoded state; ideally, this encoding learns and describes the *latent* attributes of the input data.

Other well known AE architectures are the *denoising* AE and the *sparse* AE. The former AE is largely used for the denoising of images, i.e. to reconstruct a clean image from a corrupted version of it. This task is accomplished by storing only the relevant and recurrent features of an image inside the hidden layers, so that the noise affecting it can be filtered out. The latter one, instead, makes an over-complete representation of its input available at its output.

In the following, we will take into consideration under-complete AEs only, since they are employed in various radar applications, as shown in Paragraphs 3.5.1-3.5.4. It is also worth mentioning that, in such applications, autoencoding is often employed as pre-processing method preceding supervised classification; this allows to learn repetitive structures of input data when the training dataset is not so large. The last application of AEs will be analysed in Paragraph 3.3.4 in more detail.

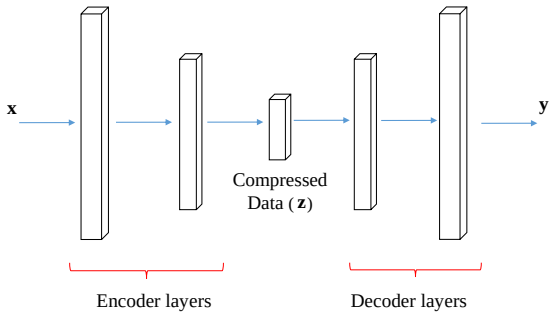


Figure 3.11: *Architecture of an under-complete autoencoder characterized by a symmetric encoding-decoding structure.*

Let us focus now on a possible application of auto-encoding to an FMCW radar system equipped with the antenna array shown in Fig. 3.1-b) ($d = \lambda/4$ is assumed) and operating in the presence of at most a single point target. In this case, the set $\{\mathbf{r}_{0,q}, \mathbf{r}_{1,q}, \mathbf{r}_{2,q}, \mathbf{r}_{3,q}\}$ (see eq.(1.9)), consisting of four N-dimensional noisy vectors becomes available in the q-th trial, with $q = 0, 1, ..., N_t - 1$. Each of these vectors

[15]Note that, in Fig. 3.11 and in the following figures, each layer is represented by a prism having a rectangular base and whose height is proportional to the overall number of its units.

undergoes the DFT processing described in Paragraph 3.2.1; this allows to compute the $4\hat{N}_0$-dimensional feature vector

$$\begin{aligned}\mathbf{R}_q &= [R_{q,0}, R_{q,1}, ..., R_{q,4\hat{N}_0-1}]^T \\ &\triangleq \left[\left(\mathbf{Y}_0^{(q)}\right)^T, \left(\mathbf{Y}_1^{(q)}\right)^T, \left(\mathbf{Y}_2^{(q)}\right)^T, \left(\mathbf{Y}_3^{(q)}\right)^T\right]^T\end{aligned} \tag{3.154}$$

for any q; here, for any v,

$$\begin{aligned}\mathbf{Y}_v^{(q)} &= [Y_{v,0}^{(q)}, Y_{v,1}^{(q)}, ..., Y_{v,\hat{N}_0-1}^{(q)}]^T \\ &\triangleq M_r[|X_{v,b_m}|, |X_{v,b_m+1}|, ..., |X_{v,b_M}|]^T\end{aligned} \tag{3.155}$$

is an $\hat{N}_0$-dimensional vector, $X_{v,k}$ is the k-th element of the N_0-dimensional vector $\mathbf{X}_v^{(q)}$ computed on the basis of eq. (3.20) (with $k = b_m$, $b_m + 1$, ..., b_M), M_r is the oversampling factor employed in DFT processing,

$$\hat{N}_0 \triangleq b_M - b_m + 1, \tag{3.156}$$

and b_m and b_M are integer parameters delimiting the portion of the received signal spectrum over which an amplitude peak, due to the presence of a possible target, is expected. Note that the couple (b_m, b_M) represents a form of a priori information and that, in general, the inequality $0 \leq b_m < b_M \leq N_0 - 1$ holds. Let assume now that the overall data set

$$\mathscr{D}_o = \{(\mathbf{R}_q, t_q); q = 0, 1, ..., \hat{N}_t - 1\}, \tag{3.157}$$

acquired in $\hat{N}_t = 2400$ independent trials, is available; here, the label $t_q = 1$ (-1) refers to the presence of a *real* (*false*) target detected on the basis of the deterministic strategy expressed by eq. (3.124). Moreover, the following assumptions are made in synthetically generating the set $\mathscr{D}_o$ (3.157):

a) Half of its data are associated with the detection of a real target, the remaining half with the detection of a false target.

b) The parameters of the employed radar system take on the same values as those selected for the example illustrated for the SVM and K–NN methods in Paragraph 3.2.2.

c) The stochastic models adopted for amplitude $a_v^{(q)}$ of the sinusoid observed on the v-th antenna in the presence of a real target, and for the range R_q and the azimuth ϕ_q of the target (if present) are the same as those defined in the example illustrated for the SVM and K–NN methods in Paragraph 3.2.2.

d) The size $\hat{N}_0$ of the vector $\mathbf{X}_v^{(q)}$ is equal to 121, since (see eq. (3.156))

$$b_m = \left\lfloor \frac{2R_m\,\mu}{c} N_0 T_s \right\rfloor = 13 \tag{3.158}$$

and

$$b_M = \left\lfloor \frac{2R_M\,\mu}{c} N_0 T_s \right\rfloor = 133, \tag{3.159}$$

where $R_m = 0.5$ m ($R_M = 5.0$ m) represent the minimum (maximum) range expected for the target.

An AE is employed in the considered radar system to reduce the dimensionality of the feature vector $\mathbf{R}_q$ (3.154) (whose size is $4\hat{N}_0 = 484$); note that, unlike the deterministic approach described in Paragraph 3.2.1 and based on a maximum search, an unsupervised data-driven method is exploited in this case. The AE architecture we adopt is similar to the one illustrated in Fig. 3.11, but includes only a single layer in its encoder and a single layer in its decoder, for simplicity. The compressed representation available at the output of the encoder layer is represented by the M-dimensional vector

$$\mathbf{z}_q \triangleq \mathbf{h}_e\left(\mathbf{W}_e \mathbf{R}_q + \mathbf{b}_e\right) \tag{3.160}$$

collecting the hidden variables; here, $\mathbf{W}_e$ is a weight matrix of size $M \times 4\,\hat{N}_0$, $\mathbf{b}_e$ is an M-dimensional bias vector and $\mathbf{h}_e\left(\mathbf{x}\right)$ is an M-dimensional vector resulting from the element-by-element application of the positive saturating linear transfer function

$$h(x) = \begin{cases} 0 \;\text{ if }\; x \leq 0 \\ x \;\text{ if }\; 0 < x < 1 \\ 1 \;\text{ if }\; x \geq 1 \end{cases} \tag{3.161}$$

to the $4\hat{N}_0$-dimensional input vector $\mathbf{R}_q$ (3.154). The decoder maps the encoded representation $\mathbf{z}_q$ (3.160) back to the $4\hat{N}_0$-dimensional vector

$$\begin{aligned} \mathbf{y}_q &= [y_{q,0}, y_{q,1}, ..., y_{q,4\hat{N}_0-1}]^T \\ &\triangleq \mathbf{W}_d\,\mathbf{z}_q + \mathbf{b}_d, \end{aligned} \tag{3.162}$$

that represents an estimate of the original input vector; here, $\mathbf{W}_d$ is a $4\hat{N}_0 \times M$ weight matrix and $\mathbf{b}_d$ is an $4\hat{N}_0$-dimensional bias vector. In our simulations, $M = 60$ has been selected; consequently, a 60-dimensional hidden vector is extracted from a 484-dimensional observation (i.e., roughly an eightfold dimensionality reduction is achieved). Moreover, the *scaled conjugate gradient* method [105] has been employed to train the AE. Training is based on the dataset $\mathscr{D}$, that contains 90% of the dataset $\mathscr{D}_o$ (3.157) and, consequently, involves $N_t = 2160$ observations; the remaining part of the dataset, whose size is $\bar{N}_t = 240$, forms the test set $\mathscr{D}_{ts}$. The effectiveness of the employed AE is exemplified by Fig. 3.12, where the output vector $\mathbf{y}_q$ generated by the autoencoder in response to a specific feature vector $\mathbf{R}_q$ of the test set is shown; this is also confirmed by the small RMSE evaluated over $\mathscr{D}_{ts}$: $RMSE = 0.1$ V is found in this case. This leads to the conclusion that the compressed representation computed by the AE and expressed by the vector $\mathbf{z}_q$ (3.160) is really able to capture all the relevant information conveyed by the input vector $\mathbf{R}_q$ (3.154).

Finally, it worth mentioning that the compressed representation $\mathbf{z}_q$ (3.160) can be exploited to train the linear SVM and K-NN methods described in Paragraph 3.2.2 and employed to discriminate between real and false targets. In our experiment, these

two supervised methods have trained on a dataset consisting of $N_t = 240$ observations ($K = 4$ has been selected for the K-NN method); half of them are associated with the detection of a real target, half with the detection of a false target. Our computer simulations have evidenced that, despite the dimensionality reduction, a slightly better accuracy is achieved by the considered classification techniques; in fact, the obtained accuracies are equal to 93% and 97% for the K-NN and the linear SVM, respectively (N-fold cross validation, with $N = 5$, has been used).

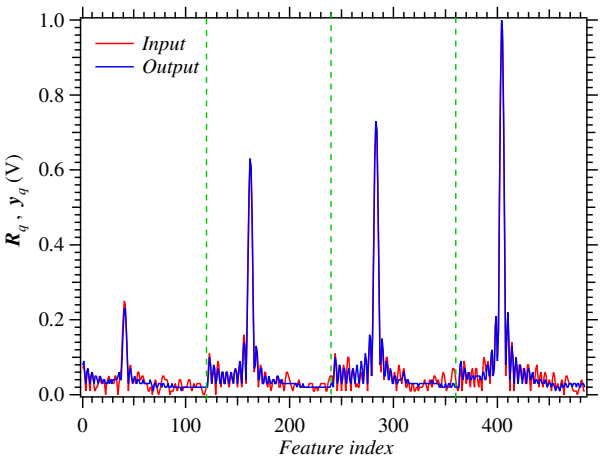


Figure 3.12: *Example of a feature vector* $\mathbf{R}_q$ *(3.154) (red line) and of the corresponding output vector* $\mathbf{y}_q$ *(3.162) (blue line) predicted by the AE employed in the example of Paragraph 3.3.4. The contributions of the four vectors* $\{\mathbf{Y}_v^{(q)};\ v = 0,\ 1,\ 2,\ 3\}$ *which* $\mathbf{R}_q$ *is made of are delimited by green dashed lines.*

Convolutional neural networks

Convolutional neural networks (CNNs) play an important role in DL applications, since they allow to exploit the spatio-temporal information available in a sequence of images [103], [106]; for this reason, they are trained using a labelled dataset. The processing performed by a CNN aims at capturing the local features of input images and is based on *spatially localized convolutional filtering*. Its typical architecture includes *convolutional*, *pooling*, *fully connected* layers, and is motivated by the fact that, in images, local groups of values may exhibit high correlation and local statistics are invariant to position. In fact, convolutional layers aim at detecting local features on the basis of the data originating from the previous layer, pooling layers at merging semantically similar features and fully connected layers at generating the final feature vector.

The processing accomplished by the convolution and pooling operations on a greyscale image is outlined in Fig. 3.13. The output of the convolution depends on both its input, represented by a small portion of the image, and the adopted convolution *kernel*, denoted $\{K[m, n]\}$; moreover, this operation is repeated on multiple disjoint portions until the whole input image is scanned. From a mathematical viewpoint, the convolution input is a matrix, consisting of $I_S \times I_S$ pixels and denoted $\mathbf{I} = [I[i, j]]$, whereas

the resulting output is a $S_Y \times S_Y$ matrix, called *activation* or *feature map* and denoted $\mathbf{Y} = [Y[i,j]]$. The (i,j)-th element (i.e., unit) of the activation map is evaluated as

$$Y[i,j] \triangleq \sigma\left(\sum_{m=-F/2}^{F/2-1}\sum_{n=-F/2}^{F/2-1} K[m,n]\, I[i-m, j-n]\right), \tag{3.163}$$

where F and $K[m,n]$ are the *size* of the convolutional filter (also known as *kernel size*) and its (m,n)-th weight, respectively, and $\sigma(\cdot)$ is a non linear *activation function*. Another relevant parameter of a convolutional layer is its *stride* S, that represents the number of pixels shifts over the input matrix when the kernel moves from a portion of the image to the next one; for instance, when the stride is one, the filter moves one pixel at a time. The area of the input image processed by the kernel can be also extended by adding a set of pixel (usually set at zero) to the border of image itself, as shown in Fig. 3.13; in that figure, the parameter P (dubbed *padding*) represents the number of zero columns and rows added to the input image. The stride, the padding and the kernel size of a convolutional layer influence the size S_Y of the output matrix; in fact, it can be shown that

$$S_Y = \frac{I_S - F + 2\,P}{S} + 1. \tag{3.164}$$

For this reason, the above mentioned parameters have to be jointly selected in a way that the RHS of last equation takes on an integer value.

Generally speaking, the convolution operation expressed by eq. (3.163) can be performed N_d times over the same image; in accomplishing this procedure, the parameters P and S do not change. This produces the output volume (i.e., matrix) $\mathbf{W}$ shown in Fig. 3.13 and having size $S_Y \times S_Y \times N_d$ (the parameter N_d is called *depth*); this matrix results from stacking N_d distinct activation maps, each representing a specific *slice*.

The convolutional layer represented in Fig. 3.13 feeds a *pooling layer*, whose task is reducing the dimensionality of each input slice and, consequently, the overall complexity of the considered CNN. The processing accomplished by the pooling layer can be easily described by referring to a single slice, denoted $\mathbf{Y}$, of the output volume $\mathbf{W}$. Similarly as the convolution operation, the pooling operation is fed by a portion, having size $F_p \times F_p$, of the considered slice and generates the $S_{Y_p} \times S_{Y_p}$ output matrix $\mathbf{Y}_p = [Y_p[i,j]]$. The most popular layers of this type are known as *max pooling* and as *average pooling*. In the former case, the (i,j)-th pixel of the output matrix $\mathbf{Y}_p$ is computed as

$$Y_p[i,j] \triangleq \max_{p,q \in \mathscr{S}_{F_p}(i,j)} Y[p,q], \tag{3.165}$$

whereas in the latter one as

$$Y_p[i,j] \triangleq \frac{1}{F_p^2} \sum_{p,q \in \mathscr{S}_{F_p}(i,j)} Y[p,q], \tag{3.166}$$

where $\mathscr{S}_{F_p}(i,j) \triangleq \{(p,q) | -F_p/2 + i \leq p \leq i + F_p/2 - 1, -F_p/2 + j \leq p \leq F_p/2 - 1 + j)\}$ and the parameter F_p is called *pool size*. It can be shown that

$$S_{Y_p} = \frac{S_Y - F_p}{S_p} + 1, \tag{3.167}$$

where S_p is the *stride* of the pooling (its meaning is similar to that illustrated above for the parameter S). Note that the depth N_d of the final output volume $\mathbf{W}_p$ generated by pooling is the same as that of $\mathbf{W}$.

Finally, it is important to point out that:

a) in CNN applications, a chain of pairs of convolutional and pooling layers is commonly used. Moreover, fully connected layers (FC) of different lengths are often added at the end of the cascade of convolutional/pooling layers, as illustrated in Fig. 3.14; this allows to combine all the extracted features in a 1D vector.

b) As shown in Paragraphs 3.5.1-3.5.4, CNNs are employed in a number of radar applications ranging from human activity characterization to autonomous driving. Some experimental results about the use of CNNs in the classification of three different human activities CNN are illustrated in Section 3.7.

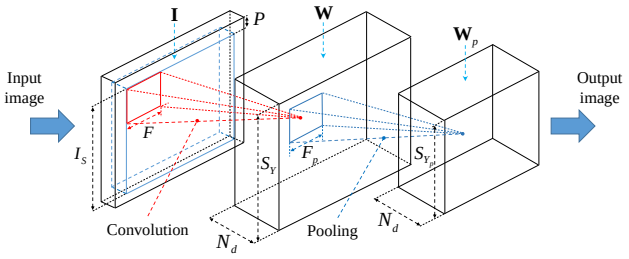


Figure 3.13: *Representation of the* convolution *and* pooling *operations accomplished by a CNN on a greyscale image. The area corresponding to the convolution input (red square) is moved from left to right, and up and down over the input image. The convolution generates the activation map* $\mathbf{Y}$*, that represents a portion of the output volume* $\mathbf{W}$*. Pooling is employed to reduce the size of the final map* $\mathbf{Y}_p$*.*

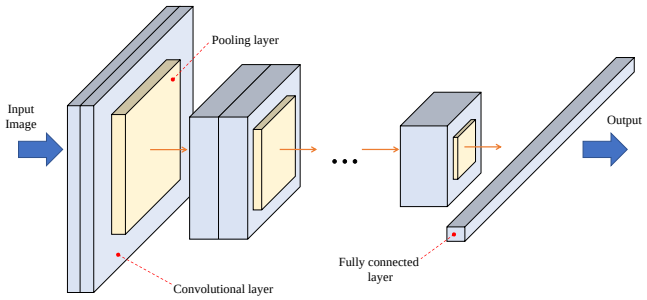


Figure 3.14: *Architecture of a CNN containing multiple convolutional layers, max pooling layers and fully connected layers.*

Convolutional autoencoders

A *convolutional autoencoder* (CAE) may represent an appealing technical option in those applications in which a large amount of labelled data is unavailable. In fact, it combines the advantages offered by unsupervised learning techniques (i.e., by AEs) with the capability of CNNs to extract the spatio-temporal information from a sequence

of images. The architecture of a CAE is exemplified by Fig. 3.15; this network consists of an encoder side, combining convolutional and pooling (i.e., *downsampling*) layers, and of a decoder side, made of transposed convolutional (also called *deconvolutional*) and unpooling (i.e., *upsampling*) layers. Each transposed convolutional layer allows to upsample its input feature map with the aim of retrieving the original shape of the image available at the input of the first convolutional layer contained in the encoder side. In each unpooling layer, instead, an upsampling procedure exploiting the positions of the maxima stored in the corresponding max pooling operation executed at the encoder side is accomplished.

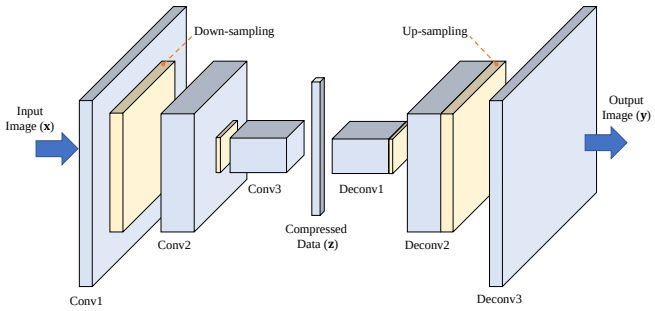


Figure 3.15: *Example of CAE architecture. The acronym* Conv$X$ *(with* $X = 1, 2$ *and* 3*) identifies the* X*-th convolutional and pooling layer, whereas* Deconv$X$ *(with* $X = 1, 2$ *and* 3*) the transpose and unpooling layer. The vector* $\mathbf{z}$ *can be considered as a compressed representation of the input image* $\mathbf{x}$*.*

Recurrent neural networks

In the neural networks treated so far, all the inputs and all the outputs are time-independent from each other. Features related to the time evolution of the observed data can be extracted through a *recurrent neural network* (RNN) [107]. A well known example of RNN is the so called *Vanilla* RNN, whose architecture is represented in Fig 3.16-a). In this network, past information contribute to the computation of its output, since they are reinjected into the network itself and stored in its internal (i.e., hidden) state. Moreover, the following three distinct weight matrices are employed by this network: a) the $M \times D_r$ matrix $\mathbf{U}$ employed in the mapping of the D_r-dimensional input vector $\mathbf{r}_q^{(t)}$ at time t to its M-dimensional hidden state $\mathbf{h}^{(t)}$; b) the $M \times M$ square matrix $\mathbf{W}$ involved in the update of its internal state; c) the $D_r' \times M$ matrix $\mathbf{V}$ employed to map $\mathbf{h}^{(t)}$ to its D_r'-dimensional output vector $\mathbf{o}^{(t)}$. In fact, based on these matrices, the state update of the network and the computation of its output at time t can be expressed as

$$\mathbf{h}^{(t)} = \boldsymbol{\phi}(\mathbf{W}\,\mathbf{h}^{(t-1)} + \mathbf{U}\,\mathbf{r}_q^{(t)}) \tag{3.168}$$

and as

$$\mathbf{o}^{(t)} = \mathbf{V}\,\mathbf{h}^{(t)}, \tag{3.169}$$

respectively; here, $\boldsymbol{\phi}(\cdot)$ denotes a non-linear activation vector function.
It is important to point out that:

a) A RNN can be thought as the result of the interconnection of multiple copies of the same network, each passing a message to a successor. In fact, *unrolling* it leads to a chain-like architecture, made of multiple replicas of the same module and such that each module passes a message to its successor.

b) The standard procedure for training a RNN is known as *backpropagation through time* (BPTT) [108]. Unluckily, it may not be so effective when training involves long time sequences, because of the so called *vanishing* and the *exploding gradient* problems [109]. The former problem refers to the exponential decrease observed in the norm of the gradient of the employed cost function during training, whereas the latter one concerns the opposite behaviour (more specifically, a large increase of the same gradient).

The problems mentioned in the last point can be circumvented by adopting a *long short term memory* (LSTM) neural network [110], whose architecture is illustrated in Fig. 3.16-b). This architecture consists of a *memory cell* and of three different *multiplicative gates*, namely an input gate, an output gate and a forget gate. The input gate, whose content at time t is denoted $\mathbf{i}^{(t)}$, represents the input of the memory cell (whose content at time t is denoted $\mathbf{c}^{(t)}$) and is employed to protect the content of this cell from perturbations due to irrelevant inputs. The output gate, whose content at time t is denoted $\mathbf{o}^{(t)}$, protects the other units connected to the output of the memory cell from perturbations due to irrelevant memory contents. Finally, the forget gate, whose content at time t is denoted $\mathbf{f}^{(t)}$, protects the contents stored in the vector $\mathbf{c}^{(t)}$ from the unwanted fluctuations of the memory at the previous instance (i.e., from $\mathbf{c}^{(t-1)}$). In summary, the cell allows for long term memory storage, whereas the gates prevent memory contents from being perturbed by irrelevant inputs and outputs.

If $\mathbf{r}_q^{(t)}$ denotes the vector of input features at time t, the time evolution of the LSTM network shown in Fig. 3.16-b) is described by the equations

$$\mathbf{i}^{(t)} = \boldsymbol{\sigma}(\mathbf{U}_i\,\mathbf{r}_q^{(t)} + \mathbf{W}_i\,\mathbf{s}^{(t-1)}), \tag{3.170}$$

$$\mathbf{f}^{(t)} = \boldsymbol{\sigma}(\mathbf{U}_f\,\mathbf{r}_q^{(t)} + \mathbf{W}_f\,\mathbf{s}^{(t-1)}), \tag{3.171}$$

$$\mathbf{o}^{(t)} = \boldsymbol{\sigma}(\mathbf{U}_o\,\mathbf{r}_q^{(t)} + \mathbf{W}_o\,\mathbf{s}^{(t-1)}), \tag{3.172}$$

$$\mathbf{g}^{(t)} = \boldsymbol{\sigma}_c(\mathbf{U}_g\,\mathbf{r}_q^{(t)} + \mathbf{W}_g\,\mathbf{s}^{(t-1)}), \tag{3.173}$$

$$\mathbf{c}^{(t)} = \mathbf{c}^{(t-1)} \odot \mathbf{f}^{(t)} + \mathbf{g}^{(t)} \odot \mathbf{i}^{(t)} \tag{3.174}$$

and

$$\mathbf{s}^{(t)} = \boldsymbol{\sigma}_c(\mathbf{c}^{(t)}) \odot \mathbf{o}^{(t)}; \tag{3.175}$$

here, $\boldsymbol{\sigma}(\cdot)$ is a logistic sigmoid vector function, $\boldsymbol{\sigma}_c(\cdot)$ is an hyperbolic tangent vector function, the operator $\odot$ denotes the Hadamard product, $\mathbf{s}^{(t)}$ is the output of the memory cell at time t, $\mathbf{U}_i, \mathbf{U}_f, \mathbf{U}_o$ and $\mathbf{U}_g$ ($\mathbf{W}_i, \mathbf{W}_f, \mathbf{W}_o$ and $\mathbf{W}_g$) are weight matrices characterizing the multiplicative gates and referring to the vector $\mathbf{r}_q^{(t)}$ ($\mathbf{s}^{(t-1)}$), and $\mathbf{g}^{(t)}$ can be interpreted as a candidate state, whose influence on the state $\mathbf{c}^{(t)}$ is controlled by the input gate through $\mathbf{i}^{(t)}$. From eqs. (3.170)-(3.175) it is easily inferred that: a) the contents of the input, output and forget gates at time t are proportional to a combination of both the vectors $\mathbf{r}_q^{(t)}$ and $\mathbf{s}^{(t-1)}$; b) the output state $\mathbf{s}^{(t)}$ depends not only on the cell content $\mathbf{c}^{(t)}$, but also on the content of the output gate (i.e. on $\mathbf{o}^{(t)}$).

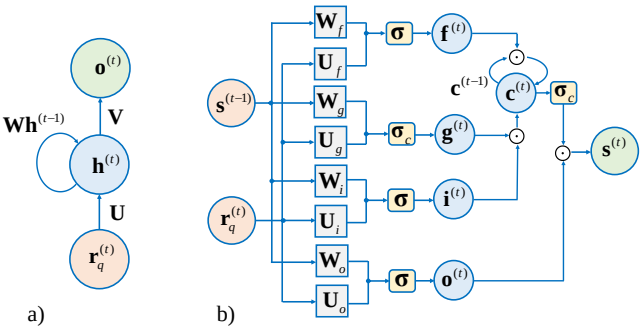


Figure 3.16: *Architecture of: a) a Vanilla RNN; b) an LSTM neural network.*

Let us focus now on the application of an LSTM neural network to an FMCW radar system equipped with a single TX-RX pair and detecting a person that accomplishes specific activities and, in particular, that runs or walks. In this case, each observation processed by the LSTM involves N_f consecutive frames, in each of which N_c chirps are transmitted (see Paragraph 1.3). For this reason, the q-th observation processed by the considered network is extracted from $N_f\,N_c$ noisy vectors, acquired over N_f consecutive frames, (i.e., over $N_f\,N_c$ consecutive chirps). In the p-th frame (with $p = 0,\ 1,\ ...,\ N_f - 1$), the set of vectors $\{\mathbf{r}_{p,0}^{(q)},\ \mathbf{r}_{p,1}^{(q)},\ ...,\ \mathbf{r}_{p,N_c-1}^{(q)}\}$, each having size N, is available; here,

$$\mathbf{r}_{p,k}^{(q)} = \left[r_{p,k,0}^{(q)}, r_{p,k,1}^{(q)}, ..., r_{p,k,N-1}^{(q)}\right], \tag{3.176}$$

represents the vector of signal samples acquired in the k-th chirp interval of the p-th frame and its n-th sample $r_{p,k,n}^{(q)}$ is expressed by a formula similar to eq. (1.23) (with $n = 0,\ 1,\ ...,\ N - 1$). In our experiment, the *Phased Array System* toolbox available in the MATLAB environment is employed to generate the useful signal component (i.e., the contribution of the detected person) to the vector $\mathbf{r}_{p,k}^{(q)}$ (3.176) [111]. This contribution is modelled as the superposition of L different echoes, each originating from a point-like target and associated with a different part of the body. Moreover, in the p-th frame contributing to the q-th observation, the l-th point target is characterized by its RCS $a_{p,l}^{(q)}$, its range $R_{p,l}^{(q)}$ and its radial velocity $v_{p,l}^{(q)}$ (with $p = 0,\ 1,\ ...,\ N_c - 1$ and $l = 0,\ 1,\ ...,\ L - 1$). These parameters are assumed to be static over each frame; in addition, the values they take on in the p-th frame are automatically computed by the above mentioned toolbox on the basis of the height h_p of the person, its position and its RCS in the previous (i.e., in the $(p-1)$-th) frame, the direction of its movement with respect to the radar system and its radial velocity v_q. The dataset processed by the network is

$$\mathcal{D}_o \triangleq \{(\mathbf{R}_q, t_q); q = 0, 1, ..., \hat{N}_t - 1\}, \tag{3.177}$$

where

$$\mathbf{R}_q \triangleq \left[\left(\mathbf{x}_0^{(q)}\right)^T, \left(\mathbf{x}_1^{(q)}\right)^T, ..., \left(\mathbf{x}_{N_f-1}^{(q)}\right)^T\right]^T, \tag{3.178}$$

is the q-th noisy observation, t_q its label,

$$\mathbf{x}_p^{(q)} \triangleq [\hat{R}_p^{(q)}, \hat{v}_p^{(q)}]^T, \tag{3.179}$$

and $\hat{R}_p^{(q)}$ and $\hat{v}_p^{(q)}$ are the estimates of the range $R_p^{(q)}$ and of the velocity $v_p^{(q)}$, respectively, of the considered person in the p-th frame (with $p = 0, 1, ..., N_f - 1$); moreover, it is assumed that $t_q = 1$ (-1) if the person is walking (running), i.e. if $|v_p^{(q)}| \leq v_{th}$ $(|v_p^{(q)}| > v_{th})$, being v_{th} a proper threshold.

In our experiment, the dataset $\mathscr{D}_o$ (3.177) has been acquired in $\hat{N}_t = 400$ independent trials; half of the labels of this dataset are associated with a walker and the remaining half with a runner; moreover, the estimates $\hat{R}_p^{(q)}$ and $\hat{v}_p^{(q)}$ are computed by the algorithm consisting of the following two steps:

1. *Range Estimation* - In this step, the N-dimensional vector $\mathbf{r}_{p,k}^{(q)}$ (3.176) undergoes zero padding; this results in the N_0-dimensional vector $\mathbf{r}_{p,k,\mathrm{ZP}}^{(q)}$, with $N_0 \triangleq M_r N$ (here, the parameter M_r represents the selected oversampling factor). The last vector feeds a N_0-th order FFT, whose output is the N_0-dimensional vector $\mathbf{X}_{p,k}^{(q)} = [X_{p,k,0}^{(q)}, X_{p,k,1}^{(q)}, ..., X_{p,k,N_0-1}^{(q)}]^T$. Then, the average power spectrum

$$P_m^{(q)} \triangleq \frac{1}{N_c} \sum_{k=0}^{N_c-1} |X_{p,k,m}^{(q)}|^2, \tag{3.180}$$

is computed for $k = 0, 1, ..., N_c - 1$. Finally, $\hat{R}_p^{(q)}$ is evaluated as (see eqs. (3.25), (3.27) and (3.29))

$$\hat{R}_p^{(q)} = \frac{c}{2\mu} \hat{f}_p^{(q)}, \tag{3.181}$$

where $\hat{f}_p^{(q)} = \hat{m}_p^{(q)} / N_0 T_s$ and

$$\hat{m}_p^{(q)} = \arg \max_{m \in \{0,1,...,N_0/2\}} P_m^{(q)}. \tag{3.182}$$

2. *Velocity Estimation* - This step is based on the N_c-dimensional vector

$$\hat{\mathbf{A}}_p^{(q)} = [\hat{A}_{p,0}^{(q)}, \hat{A}_{p,1}^{(q)}, ..., \hat{A}_{p,N_c-1}^{(q)}]^T, \tag{3.183}$$

where

$$\hat{A}_{p,k}^{(q)} = M_r \, X_{p,k,\hat{m}_p^{(q)}}^{(q)} \tag{3.184}$$

and $\hat{m}_p^{(q)}$ is expressed by eq. (3.182). Applying zero padding to this vector produces the $N_0^{'}$-dimensional vector $\hat{\mathbf{A}}_{p,\mathrm{ZP}}^{(q)}$, with $N_0^{'} \triangleq M_A N_c$ (here, the parameter M_A represents the selected oversampling factor); the last vector feeds a $N_0^{'}$-th order FFT, whose output is the $N_0^{'}$-dimensional vector

$$\mathbf{d}_p^{(q)} \triangleq \left[d_{p,0}^{(q)}, ..., d_{p,N_0^{'}/2}^{(q)}, d_{p,-N_0^{'}/2+1}^{(q)}, ..., d_{p,-1}^{(q)} \right]^T. \tag{3.185}$$

After solving the maximization problem

$$\hat{k}_p^{(q)} = \arg \max_{\tilde{k} \in \{-N_0'/2+1, -N_0'/2+2, ..., N_0'/2\}} \left| d_{p,\tilde{k}}^{(q)} \right|, \tag{3.186}$$

the estimate (see eqs. (1.24) and (3.27))

$$\hat{v}_p^{(q)} = \frac{1}{2} f_p^{(q)} \lambda \tag{3.187}$$

of the person velocity is evaluated; here,

$$f_p^{(q)} \triangleq \frac{2\hat{k}_p^{(q)}}{N_0' T_0} \tag{3.188}$$

represents the Doppler frequency estimated in p-th frame.

For any q, in generating the sequence of pairs $\{(\hat{R}_p^{(q)}, \hat{v}_p^{(q)}); t = 0, 1, ..., N_f - 1\}$, the following assumptions have been made about the detected person:

a) its response to the signal radiated by the radar system consists of $L = 16$ echoes;

b) its height h_p is uniformly distributed over the interval (1.70 , 2.0) m;

c) its initial coordinates in a 3D space are $(x_0^{(q)}, y_0^{(q)}, z_0^{(q)})$ = (0,10,0) m, whereas the coordinates of the employed radar device in the same reference system are (x_r, y_r, z_r) = (0,0,1) m;

d) the angle ϕ_i representing the initial direction of its velocity is uniformly distributed over the domain $(60° , -60°) \cup (120°, 180°)$ (the reference line, with respect to which this angle is measured, is perpendicular to the array of the radar system).

e) the radial velocity $v_p^{(q)}$ is uniformly distributed over the interval (0.1, 2.1) m/s ((−2.1, −0.1) m/s) if $\phi_i \in (60° , -60°)$ ($\phi_i \in (120°, 180°)$) for any p and, in each frame, changes in an independent fashion;

f) the initial amplitude $a_{0,l}^{(q)}$ is equal to 1V (with $l = 0, 1, ..., L - 1$).

Moreover, the following choices have been made for the employed radar system:

a) the transmitted waveform is characterized by $\lambda = 4$ mm, $\mu = 1.5625 \cdot 10^{13}$ Hz s^{-1}, $T = 64$ μs and $T_0 = 72$ μs;

b) each frame consists of $N_c = 128$ chirps;

c) consecutive frames are separated by a time interval lasting $\Delta t = 40$ ms;

d) the sampling period employed at the receive side is $T_s = 12.5\,\mu$s;

e) the overall number of time-domain samples acquired in each chirp interval is $N = 1024$ and the standard deviation of the noise affecting each sample is $\sigma_w = 0.1$ V (see eq. (1.23));

f) the oversampling factors $M_r = 2$ and $M_A = 8$, and the threshold $v_{th} = 1.1$ m/s are selected for the range/estimation algorithm illustrated above;

g) each observation refers to $N_f = 30$ consecutive frames.

The representation, on a Cartesian plane, of two different feature vectors (see eqs. (3.178)-(3.179)), is provided in Fig. 3.17. These vectors are denoted $\mathbf{R}_0$ and $\mathbf{R}_1$; the former refers to a runner, whereas the latter to a walker. Note that the range difference $\Delta\hat{R}_0 \triangleq |\hat{R}_{N_f-1}^{(0)} - \hat{R}_0^{(0)}|$ referring to the runner is greater than the corresponding quantity (i.e., $\Delta\hat{R}_1 \triangleq |\hat{R}_{N_f-1}^{(1)} - \hat{R}_0^{(1)}|$) referring to the walker. In this case, the proposed LSTM

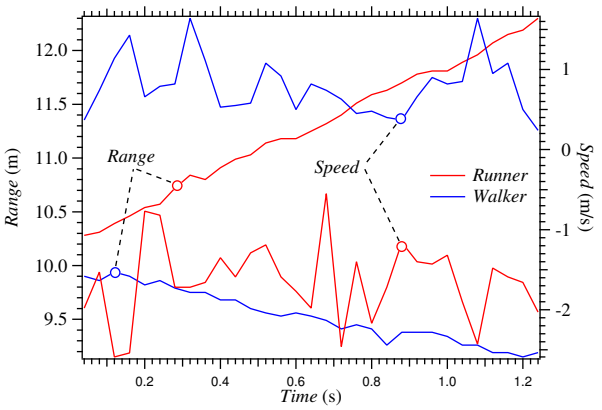


Figure 3.17: *Representation of the elements of the two feature vectors* $\mathbf{R}_0$ *and* $\mathbf{R}_1$*; one refers to a runner (red lines), the other one to a walker (blue lines).*

network is employed to discriminate a walker from a runner. The core of its architecture is characterized by an LSTM layer, able to learn the long term dependencies between different frames. The behaviour of network is described by the block diagram shown in Fig. 3.16 and by eqs. (3.170)-(3.174) (the time index t corresponds to the frame index p in this case). Moreover, in our experiment, the following choices have been made:

1) the size of the input vector is $D_r = 2$, whereas that of the inner state is $M = 10$;

2) the non-linear gate activation function $\sigma(x) = [1 + \exp(-x)]^{-1}$ is used;

3) the size of each of the weight matrices $\{\mathbf{U}_i, \mathbf{U}_f, \mathbf{U}_g, \mathbf{U}_o\}$ is $M \times D_r = 10 \times 2$, whereas that of the weight matrices $\{\mathbf{W}_i, \mathbf{W}_f, \mathbf{W}_g, \mathbf{W}_o\}$ is $M \times M = 10 \times 10$;

5) both the initial cell content $\mathbf{c}^{(0)}$ and the initial state $\mathbf{s}^{(0)}$ are independently chosen as random vectors of size $M = 10$;

6) a fully connected layer and a softmax layer[16] have been added at the output of the LSTM layer to perform classification.

The *adaptive moment estimation* (briefly, *adam*) optimizer [112] has been exploited to train the proposed network (i.e., to tune all the above mentioned weighted matrices); the batch size, the (constant) learning rate and the number of epochs selected for this procedure are $N_S = 32$, $\gamma = 10^{-3}$ and $N_E = 50$, respectively (see eq. (3.73)). Moreover, a training set $\mathcal{D}$ of size $N_t = 300$, corresponding to 75% of the dataset $\mathcal{D}_o$ (3.177) has been employed for training; the remaining part $\mathcal{D}_{ts}$ of $\mathcal{D}_o$ has been used as a test set (collecting $\bar{N}_t = 100$ observations)). Our results have evidenced that a 98% accuracy is achieved by the adopted LSTM network. These results suggest that:

a) Combining deterministic estimators with deep learning methods can result in classification techniques achieving excellent performance;

b) Merging range and velocity information can enhance the discrimination capability of the network;

c) Observing range/velocity evolution over time (i.e., over multiple consecutive frames) significantly contribute to improve network accuracy.

[16] See Par. 3.3.4 for further details about this layer.

Generative adversarial networks

A *generative adversarial network* (GAN) is a probabilistic generative method consisting of two deep neural networks, called *generator* and *discriminator*, and competing one against each other [113]; its architecture is shown in Fig. 3.18. The generator produces a sample[17] $\mathbf{x} = \mathbf{G}(\mathbf{z}, \boldsymbol{\theta}_g)$ from a pdf $f_g(\mathbf{x})$, starting from an input noise variable $\mathbf{z} \sim f_z(\mathbf{z})$; here, $\mathbf{G}(\cdot, \cdot)$ is called *generative model* and is typically implemented through a neural network, whereas $\boldsymbol{\theta}_g$ is the vector of training parameters.

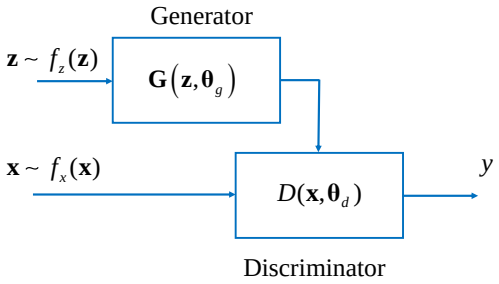


Figure 3.18: *Architecture of a generative adversarial network.*

The discriminator, instead, generates the output $y = D(\mathbf{x}, \boldsymbol{\theta}_d)$, that represents the probability that an input $\mathbf{x}$ originates from the training data (i.e., from their pdf $f_d(\mathbf{x})$) rather than from the generator pdf $f_g(\mathbf{x})$; here, $D(\cdot, \cdot)$ represents the *discriminative model* and $\boldsymbol{\theta}_d$ is the vector of the training parameters characterizing the network that implements the model itself. In practice, the generative model can be thought as a team of counterfeiters, trying to produce fake currency for fooling the police, while the discriminator, acting like the police, tries to detect the counterfeit currency. Competition in this game drives both teams to improve their methods. In fact, the objective of the training of the generative network is minimizing the accuracy of the discriminative network when the data generated by the former network are provided to the latter one; on the contrary, the objective of the discriminator is maximizing the probability of assigning the correct label to both the real data of the training set and the fake samples originating from the generator. For this reason, the interaction between the discriminator and the generator can be modelled as a *two-player minimax game*. This leads to formulating the optimal strategy of these networks as the solution of the minimax problem

$$\min_{\mathbf{G}} \max_{D} V(D, \mathbf{G}) \tag{3.189}$$

$$= \min_{\boldsymbol{\theta}_g} \max_{\boldsymbol{\theta}_d} V(D, \mathbf{G}), \tag{3.190}$$

where

$$\begin{aligned} V(D, \mathbf{G}) \triangleq\ & \mathrm{E}_{\mathbf{x} \sim f_d(\mathbf{x})} \left\{ \log D\left(\mathbf{x}, \boldsymbol{\theta}_d\right) \right\}] + \\ +\ & \mathrm{E}_{\mathbf{z} \sim f_z(\mathbf{z})} \left\{ \log\left(1 - D\left(\mathbf{G}\left(\mathbf{z}, \boldsymbol{\theta}_g\right), \boldsymbol{\theta}_d\right)\right) \right\}. \end{aligned} \tag{3.191}$$

[17] Scalar variables are considered in this paragraph, for simplicity.

The backpropagation algorithm can be used for training a GAN; the training process allows the discriminator of the considered GAN to learn, through a proper feature representation, how to identify real inputs among the generated data and, similarly, the generator how to generate realistic data.

Generative adversarial networks have the favourable property that a wide variety of functions can be incorporated into their model; these make them able to represent very sharp (and even degenerate) data distributions. However, their use require the availability of efficient tools to solve the minimax optimization problem (3.189). Moreover, a tight synchronization between the generator and the discriminator has to be guaranteed during training; in fact, if one of the two networks learns too quickly, the other one may fail to learn.

Softmax Classification Layer

Generally speaking, the DL methods illustrated above can be employed to extract the relevant features of an image. Once this result has been obtained, any multi-class problem referring to that image can be solved by adding a *softmax layer* to the employed network. If K classes are assumed, the target of this layer is generating the posterior probability

$$p_i = \frac{\exp(a_i(\mathbf{r}^{(L)}))}{\sum_{j=0}^{K-1} \exp(a_j(\mathbf{r}^{(L)}))} \tag{3.192}$$

for the i-th class, with $i = 0, 1, ..., K-1$; here,

$$a_j(\mathbf{r}^{(L)}) = \mathbf{w}_j^T \mathbf{r}^{(L)} + w_{j,0} \tag{3.193}$$

and $\mathbf{r}^{(L)}$ is an L-dimensional feature vector made available by the previous hidden (convolutional or LSTM) layer, and $\mathbf{w}_j$ and $w_{j,0}$ are an L-dimensional weight vector and a bias term, respectively, characterizing the softmax layer.

3.4 Comparison of ML and DL techniques

The ML and DL methods described in Secs. 3.2 and 3.3 are compared in Table 3.1 in terms of: a) the type of their training procedure (*supervised*, S, or *unsupervised*, U); b) the complexity of their training procedure (*low*, L, *medium*, M, or *high*, H); c) their classification accuracy; d) their sensitivity to clutter and noise; e) the method they use for extracting the salient features (*manual*, M, or *automatic*, A); f) the size of the dataset they require in order to achieve a good generalization capability (*small*, S, or *large*, L). If these methods are employed for image classification, it should be always kept in mind that:

1. The K-NN and SVM methods require a limited computational load, but achieve low classification accuracy. They are outperformed by CNNs and CAEs at the price, however, of a substantially larger computational effort. Moreover, the last methods are insensitive to a spatial transformation of input data. To understand the importance of the last property (known as *invariance property*), let us take into consideration a CNN employed to classify different objects in a radar image.

This network, thanks to the above mentioned property, is able to select only those portions of the images relevant for its task and its behaviour is not influenced by other irrelevant characteristics, such as the position of a given target or its rotation.

2. *Long short term memory networks* are able to cope with a sequence of signals evolving over time.

3. *Generative adversarial networks* are able to generate synthetic images on the basis of a set of noisy input data. This property can be exploited in radar systems to de-noise images [114] or to detect abnormalities.

3.5 Applications of Machine and Deep Learning Techniques to MIMO Radars

In this section we focus on some applications of the learning methods illustrated in Sections 3.2 and 3.3 to MIMO radar systems. More specifically, we illustrate the exploitation of these methods in the following fields: a) human motion characterization; b) *human gesture recognition* (HGR); c) fall detection and health-care monitoring; d) autonomous driving. Various research results are available in the technical literature about these fields; some essential manuscripts concerning each of them and the use of specific learning methods are listed in Table 3.2.

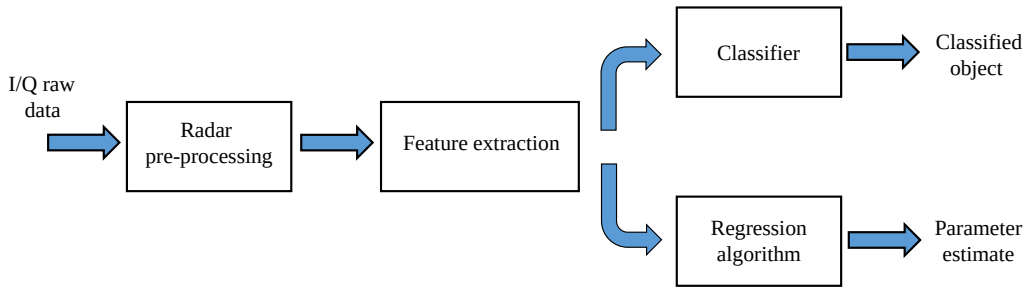


Figure 3.19: *Block diagram representing the signal processing chain of a MIMO radar system that employs a learning method.*

Before delving into the analysis of each application, it is worth pointing out that the processing accomplished at the receive side of any MIMO radar system employing a learning method for classification and/or regression is based on the block diagram shown in Fig. 3.19. First, the received signal undergoes frequency downconversion to generate its in phase and quadrature components. Sampling these components produces a stream of raw data, which is pre-processed (e.g., it may undergo FFT processing; see Paragraphs 1.3 and 3.2.1) before extracting relevant features from it. Finally, these features are processed by a classifier or by a regression algorithm; in the former case, a specific object class is selected, whereas, in the latter one, an estimate of the parameters of interest is evaluated. Feature extraction is based on our prior knowledge about the employed radar system if a ML method is exploited; on the

contrary, features are automatically selected and extracted from pre-processed data if a DL method is adopted. In the following paragraphs, various details about the processing accomplished by the blocks appearing in Fig. 3.19 are provided.

Methods	Learning (S/U)	Complexity (L/M/H)	Performance (L/H)	Clutter Sensitivity (L/H)	Features Extraction (M/A)	Dataset size (S/L)
KNN	S	L	L	L	M	S
SVM	S	L	L	L	M	S
Adaboost	S	M	H	H	M	S
PCA	U	L	L	L	A	S
K-Means	U	L	L	L	A	S
AE	U	H	L	H	A	S
CNN	S	H	H	H	A	L
CAE	U & S	H	H	H	A	L
RNN	S	H	H	H	A	L
GAN	S	H	H	H	A	L

Table 3.1: *Overview of the main characteristics of the learning techniques described in this chapter.*

Learning method Field/Manuscript no.	Clustering	NB	K-NN	SVM	PCA	HMM	AE	CNN	CAE	LSTM	GAN
a) [114–129]		✓	✓	✓	✓		✓	✓	✓		✓
b) [130–134]						✓		✓		✓	
c) [88, 135–140]			✓	✓			✓	✓	✓		
d) [92, 141–154]	✓		✓	✓	✓		✓	✓	✓	✓	

Table 3.2: *Specific learning methods investigated in various manuscripts that concern the four application fields considered in Section 3.5.*

3.5.1 Human motion characterization

Human motion characterization aims at recognizing and classifying different human activities on the basis of the micro-Doppler fluctuations observed in the spectrograms of radar signals. These fluctuations are known as *micro-Doppler signatures* (see Paragraph 1.3). Classifiers employed for this application aim at: a) identifying different types of human motion (e.g., walking, running and sitting) [115]; b) differentiating human

motion from that of other living animals [116]; c) remotely identifying potential active shooters [117, 118].

In the technical literature about this application, the following two methods are exploited to extract relevant features from spectrograms: a) manual extraction of *hand-crafted* features; b) automatic extraction of features based on a *data-driven* approach. Machine learning methods exploiting manual extraction of features have been investigated in [118–126], whereas the automatic extraction of features from micro-Doppler signatures or spectrograms through DL methods has been proposed in [114, 115, 117, 127–129]. It is important to keep in mind that:

1. Machine learning methods relying on spectrogram information usually exploit: a) *physical features* related to the characteristics of the observed motion; b) *transform-based* features; c) *speech-inspired* features. Physical features include the frequency and the bandwidth of the received waveforms, the offset and the signal intensity of the associated signature (see Paragraph 1.3). The first two physical features are strictly related to motion dynamics, whereas the other types of features to the RCS of the body of the observed person [116].

2. Transform-based features exploited by ML methods can be extracted from a received waveform by evaluating a) its spectral coefficients (e.g., its *discrete cosine transform* coefficients) or b) other signal-dependent coefficients. As far as point b) is concerned, the use of *linear predictive coding* (LPC) has been proposed in ref. [120] to transform a time-varying Doppler signal in a low dimensional set of prediction coefficients. A different approach, based on the computation of pseudo-Zernike moments, is illustrated in ref. [121, Sect. II-A, eqs. (10)-(12)]; this allows to extract relevant characteristics from micro-Doppler signatures, such as translational and scale invariance.

3. In many cases, the dimensionality of the feature space can be substantially reduced (see Paragraph 3.2.3). An interesting example of this approach is offered in ref. [122], where the use of a 1D standard PCA and of a robust PCA for extracting physical features from a Doppler radar signal is investigated.

Frequently used ML methods for the classification of human motion based on a set of handcrafted features include the *naive Bayes* (NB) [123], the non linear SVM [124] and the K-NN methods [125]. The use of a binary SVM classifier for multi-class problems in human motion characterization is investigated in ref. [126], where a classification procedure based on a *decision-tree* is proposed. This procedure is based on the idea of representing a classification problem involving multiple choices as a set of binary classification problems, each of which is solved through a binary SVM. This approach is exemplified in [126, Fig. 8], where a decision tree referring to the case of seven classes is illustrated. In practice, a binary SVM is employed for each node of the employed decision tree in order to separate the possible activities in two groups; if each of the two groups is further divided, another SVM classifier is used at an underlying node.

The most relevant problems emerging from the study of ML-based classification of human motion concern the processing methods to be employed for the extraction of hand-crafted features from raw micro-Doppler signals, the sensitivity of these methods

to noise and clutter, and the impact of similarities among the considered classes on their performance. The ability of a *deep neural network* to learn the relevant features directly from the available raw data allows to solve the above mentioned problems. This consideration has motivated the investigation of *deep* CNNs (DCNNs; see Paragraph 3.3.4) for the automatic extraction of features in human motion characterization. The use of a DCNN, fed by spectrograms (converted in *red green blue*, RGB, or greyscale images), and employing convolutional layers and pooling layers of small size, has been proposed in ref. [127]. A different DL method, based on the same principles as convolutional autoencoding (see Paragraph 3.3.4), has been developed in ref. [128, Par. IV-C, Fig.8]. It combines the ability of a DCNN to capture local features of input images with that of an AE to directly learn features through an unsupervised pre-training procedure. In this case, after an initial and unsupervised pre-training stage, the decoder of a CAE is substituted by a few fully connected layers and a softmax classifier. This procedure allows the resulting DCNN to learn specific patterns from the processed signatures, so easing training for supervised classification. The performance results obtained in this case lead to the conclusion that a CAE not only is able to outperform conventional classification methods based on handcrafted features (e.g., SVM), but also a standard DCNN.

Finally, it is useful to mention that another important research problem investigated in the field considered in this paragraph is represented by the *de-noising of micro-Doppler spectra*. In this case, the training set includes two different types of images: a) perfectly clean spectrograms; b) the same spectrograms affected by background noise. The use of a deep GAN, based on a convolutional encoder-decoder structure, has been proposed in ref. [114] for this application. The performance results obtained in this case evidence that this network does not affect the relevant components of micro-Doppler spectra and is able to outperform other classic de-noising techniques commonly used for the suppression of background noise.

3.5.2 Human gesture recognition

The significant attention paid to HGR is due to its exploitation in advanced *human computer interfaces* (HCIs), that are employed in a number of control, infotainment and security applications. Relevant information about the dynamics of human gestures are typically contained in the micro-Doppler signatures acquired over consecutive transmitted frames. Therefore, similarly as human activity characterization, relevant physical features can be easily extracted from spectrograms. A commonly employed ML tool for classifying vectors of handcrafted features in HGR systems is represented by *hidden Markov modelling* [130]. This approach leads to classifying a new sequence of data, called observation, on the basis of a stochastic model, called *hidden Markov model* (HMM), which has been extracted from past observations and describes their generation. If an HMM of a given random phenomenon is available, the probability of observing a specific realization (e.g., a specific gesture), conditioned on a given sequence of hidden states, can be computed. In this case, model training aims at estimating the so called *transition* and *emission* probability matrices of the developed HMM; the former matrix collects the probabilities to move from a given state to another one, while the latter one the probabilities that a given observation is generated in each specific

state. The efficacy of a HMM-based classifier depends on the overall number of states characterizing the model; in general, a larger number of states allows to model a more complicated process and to improve prediction accuracy. However, a discrete state space of small size is often adopted in HGR applications in order to mitigate the overall complexity of the developed HGR system. This choice makes the resulting classifier unable to distinguish gestures characterized only by subtle differences in their spectrograms. For this reason, DCNNs are usually preferred. One of the first important research activities focusing on the exploitation of this type of networks in HGR is the well known Google's *Soli project* [131], whose scope has been the development of a HGR mobile and wearable device based on a RF sensor. Various research results about this research field can be found in ref. [132], where it is shown that the accuracy of these deep classifiers gets worse if: a) the number of classes[18] increases; b) the incident angle and/or the distance between the gesture and the employed radar device get larger. The accuracy of a classifier based on a DCNN can be improved by extracting features not only from spectrograms, but also from range-Doppler maps [133]. Another DL architecture, specifically developed for RF HGR and combining the ability of a CNN network of capturing local features of input images with that of coping with time-varying signals, has been investigated in ref. [134]. This architecture consists of a 3D-CNN for spatial-temporal modelling of short consecutive frames, an LSTM for extracting global temporal features and a final classification layer (a detailed block diagram is illustrated in ref. [134, Fig. 7]). This architecture achieves a very high recognition accuracy, and outperforms other conventional ML and DL methods used in HGR applications, like HMMs or 2D-CNNs.

3.5.3 Fall detection and health-care monitoring

Human falls represent a worldwide health problem and are known to be one of the main causes of unintentional injury death in seniors; this motivates the recent interest in devising electronic systems able to detect their occurrence. Another important problem in the field of technology for human health concerns the development of non-invasive and non-contact devices for monitoring human vital signs, such as breath and heart rates, and sleep quality. Various results in both research areas have evidenced that innovative solutions to both problems can be developed by exploiting ML and DL methods fed by the micro-Doppler signatures acquired through a radar system. In any case, when the overall number of classes to be identified increases and the degree of dissimilarity between the Doppler signatures characterizing them reduces, DL methods are preferred, since they achieve better accuracy.

An interesting study on the dynamics of human falls analysed through micro-Doppler signatures can be found in in ref. [135], where it is shown that fall accidents can be distinguished from normal activities on the basis of: a) the strength of the received echo (i.e., the RCS of the subject under test); b) the distance of the radar device from the body of the subject under test during a fall; c) the Doppler information acquired during the movement of the subject itself. Experimental results have evidenced that, when a subject starts falling, the observed Doppler frequency increases steeply; on the contrary, the RCS of the human subject gradually decreases since its tilt angle gets

[18] The maximum number of distinct hand gestures considered in ref. [132] is equal to 10.

larger. In this case, ML and DL algorithms can be trained to detect a fall on the basis of the time variations of Doppler signatures. A specific DL classifier based on a stacked AE and exploiting a range-Doppler radar has been developed in ref. [136], where it is shown that the proposed solution is more accurate than PCA-based methods in detecting different actions, such as falling, walking, sitting and bending.

The use of learning techniques in the analysis of sleep stages has been investigated in ref. [137], where a solution based on a K-NN classifier has been proposed.

The exploitation of learning techniques for heart and breath rate estimation represents a challenging problem, because a large and heterogeneous datasets for network training cannot be easily built and contactless systems for vital sign monitoring are strongly limited by body movements. Some interesting contributions to this field are provided by refs. [138], [139] and [140]. More specifically, a method based on a classical feed-forward NN for hearth rate estimation is proposed in ref. [138], whereas a DL method for body movement compensation is investigated in ref. [139]. Finally, a contactless breathing disorder recognition system using 2.4-GHz Doppler radar and based on a linear SVM classifier is developed in ref. [140].

The real-time implementation of radar sensing methods for HGR, health monitoring and fall detection can be computationally intensive. This problem becomes more relevant in all those applications in which multiple persons have to be monitored in the same environment; in fact, in such cases, the exploitation of the MIMO technology becomes mandatory, because of the need of localising multiple agents. This explains why an important technical challenge is represented by the exploitation of hardware platforms that support parallel computing (namely, FPGAs and GPUs), require a limited power consumption and can manage a large data rate at their inputs [88].

3.5.4 Autonomous driving

Automotive radar represents one of the key enabling technologies for autonomous driving. The typical processing chain employed for target detection in a MIMO FMCW radar system for automotive applications is represented in Fig. 3.20. The signals ac-

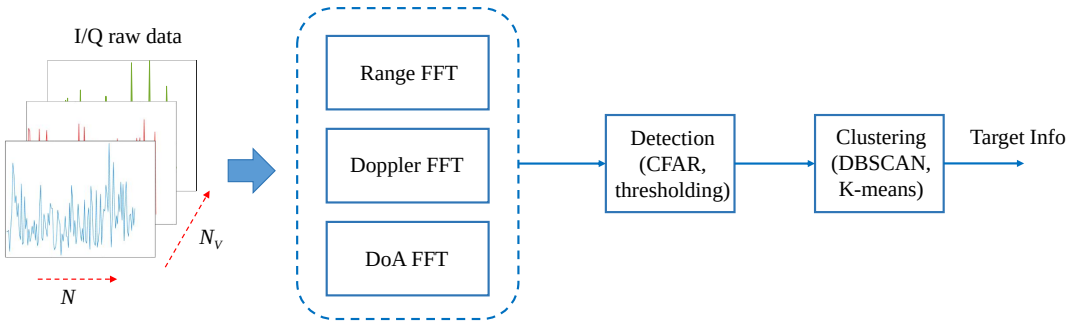


Figure 3.20: *Typical processing chain employed in automotive radar systems [53].*

quired through multiple receive antennas undergo multidimensional FFT processing; this allows to extract range, Doppler and DoA information. The data generated by the FFT blocks are processed by a detection algorithm, whose objective is identifying

the presence of multiple extended targets, and estimating their spatial coordinates and their radial velocity (i.e., the Doppler shift characterizing them). Each of the detected targets (e.g., pedestrians, cars or bicycles) usually appears as a cloud of point targets; the association of each point to a given extended target is called *clustering*. The simplest unsupervised strategy that can be adopted for target detection is *thresholding*; in this case, a target is detected if the amplitude spectrum of the acquired signals exceeds a fixed threshold, as illustrated in Paragraphs 3.2.1 and 3.2.2 for specific FMCW radar systems. A more refined alternative is represented by the *constant false alarm rate* (CFAR) technique [155]. This method consists in estimating the level of interference in each cell in the range domain of interest and in exploiting these information for the detection of the presence of a target in each cell of a radar image. Unluckily, due to the high resolution achieved by automotive radars, a single target can occupy multiple adjacent cells; when this occurs, the CFAR technique undergoes performance degradation because of the contamination affecting the estimated interference level. Clustering techniques rely on the key idea that each cluster of points is a region containing a group of detected targets, whose center typically corresponds to the point target characterized by the strongest reflectivity (see Par. 3.2.3). This means that each cluster has a density (in terms of targets per region) which is considerably larger than that outside it; for this reason, a given point is expected to be part of a cluster if the number of its neighbours is greater than a proper threshold. Learning methods for unsupervised clustering include the *density based clustering algorithm* (DBSCAN) [141], [142], and the K-means algorithm [143]. The main difference between these two methods consists in the fact that the former method, unlike the latter one, does not require prior knowledge of the number of clusters and their shape.

All the techniques described above (namely, thresholding, CFAR and clustering) allow to detect multiple point targets and to cluster them. In general, learning methods can be adopted to improve detection performance. A number of technical problems have been identified in this area; most of them require the development of sophisticated signal processing algorithms. Specific contributions about the use of ML methods in target detection can be found in refs. [144–146]. In particular, a K-NN classifier is proposed as an alternative to robust CFAR detection in ref. [144], whereas the use of the SVM and PCA techniques for improving angular resolution is investigated in ref. [145]. The use of DL methods for target classification in a 2D space, instead, have been studied in refs. [92, 147–154]. It is worth mentioning that, in the technical literature, the first results about the use of DL methods in automotive radar systems appeared after 2015, when it was found that DCNNs were able to simultaneously detect, localize and classify multiple targets by simply analysing 2D range-azimuth (or range-Doppler) maps. Networks originally developed for computer vision applications, like *AlexNet* [156] or *ImageNet* [157], have inspired the architecture of various networks for automatically extracting features from automotive radar images [92, 147, 148]. Despite this, the CNNs usually devised for automotive applications are not as deep as those employed in computer vision. This difference is mainly due to the fact that: a) the information provided by range-azimuth or range-Doppler maps are not as rich as traditional RGB images; b) the employed inference procedure has to be as fast as possible [149]. These ideas are exemplified by the CNN proposed in ref. [150] for the classification of automotive targets, like motorcycles, cars, bicycles and pedestrians; its architecture, illustrated in

ref. [150, Fig. 2], consists of three convolutional layers and filters of size 3×3 (whose depths are equal to 32, 64 and 128, respectively). Moreover, each convolutional layer is followed by a 2×2 average-pooling layer, two fully-connected layers and a softmax layer, which is used at the end for classification. A relevant novelty introduced in this work (and in ref. [151] too) consists in considering a certain *region of interest* (ROI) around the desired targets in the analysed scene as prior information to be used during training, in order to improve the learning procedure.

Deep learning methods can be also employed to solve the problem of *scene understanding*, i.e. of correctly interpreting the events occurring around it (e.g., the event of a vehicle passing near a pedestrian that crosses a road). In this case, improving the prediction accuracy of the employed NN requires exploiting the information contained in the frames preceding and following the frame under test because of the high variability of the data provided by MIMO radar systems. An architecture based on the cascade of a LSTM module with a CNN has been proposed in ref. [152]; this exploits the temporal information provided by radar signals and is able to capture the dynamics of the surrounding scene.

Finally, it is worth mentioning that learning methods can be also employed to detect the fatigue of the driver's eyes [146] and to mitigate the interference orginating from the transmission of multiple MIMO radars in the same area. In general, the interference affecting a MIMO radar system can be due to the system itself (*self-interference*) or from other radar systems placed on the same vehicle or on other vehicles (*cross or mutual interference*); in both cases, this phenomenon results in an increase of the observed noise floor and, consequently, affects the detectability of targets. The use of RNNs for interference mitigation has been investigated in refs. [153] and [154].

3.6 Current Trends in Research on MIMO Radars

In this section, a short description of three research trends in the field of DL techniques for MIMO radars is provided. More specifically, we first focus on *transfer learning*, and recent DL methods for object detection and classification. Then, we discuss the role that *explainable artificial intelligence* (XAI) may play in the radar field.

3.6.1 Transfer learning

The minimization procedure accomplished by a deep NN trained from scratch (through random initialization) may lead to a local minimum which is far from the globally optimal solution if the involved cost function is highly non-convex. Moreover, if the dataset employed in network training is not large enough, the risk of over-fitting is quite high. These problems are likely to arise in radar applications. When this occurs, *transfer learning* could represent a tool to solve them; in fact, this method often allows to achieve a good generalization capability even if the available dataset is limited [158, 159]. Transfer learning is based on the idea of exploiting the knowledge gained from a different domain to solve other related classification problems. Two approaches to the exploitation of this method in radar applications have been recently proposed. The first approach, developed for the classification of human activities, is based on training an *unsupervised network*, characterized by an encoder-decoder structure and employed

to learn specific patterns appearing in the available dataset [160]. When the decoder becomes able to reconstruct the input data with a reasonable accuracy, it is removed, and fully connected and softmax layers are added in cascade to the associated encoder. Finally, the resulting network is trained in a supervised manner with a smaller, but labelled dataset: this procedure is called *fine-tuning*.

The second approach is based on the architecture represented in Fig. 3.21 and developed in ref. [161]. In this case, a DCNN network trained on a large dataset of RGB images is combined with fully connected and softmax layers initialized from scratch; this results in a new network, which is fine-tuned on a small dataset.

The decision about which type of transfer learning has to be preferred is based on the size of the available dataset and on the similarity of the last dataset with the one used for pre-training the selected network architecture. It has been shown that the final score of a DCNN-based classifier can be improved either by exploiting a pre-training procedure based on a simulated radar dataset [162] or by employing a pre-trained DCNN on a separate large scale RGB dataset [161].

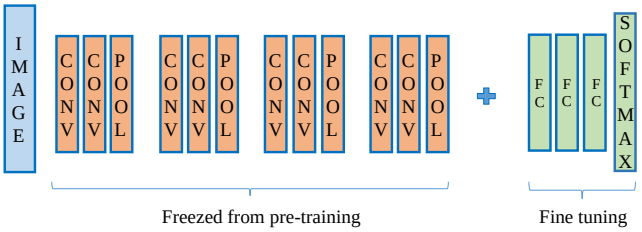


Figure 3.21: *Architecture of a DCNN pre-trained on dataset of RGB images and fine-tuned on a small dataset of radar images.*

3.6.2 Object detection and classification

The aims of object detection and classification are the labelling of all the objects appearing in a given image and the generation of a bounding box identifying their position. The *fast* R-CNN [163] and *faster* R-CNN [164] are examples of *region-based* CNNs for object detection based on bounding boxes. Another relevant solution of this type is represented by the solo called *You only look once* (YOLO) network[19]. When building up a dataset for training this network, each detectable target is bounded with a box characterized by specific size and position in the whole image. If an object detection problem in which different targets can be associated to several (say, K) classes is considered, the YOLO network should be preferred to the other methods mentioned above because of: a) its ability to predict not only the size and the position of the bounding box associated with a given target, but also the probability that the target inside a given box belongs to a certain class; b) its architecture which, being based on a CNN, is simple and fast; c) its ability to learn very general representations of objects. The results illustrated in ref. [165] for various applications evidence that a YOLO network outperforms a R-CNN in terms of detection ability, since it produces a lower number of false negatives. However, it is important to remember that a YOLO network usually

[19]The name of this network has been inspired by the human ability of looking once at an image and instantly recognizing the objects it contains.

makes a significant number of localization errors and, consequently, achieves a limited accuracy. Better results are obtained if an improved architecture, known as YOLO v2 and originally proposed in ref. [166], is adopted. This new version of the YOLO network is still based on a convolutional architecture, but employs *anchor boxes*[20] in predicting the position of objects. The use of anchor boxes makes the learning procedure easy, since the network has only to adjust and refine their size in order to fit an object detected in the processed image. A specific application of the YOLO v2 network to a MIMO radar system is illustrated in Paragraph 3.7.2.

A recent research topic in the field of target detection and classification is represented by the use of *semantic segmentation*, that represents a powerful technique adopted for classifying the pixels of an image (a fixed set of classes is assumed in this case). The state of the art in semantic segmentation for image processing is represented by: a) *fully convolutional networks* (FCNs) [167], in which a convolutional network endowed with a pixel classification layer (instead of a fully connected layer) is used; b) SegNet [168] and U-Net [169], both based on a symmetrical encoder-decoder architecture. A more complicated method is represented by *instance segmentation*, whose aim is not only detecting and classifying all the objects appearing in an image, but also generating the segmentation of each instance appearing in the bounding box associated with each detected object. To accomplish the last task, the Facebook AI research group has proposed a new method called, *Mask*-R-CNN, that extends a Faster R-CNN by adding a branch for the prediction of the segmentation mask in each ROI [170].

It is important to note that the application of the above mentioned DL techniques to object detection and localization in radar images is still at an early stage. Despite this, specific DL methods inspired by FCNs and U-Net have been already implemented for detecting and estimating the position of different targets (like cars and other automotive targets) on the basis of range-Doppler-azimuth radar maps [171–173]. Moreover, the use of semantic segmentation in the radar field has been already investigated for the classification and localization of 3D point clouds of automotive targets, like cars, tractors and pedestrians; various results referring to automotive MIMO radars that operate at 77 GHz can be found in refs. [174] and [175]. The experimental results shown in these manuscripts evidence that the performance of the NNs employed for semantic segmentation substantially improves if radar data are fused with those one provided by optical sensors. It should not be forgotten that radar information can be augmented by an highly dense point cloud generated by a lidar device and that lidar data can be replaced by radar data in case of adverse weather or lighting conditions. An example of radar-centric automotive dataset based on radar, lidar and camera data for is described in ref. [176]; this dataset has been exploited in ref. [177] to test DL algorithms for 3D object detection.

3.6.3 Explainable artificial intelligence

Neural networks and sophisticated decision methods are currently employed in a number of applications to solve complicated tasks. The requirement of *transparency* is becoming more and more important in AI, especially when it is employed in autonomous systems. Unluckily, understanding which features are evaluated by a DNN in taking its

[20] Anchor boxes are a set of predefined bounding boxes having certain height and width.

decision is a complicated problem. Explainable artificial intelligence is a new branch of AI and concerns the problem of how the effectiveness of a deep network can be guaranteed [178]. An interesting method to improve the transparency of a DNN is based on the visualization of the features learned by each layer of the network [179]. The first layers of a DCNNs tested on radar images typically learn basic features, that depend on the size of their convolution filters. In fact, large (small) filters memorize general shapes (more specific properties), whereas some filters are also able to learn noise and clutter [128]. An alternative method to get some insight on the learning process of a CNN is based on the idea of identifying the parts of a radar image that are relevant for the classification of the object under test; such parts are also known as *spatial supports*. This approach allows to assess if a specific network is robust in taking its decision on the basis of a correct analysis of the given image. A specific technique, called *saliency extraction*, is based on this idea and, in particular, on the evaluation of the so called *saliency map*, as illustrated in ref. [180].

3.7 Experimental Results

In this section we show how specific ML and DL methods can be employed in a commercial colocated MIMO radar system to: a) classify three different human activities; b) estimate the range and DoA (azimuth) of a single target in a 2D propagation scenario. In both case, such methods are compared, in terms of accuracy and processing time; moreover, in case b), a comparison with deterministic methods is also made.

It is worth stressing that, unlike the previous sections, the results illustrated below do not originate from a synthetically generated dataset. In fact, the following tools have been exploited to generate them:[21]

1. A colocated FMCW MIMO radar manufactured by *Inras GmbH* [181]. This radar device, shown in Fig. 3.22-a) and employed to acquire all our measurements, operates in the E-band (the center frequency of its transmitted signal is $f_0 = 77$ GHz) and is equipped with a TX ULA and an RX ULA, consisting of $N_T = 2$ and $N_R = 16$ antennas, respectively (see Fig. 3.23-a)); even if, in principle, $2 \cdot 16 = 32$ virtual channels are available (see Paragraph 1.3), only $N_V = 31$ of them are exploited in our work, since two elements of the virtual array overlap.

2. A *pico-flexx camera* manufactured by *PMD Technologies Inc.* [84]. This time-of-flight camera, shown in Fig. 3.22-b) and employed as a reference sensor in our experiments, is based on a near-infrared vertical cavity surface emitting laser, and is able to provide a depth map or, equivalently, a three-dimensional point-cloud of a small region of the observed environment (its maximum depth is equal to 4 m, whereas its FOV is $62° \times 45°$).

3. A desktop computer equipped with a single i7 processor. All our software has been developed in the MATLAB and/or Python environment and run on this computer.

[21] The datasets employed in our experiments are available at: `http://www.sigcomm.unimore.it/downloads/`.

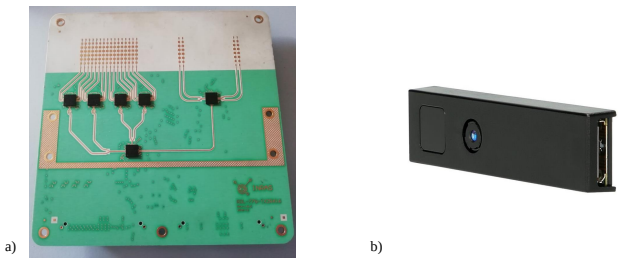


Figure 3.22: *a) Colocated MIMO radar system and b)* pico-flexx camera *employed in our experiments.*

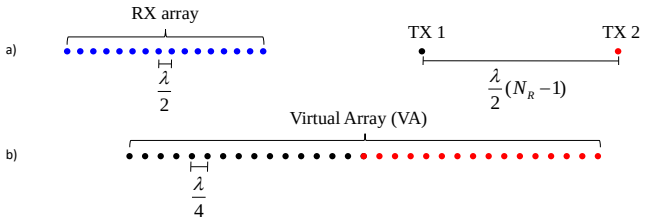


Figure 3.23: *Geometry of a) the physical TX and RX arrays and b) the corresponding virtual array of the radar device shown in Fig. 3.22-a).*

In the following two paragraphs, we provide various details about the experiments accomplished for the two specific applications mentioned above and illustrate the most relevant results we obtained.

3.7.1 Human activity classification

Our first experiment concerns the classification of following three different human activities: *walking*, *running* and *jumping*. The following choices have been made in the acquisition of our measurements:

1. The person whose activity has to be classified is alone and is in front of the employed radar device.

2. A single pair of TX-RX antennas is used (since angular information is not required).

3. The transmitted waveform is characterized by the following parameters: $N_c = 128$, $T = 128$ μs, $T_R = 32$ μs and $B = 1$ GHz (consequently, $\mu = 7.8 \cdot 10^{12}$ GHz/s; see eq. (1.4)).

4. At the receive side, analog-to-digital conversion is accomplished at the sampling frequency $f_s = 80$ MHz and $N = 1024$ samples are acquired over each chirp period and an oversampling factor $M_r = 4$ is considered for our processing.

Different classification methods have been tested for this application. First of all, we took into consideration the following five ML methods: a linear SVM technique, the K-NN technique (with $K = 4$), an Adaboost classifier with decision stumps as

weak learners (see Paragraph 3.2.2), a *customised* double stage SVM binary classifier (CSVM) and a specific version of the Adaboost, called *Stagewise Additive Modeling using a Multi-class Exponential loss function* (SAMME) [182]. As far as DL methods are concerned, we have taken into consideration a specific CNN only, since, as shown below, the preprocessed data feeding it can be interpreted as 2D images.

All these methods are fed by the matrices **E** and **G** defined at the end of Paragraph 1.3 (see eqs. (1.30) and (1.31)) and whose sizes are $N_f \times N_0'$ and $N_f' \times N_0'$, respectively (in all the experiments made for the considered application, $N_f = 143$, $N_0' = 256$ and $N_f' = 512$ have been selected). It is also worth remembering that the former (the latter) matrix is used to generate the spectrogram (the CVD) of the received signal. Examples of the spectrograms associated with the three possible activities are shown in Figs. 3.24-(a), -(b) and -(c) (note that the same time scale is used in all these figures), whereas an example of CVD is illustrated in Fig. 3.25. Moreover, in the last figure, two additional plots, one referring to the cadence frequency of the observed motion (left), the other one to its velocity (bottom), are also given for completeness. From Figs. 3.24-3.25 it is easily inferred that:

a) The period of the spectrogram (i.e., the distance between its consecutive peaks) is inversely proportional to the speed of the observed motion.

b) The shape of the spectrogram is influenced by the type of motion.

c) The CVD diagram contains important information regarding the motion and it is strictly related to the shape of the spectrogram. In fact, the principal components characterizing the observed motion can be identified in the CVD diagram in correspondence of the so-called *cadence frequencies*; each of these frequencies indicates how frequently a specific velocity component repeats in the observation interval.

An experimental campaign has been accomplished to build up an experimental dataset, that collects $N_t = 150$ observations equally divided among the three classes. Each observation refers to N_f consecutive frames, each consisting of N_c chirps, and is acquired over an observation interval whose duration is $T_O = 3$ s (each frame lasts $T_F = T_O/N_f = 21$ ms). Moreover, the q-th entry of the dataset $\mathcal{D}_o$ processed by the above mentioned ML methods is represented by the couple $(\mathbf{r}_q, t_q)$ (see eq. (3.41)), where

$$\mathbf{r}_q = [r_{q,0}, r_{q,1}, r_{q,2}, r_{q,3}] \tag{3.194}$$

is a 4D feature vector (so that $D_r = 4$) and t_q is a integer label identifying the specific activity which the vector $\mathbf{r}_q$ is associated with ($t_q = 0$, 1 and 2 if the observed person is walking, running or jumping, respectively). The first three elements of the vector $\mathbf{r}_q$ (3.194) depend on the value $\mathbf{G}_q = [G_{l,m}^{(q)}]$ of the matrix **G** computed for the q-th observation, since

$$r_{q,0} \triangleq \frac{\hat{l}_q}{N_f' \, T_F}, \tag{3.195}$$

$$r_{q,1} \triangleq \frac{1}{N_0'} \sum_{m=0}^{N_0'-1} \left(G_{\hat{l}_{q,0},m}^{(q)} - \mu_0^{(q)} \right) \cdot \left(G_{\hat{l}_{q,1},m}^{(q)} - \mu_1^{(q)} \right) \tag{3.196}$$

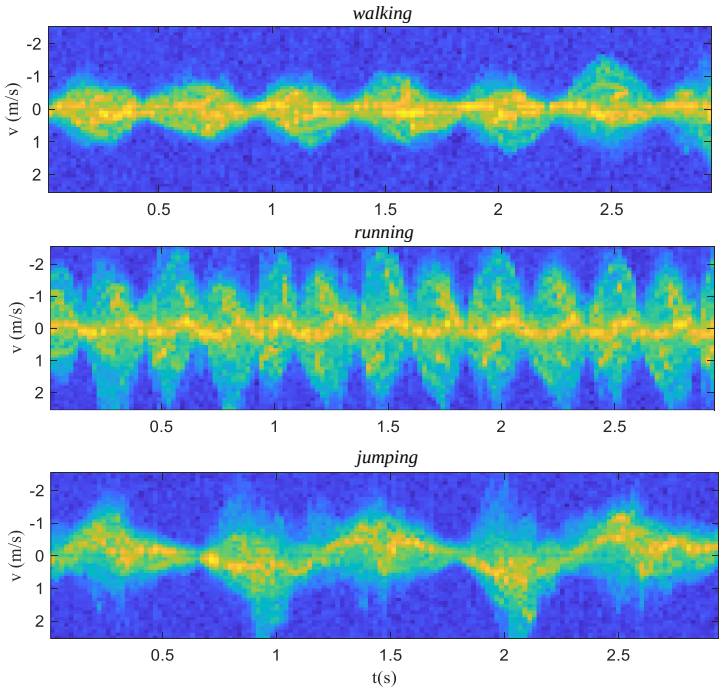


Figure 3.24: *Spectrograms observed for the following three different activities: walking (top), running (center) and jumping (bottom).*

and

$$r_{q,2} \triangleq \frac{1}{N_0'} \sum_{m=0}^{N_0'-1} \left(\left(G^{(q)}_{\hat{l}_{q,0},m} \right)^2 + \left(G^{(q)}_{\hat{l}_{q,1},m} \right)^2 \right); \tag{3.197}$$

here,

$$\hat{l}_q \triangleq \arg \max_{l \in \{0,1,...,N_f'-1\}} V^{(q)}_l, \tag{3.198}$$

$\hat{l}_{q,k}$ is the index identifying the k-th largest peak appearing in the sequence $\{V^{(q)}_l;\, l = 0,$ $1,\, ...,\, N_f' - 1\}$ (with $k = 0$ and 1),

$$V^{(q)}_l \triangleq \sum_{m=0}^{N_0'-1} G^{(q)}_{l,m}, \tag{3.199}$$

and $\mu^{(q)}_k$ is the mean of the elements of the $\hat{l}_{q,k}$-th row of the matrix $\mathbf{G}_q$, i.e. of the vector

$$\mathbf{G}^{(q)}_{\hat{l}_{q,k}} \triangleq \left[G^{(q)}_{\hat{l}_{q,k},0}, G^{(q)}_{\hat{l}_{q,k},1}, ..., G^{(q)}_{\hat{l}_{q,k},N_0'-1} \right]^T, \tag{3.200}$$

with $k = 0$ and 1. The last feature of $\mathbf{r}_q$ (3.194) (namely, the quantity $r_{q,3}$) depends on the value $\mathbf{E}_q$ of the matrix $\mathbf{E}$ computed for the q-th observation, since it represents the period of the spectrogram, i.e. the distance between two consecutive peaks observed along the time dimension. It is important to point out that:

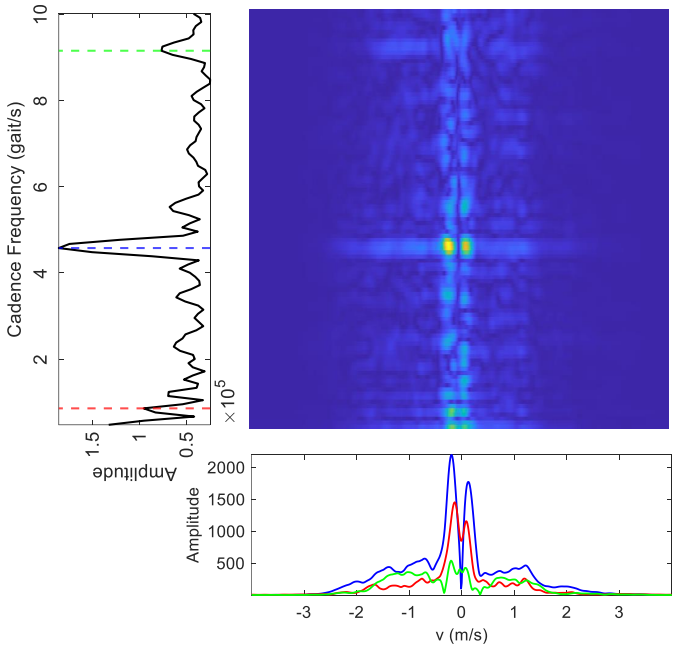


Figure 3.25: *Representation of a CVD and of two diagrams extracted from it (one providing information about cadence frequencies, the other one about velocities). In the diagram appearing on the left, the three strongest frequency components are identified by blue, red and green dashed lines; each line is associated with the velocity profile shown in the other diagram and having the same colour.*

1. The parameter $r_{q,0}$ (3.195) represents the strongest frequency component detected in the CVD diagram (see Fig. 3.25). The value of this parameter is expected to increase with the speed of the observed person.

2. The parameters $\hat{l}_{q,0}$ and $\hat{l}_{q,1}$ identify the two strongest frequencies (denoted $\hat{f}_{q,0}$ and $\hat{f}_{q,1}$, respectively) detected in the CVD referring to the q-th observation; such frequencies are evaluated as

$$\hat{f}_{q,k} = \frac{\hat{l}_{q,k}}{N'_f \, T_F}, \tag{3.201}$$

   with $k = 0$ and 1.

3. The parameter $\mu_k^{(q)}$ is the mean of the *velocity profile* expressed by the N'_0-dimensional vector $\mathbf{G}^{(q)}_{\hat{l}_{q,k}}$ (3.200).

4. The parameter $r_{q,1}$ (3.196) represents the covariance between the velocity profiles $\mathbf{G}^{(q)}_{\hat{l}_{q,0}}$ and $\mathbf{G}^{(q)}_{\hat{l}_{q,1}}$, whereas $r_{q,2}$ is the overall energy associated with both profiles; our experimental data have evidenced that the value of $r_{q,1}$ ($r_{q,2}$) decreases (increases) as the speed of the observed person gets larger (smaller).

5. The value of $r_{q,3}$ is inversely proportional to the speed of the observed person, since an increase of the speed shortens the period of the spectrogram.

As far as the adopted ML methods are concerned, the following choices have been made:

a) The K-NN classifier is structured as illustrated in Paragraph 3.2.2.

b) The classifiers based on the SVM and the Adaboost methods exploit the pairwise classification approach illustrated at the end of Paragraph 3.2.2. For this reason, they combine $L = K(K-1)/2 = 3$ *identical* binary classifiers (i.e., base learners).

c) The CSVM method is obtained by cascading two linear SVM binary classifiers (whose behaviour is described in Paragraph 3.2.2). The first SVM classifier (denoted SVM #1) distinguishes jumping from the rest of the activities and is fed by the feature vector $\mathbf{r}'_q = [r_{q,0}\,,\, r_{q,1}]$ (in this case, the scalar labels $t'_q = 1$ and $t'_q = -1$ are associated with jumping, and with walking and running, respectively). The second classifier (SVM #2) processes the observations related to running and walking only and is fed by the feature vector $\mathbf{r}''_q = [r_{q,2}\,,\, r_{q,3}]$ (in this case, the scalar labels $t''_q = 1$ and $t''_q = -1$ are associated with walking and running, respectively). The final predictions of the CSVM are generated on the basis of the SVM #1 (SVM #2) predictions for jumping (running and walking).

d) The employed version of the SAMME method is the one implemented in the Python library Scikit-learn [183] (namely, `sklearn.ensemble.AdaBoostClassifier`) and represents a specific version of the Adaboost technique for solving multi-class problems; in practice, it is based on a decision tree classifier characterized by two nodes (instead of a simple decision stump). This methods outperforms a classical Adaboost technique by simply emphasizing the weights assigned to misclassified points.

In our experiment, a N-fold cross-validation, with $N = 5$, has been employed. The accuracy achieved by the considered ML methods and the processing time they have required for training and prediction are listed in Table 3.3.

From these results, it is easily inferred that:

a) The accuracy is reasonably good in all cases (slightly above the 90%).

b) The Adaboost performs marginally better than the K-NN and SVM methods, at the price of substantially larger computation time.

c) The best trade-off in terms of performance and computation time is achieved by the K-NN technique.

d) The CSVM method requires a lower computational effort (especially in training) with respect to the method based on SVM and round-robin binarization. This is mainly due to the fact that the former approach employs only two learners, whereas the latter one three binary classifiers.

e) The SAMME algorithm achieves the same accuracy as the round-robin binarization of the classic Adaboost, even if its computation time (in both training and prediction) is approximately ten times smaller.

The ML methods tested in the first part of our experiment exploit a dataset of manually extracted features (see eqs. (3.194)-(3.197)). On the contrary, the CNN employed in the second part of our experiment is able to classify human activities by recognizing specific patterns directly in the matrix $\mathbf{E}$. A description of its architecture

	SVM	K-NN	ADA	CSVM	SAMME
Accuracy (%)	89	90	91	90	91
Training time (s)	0.1	0.03	4.5	0.06	0.45
Prediction time (s)	0.01	0.01	0.5	0.01	0.05

Table 3.3: *Accuracy, training time and prediction time evaluated for each of the ML methods considered for human activity classification.*

is provided in Table 3.4. The first three layers of the employed network are represented by three convolutional 2D filters, having size 15×5 and depths 4, 8 and 16; moreover, each filter feeds a linear rectifier, followed by a max pooling layer. Each max pooling layer allows to halve the size of the image made available by the previous layer, so that a significant dimensionality reduction is obtained. The first three layers are followed by another 2D convolutional filter with a batch normalization layer. The last layers are represented by a *fully-connected* (FC) and a *softmax* (Soft) layer transforming the residual 2D image in a vector of size 3, since three classes are considered. It is worth noting that the adoption of a CNN having a small depth is justified by the fact that spectrograms referring to the three activities are quite different, as exemplified by Fig. 3.24.

Layers	Filters	Size	Stride	Output
Convolutional + ReLu	4	15×5	1	$143 \times 53 \times 4$
Max pooling	-	15×5	2	$65 \times 25 \times 4$
Convolutional + ReLu	8	15×5	1	$65 \times 25 \times 8$
Max pooling	-	15×5	2	$26 \times 11 \times 8$
Convolutional + ReLu	16	15×5	1	$26 \times 11 \times 16$
Max pooling	-	3×3	2	$12 \times 5 \times 16$
Convolutional + BN + ReLu	3	3×3	1	$12 \times 5 \times 3$
FC + Soft	3	-	-	$1 \times 1 \times 3$

Table 3.4: *Architecture of the CNN employed for the classification of three human activities.*

The q-th entry of the dataset $\mathscr{D}_o$ processed by the employed CNN is represented by the couple $(\mathbf{r}_q, t_q)$, where, however, the observation $\mathbf{r}_q$ is represented by the value $\mathbf{E}_q$ taken on by matrix $\mathbf{E}$ in the q-th acquisition (the label t_q, instead, has the same meaning as in the ML case). Moreover, $N_f = 143$ and $\bar{N}_0' = 53$, and $\hat{N}_t = 150$ are assumed for the size of the matrix $\mathbf{E}$ and for the dataset $\mathscr{D}_o$, respectively. Network training is based on 60% of the whole dataset (the remaining part of the dataset has been equally divided to generate a validation set and a test set); moreover, it has been accomplished by an

SGD minimization procedure, which is characterized by a subset S of 4 training data samples, a learning rate $\gamma^{(i)} = 10^{-3}$ for any i and an overall number of epochs $N_E = 50$ (see eq. (3.73)). A 96% classification accuracy has been achieved in this case; therefore, the proposed DL method achieves a substantially better generalization capability that the ML counterparts described above. We should not forget, however, that this result is achieved at the price of a training time of about 25 s; this is substantially larger than that required by the considered ML methods (see Table 3.3). Finally, it is important to mention that the computation time required by the employed CNN for evaluation a new prediction is about 0.03 s and, consequently, is reasonably short and comparable with the one characterizing the considered ML methods.

3.7.2 Estimation of the range and azimuth of a single target

The second application we have investigated concerns the detection of a specific target moving on a 2D multi-target scenario, and the estimation of its range and azimuth. In our experiment, the target to be detected is an omnidirectional reflector, obtained by putting together eight corner reflectors (and inspired by the architecture of the *echo-master* corners used for maritime applications). This target is mounted, through a vertical carton support, on a Propeller Scribbler 3 mobile robot manufactured by Parallax Inc [184]. This robot has been programmed to move randomly inside a square white region delimited by four opaque black lines and whose side is equal to 2.5 m, as shown in Fig. 3.26; note that two corner reflectors have placed on the borders of this region in order to build a multi-target scenario. The following choices have been made in the acquisition of our measurements:

1. The whole antenna array shown in Fig. 3.23-a) is exploited, so that $N_V = 31$ distinct virtual channels are available at the receive side.

2. The waveform radiated by each TX antenna is characterized by the following parameters: $N_c = 1$, $T = 64$ μs, $T_R = 32$ μs and $B = 2$ GHz (consequently, $\mu = 3.13 \cdot 10^{13}$ GHz/s; see eq. (1.4)).

3. At the receive side, analog-to-digital conversion is accomplished at the sampling frequency $f_s = 40$ MHz and $N = 2048$ samples are acquired over each chirp period.

4. The reference position of the target with respect to a three-dimensional reference system is evaluated by means of the pico-flexx camera. This sensor is aligned with the radar system, being mounted on the same plastic support of the radar device and at a fixed distance from it (about 10 cm) along the vertical direction.

The following two supervised DL methods have been tested: a) a feed-forward NN exploiting some manually extracted features (further details about this method are provided below); b) a YOLO v2 NN for object detection (see Paragraph 3.6.2).

The q-th entry $(\mathbf{r}_q, \mathbf{t}_q)$ of the dataset $\mathcal{D}_o$ processed by the employed feed-forward NN is generated as follows. The label $\mathbf{t}_q$ associated with the q-th observation $\mathbf{r}_q$ is defined as

$$\mathbf{t}_q \triangleq [\hat{R}_q, \hat{\phi}_q] \tag{3.202}$$

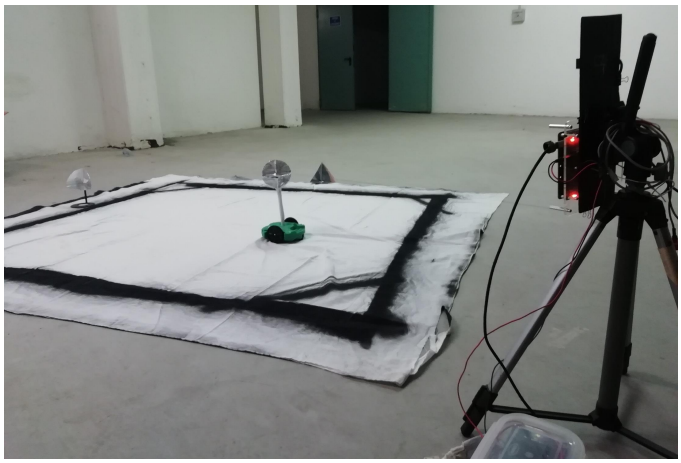


Figure 3.26: *Experimental-setup developed for our second application. The region of interest is delimited by an opaque and black line; two coner reflectors are located on its border. A robot, equipped with corner reflectors, moves randomly inside that area. The employed radar system and pico-flexx camera are placed on the tripod visible on the right.*

where $\hat{R}_q$ and $\hat{\phi}_q$ represent the estimates of the target range R_q and azimuth ϕ_q, respectively, evaluated on the basis of the point-cloud made available by our pico-flexx camera. Such a camera generates the $N_p \times 3$ matrix

$$\mathbf{P} \triangleq [\mathbf{x}\ \mathbf{y}\ \mathbf{z}], \tag{3.203}$$

collecting the 3D coordinates of $N_p = 38304$ distinct points; here, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are N_p-dimensional column vectors. The deterministic algorithm developed for the estimation of the target range and azimuth involves the computation of the estimates $(\hat{x}_q,\ \hat{y}_q, \hat{z}_q)$ of the target coordinates $(x_q,\ y_q, z_q)$ in the q-th observation; note that z_q (i.e., the target height) is assumed to be approximately known ($z_q \cong 0.4$ m). This algorithm consists of the following three consecutive steps:

1. The size of the search space for the couple $(\hat{x}_q,\ \hat{y}_q)$ is reduced by extracting the set[22]

$$\mathscr{S}_q \triangleq \{(x_{q,n}, y_{q,n}, z_{q,n}) | z_{\min} \leq z_{q,n} \leq z_{\max}; n \in \Delta_q\}, \tag{3.204}$$

   from the matrix $\mathbf{P}$ (3.203); here, Δ_q is a proper subset of the set of integers {0, 1, ..., $N_p - 1$} and consists of $\bar{N}_q$ elements, whereas $z_{\min} = 0.3$ m and $z_{\max} = 0.5$ m represent two thresholds.

2. The estimates

$$\hat{x}_q = 1/\bar{N}_q \sum_{n=0}^{\bar{N}_q - 1} x_{q,n}, \tag{3.205}$$

$$\hat{y}_q = 1/\bar{N}_q \sum_{n=0}^{\bar{N}_q - 1} y_{q,n} \tag{3.206}$$

[22]Note that, in this step, our prior knowledge about the target height is exploited.

and

$$\hat{z}_q = 1/\bar{N}_q \sum_{n=0}^{\bar{N}_q - 1} z_{q,n} \tag{3.207}$$

are computed. The estimate $\hat{z}_q$ (3.207) is exploited only to check if the vector $(\hat{x}_q, \hat{y}_q, \hat{z}_q)$ is meaningful, i.e. if the condition $\hat{z}_q \approx 0.4$ m is satisfied; if this does not occur, the thresholds appearing in the RHS of eq. (3.204) should be properly adjusted (i.e., $z_{\min}$ should be increased and/or $z_{\max}$ reduced) in order to improve the obtained accuracy.

3. The estimates

$$\hat{R}_q = \sqrt{\hat{x}_q + \hat{y}_q} \tag{3.208}$$

and

$$\hat{\phi}_q = \arctan\ (\hat{y}_q / \hat{x}_q) \tag{3.209}$$

are evaluated.

The observation $\mathbf{r}_q$ labelled by $\mathbf{t}_q$ (3.202) is defined as[23]

$$\mathbf{r}_q \triangleq \left[\hat{\psi}_{q,0}, \hat{\psi}_{q,1}, ..., \hat{\psi}_{q,N_V-1}, \hat{f}_q\right]^T, \tag{3.210}$$

where $\hat{f}_q$ is the frequency associated with the detected target (and estimated on the whole array) and $\hat{\psi}_{q,v}$ is the phase of the signal spectrum computed at the frequency $\hat{f}_q$ for the v-th virtual element (with $v = 0, 1, ..., N_V - 1$); note that the size of the vector $\mathbf{r}_q$ (3.210) is $D_r = N_V + 1 = 32$. The deterministic algorithm employed for the computation of the frequency $\hat{f}_q$ and the phases $\{\hat{\psi}_{q,v}\}$ forming $\mathbf{r}_q$ (3.210) consists of the following two steps:

1. *Coarse estimation of the target position* - The N-dimensional vector of the time domain samples acquired over the v-th virtual antenna (see eq. (3.18)) undergoes zero padding and FFT processing of order $N_0 = N \cdot M_r$ (in our experiment, $N_0 = 8192$, since $M_r = 4$). This produces the N_0-dimensional vector $\mathbf{X}_v^{(q)}$ (see eq. (3.20)), which is employed to compute the *power spectrum* $\mathbf{P}_v^{(q)} = [P_{v,0}^{(q)}, P_{v,1}^{(q)}, ..., P_{v,N_0-1}^{(q)}]$ on the basis of eq. (3.23). Then, given (see eq. (3.25))

$$\hat{l}_v^{(q)} \triangleq \arg \max_{\tilde{l} \in \{b_m, ..., b_M\}} P_{v,\tilde{l}}^{(q)}, \tag{3.211}$$

a target is detected on the v-th antenna if $P_{v,\hat{l}_v^{(q)}}^{(q)} > P_d$, where P_d is a proper threshold; here, the integer parameter b_m (b_M) identifies the frequency bin corresponding to the minimum (maximum) measurable range R_m (R_M). In our experiment, $P_d = 0.9$, and

$$b_m = \left\lfloor \frac{2\mu N_0 T_s R_m}{c} \right\rfloor = 42 \tag{3.212}$$

[23] *Unwrapped* phases are employed in this case, since they ease network training

and

$$b_M = \left\lfloor \frac{2\mu N_0 T_s R_M}{c} \right\rfloor = 147, \tag{3.213}$$

since $R_m = 1.0$ m and $R_M = 3.5$ m have been assumed. The procedure illustrated above is accomplished for each virtual channel (i.e., for $v = 0, 1, ..., N_V - 1$) and is employed to generate the set

$$\mathscr{S}_{\hat{l}} \triangleq \{\hat{l}^{(q)}_{v_k}; k = 0, 1, ..., \bar{N}_V - 1\}, \tag{3.214}$$

with $v_k < v_{k+1}$ for any k; the size $\bar{N}_V$ of this set is usually smaller that N_V, since: a) the target may be missed on one or more virtual channels (this occurs when the condition (3.211) is not satisfied); b) the elements of $\mathscr{S}_{\hat{l}}$ are required to be distinct. The elements of $\mathscr{S}_{\hat{l}}$ are collected in the vector $\hat{\mathbf{l}}_q = [\hat{l}^{(q)}_{v_0}, \hat{l}^{(q)}_{v_1}, ..., \hat{l}^{(q)}_{v_{\bar{N}_V-1}}]^T$. Then, the following vectors are computed: a) the $\bar{N}_V$-dimensional vector $\hat{\mathbf{f}}_q = [\hat{f}^{(q)}_{v_0}, \hat{f}^{(q)}_{v_1}, ..., \hat{f}^{(q)}_{v_{\bar{N}_V-1}}]^T$ and $\hat{\mathbf{R}}_q = [\hat{R}^{(q)}_{v_0}, \hat{R}^{(q)}_{v_1}, ..., \hat{R}^{(q)}_{v_{\bar{N}_V-1}}]^T$, that collect $\bar{N}_V$ estimates of the target frequency and range, respectively (these quantities computed on the basis of eqs. (3.27) and (3.29), respectively); b) the set of $\bar{N}_V$ vectors $\{\hat{\mathbf{A}}^{(q)}_{v_k}$; $k = 0, 1, ..., \bar{N}_V - 1\}$, where $\hat{\mathbf{A}}^{(q)}_{v_k} = [\hat{A}^{(q)}_{v_k,0}, \hat{A}^{(q)}_{v_k,1}, ..., \hat{A}^{(q)}_{v_k,N_V-1}]^T$ is made of the complex amplitudes evaluated over the whole virtual array on the basis of eq. (3.28) under the assumption that $\hat{l} = \hat{l}^{(q)}_{v_k}$ for any k; c) the set of $\bar{N}_V$ vectors $\{\hat{\boldsymbol{\psi}}^{(q)}_{v_k}; k = 0, 1 , ..., \bar{N}_V - 1\}$, where

$$\hat{\boldsymbol{\psi}}^{(q)}_{v_k} = [\hat{\psi}^{(q)}_{v_k,0}, \hat{\psi}^{(q)}_{v_k,1}, ..., \hat{\psi}^{(q)}_{v_k,N_V-1}]^T \tag{3.215}$$

and $\hat{\psi}^{(q)}_{v_k,l}$ is equal to the phase of the complex gain $\hat{A}^{(q)}_{v_k,l}$ for any k and l (see eq. (3.7)). Finally, each of the vectors $\{\hat{\mathbf{A}}^{(q)}_{v_k}\}$ undergoes zero padding, that increases their size to $\bar{N}_0 = 128$, and $\bar{N}_0$-th order FFT processing for azimuth estimation (see eqs. (3.37)-(3.39)). This produces the vector $\hat{\boldsymbol{\phi}}_q = [\hat{\phi}^{(q)}_{v_0}, \hat{\phi}^{(q)}_{v_1}, ..., \hat{\phi}^{(q)}_{v_{\bar{N}_V-1}}]^T$, collecting $\bar{N}_V$ different estimates of the target azimuth. Therefore, this step produces $\bar{N}_V$ distinct estimates $\{(\hat{f}^{(q)}_{v_k}$, $\hat{R}^{(q)}_{v_k}, \hat{\phi}^{(q)}_{v_k}); k = 0, 1 , ..., \bar{N}_V - 1\}$ of the target frequency, range and azimuth, respectively.

2. *Fine estimation of the target position* - A *single* estimate of the target frequency, range and azimuth is evaluated in this step on the basis of the $\bar{N}_V$ estimates $\{(\hat{f}^{(q)}_{v_k}$, $\hat{R}^{(q)}_{v_k}, \hat{\phi}^{(q)}_{v_k})\}$ available at the end of the previous step. This estimate is computed as follows. First, we compute

$$\hat{v}_q = \min_{\tilde{l} \in \mathscr{S}_{\hat{l}}} \left| \hat{\phi}_q - \hat{\phi}^{(q)}_{v_{\tilde{l}}} \right|, \tag{3.216}$$

under the constraint

$$\left| \hat{R}_q - \hat{R}^{(q)}_{v_{\tilde{l}}} \right| < R_{th}, \tag{3.217}$$

with $R_{th} = 0.3$ m; here, the quantities $\hat{R}_q$ and $\hat{\phi}_q$ are expressed by eq. (3.208) and eq. (3.209), respectively, and $\mathscr{S}_{\hat{l}}$ is the set defined by eq. (3.214). Then, the vector $\mathbf{r}_q$ (3.210) is evaluated as

$$\mathbf{r}_q = [\hat{\boldsymbol{\psi}}^{(q)}_{\hat{v}_q}, \hat{f}^{(q)}_{\hat{v}_q}]^T \tag{3.218}$$

where the vector $\hat{\boldsymbol{\psi}}_{\hat{v}_q}^{(q)}$ is expressed by eq. (3.215) with $v_k = \hat{v}_q$.

The entire dataset is generated by accomplishing the feature selection procedure expressed by eqs. (3.216)-(3.217) for any q. It is important to stress that, in our experiment, only a specific target must be selected for each observation. In fact, in a multiple target scenario like the one we are considering, it is hard to understand which elements of the set $\mathscr{S}_{\hat{l}}$ (3.214) are associated with the target of interest.

The vector $\mathbf{r}_q$ (3.218) generated by the deterministic procedure described above represents the input of our feed-forward NN, whose response is the bidimensional vector $\hat{\mathbf{t}}_q \triangleq [\hat{t}_{q,0}, \hat{t}_{q,1}]$; the elements of this vector represent the estimates of the range and the azimuth, respectively, of the target detected on the basis of the q-th observation (see eq. (3.202)). This network contains three hidden layers, consisting of $M_1 = 30$, $M_2 = 20$ and $M_3 = 10$ neurons (see Fig. 3.10). Each of them employs a ReLu, characterized by the transfer function

$$h(x) = x\,\mathrm{u}\,(x)\,, \tag{3.219}$$

where $\mathrm{u}\,(\cdot)$ denotes the *unit step function*. The estimates of the target range and azimuth are computed by the output layer, that contains two neurons only.

The size of the whole dataset acquired in our experiment is $\hat{N}_t = 1438$; 80% of it has been exploited for training the considered NN and the remaining part for its test (therefore, the size of the training set and that of the test are $N_t = 1150$ and $\bar{N}_t = 288$, respectively). Moreover, training has been accomplished by an *adam* optimizer; the batch size, the (constant) learning rate and the number of epochs selected for this procedure are $N_S = 4$, $\gamma = 10^{-3}$ and $N_E = 50$, respectively (see eq. (3.73)). The elements of the feature vector $\mathbf{r}_q$ (with $q = 0, 1, ..., N_t - 1$) have been scaled before applying it to the network (more specifically, a min-max normalization has been employed [185]); this ensures that the absolute value of such elements belongs to the interval [0, 1] and makes the training procedure more effective. The accuracy achieved by the network over the test set has been assessed by evaluating the RMSEs

$$\hat{\varepsilon}_R = \frac{1}{\sqrt{\bar{N}_t}} \left\| \hat{\mathbf{R}} - \hat{\mathbf{R}}_{NN} \right\| \tag{3.220}$$

and

$$\hat{\varepsilon}_\phi = \frac{1}{\sqrt{\bar{N}_t}} \left\| \hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}_{NN} \right\|, \tag{3.221}$$

where $\hat{\mathbf{R}}$ ($\hat{\boldsymbol{\phi}}$) is the $\bar{N}_t$-dimensional vector collecting the values of the target range (azimuth) estimated by means of the pico-flexx camera over the test set and $\hat{\mathbf{R}}_{NN}$ ($\hat{\boldsymbol{\phi}}_{NN}$) is the corresponding prediction computed by our NN ($||\boldsymbol{x}||$ denotes the Euclidean norm of the vector $\boldsymbol{x}$). The network performance has been also assessed by evaluating its *detection score*

$$A_c = \frac{N_C}{N_C + N_W}, \tag{3.222}$$

where N_C (N_W) is the number of trials in the test set in which both target azimuth and range have been correctly (wrongly) estimated (note that $N_C + N_W = \bar{N}_t$). In the q-th trial, estimation is deemed correct if $|\hat{R}_q - \hat{t}_{q,0}| \leq \Delta R$ and $|\hat{\phi}_q - \hat{t}_{q,1}| \leq \Delta\phi$, where $\Delta R = 20$ cm and $\Delta\phi = 5.5°$. It is worth pointing out that the values selected for the parameters ΔR and $\Delta\phi$ account for the limited resolution of the employed camera

Methods	$\hat{\varepsilon}_R$ (m)	$\hat{\varepsilon}_\theta$ (°)	AC (%)	*Training* (sec)	*Prediction* (msec)
FFT based	0.09	3.0	88	-	5
ANN	0.07	3.5	92	8	10
YOLO v2	0.03	1.5	98	398	20

Table 3.5: *Accuracy, detection score, training and prediction time of a deterministic estimation algorithm, a feed-forward NN and a YOLO v2 network.*

and radar system. Actually, the value selected for ΔR may look larger than expected, because of the high resolution that can be potentially achieved by both our radar device and pico-flexx camera. However, readers should not forget that the algorithm employed for the computation of $\hat{R}_q$ is not error free (see eq. (3.208) and (3.209)), especially when the cluster of points Δ_q (3.204) is not so dense or when the size $\bar{N}_q$ in eq. (3.207) is large. A low density in the set $\mathscr{S}_q$ could be observed when, for instance, the robot reaches the corners of the delimited area or in presence of optical disturbances.

The estimated accuracy and precision achieved by the adopted NN together with the time required for its training and testing are listed in Table 3.5. In the same table, the values of the same parameters evaluated on the basis of the deterministic algorithm employed for feature extraction are also provided; note that, for any q, this algorithm can be exploited to generate the estimates $\hat{R}_{\hat{v}_q}^{(q)}$ and $\hat{\phi}_{\hat{v}_q}^{(q)}$ of the target range and azimuth, on the basis of $\hat{v}_q$ (3.216) (note that $\hat{R}_{\hat{v}_q}^{(q)}$ and $\hat{\phi}_{\hat{v}_q}^{(q)}$ represent the $\hat{v}_q$-the element of the vectors $\hat{\mathbf{R}}_q$ and $\hat{\boldsymbol{\phi}}_q$, respectively). From these results it is easily inferred that: a) the NN is able to accurately predict the position of the target; b) it outperforms the deterministic algorithm in terms of both accuracy and precision; c) its prediction time is comparable with the computation time required by the deterministic algorithm.

In general, feed-forward NNs require a clever selection of their feature vector; for this reason, some expertise in radar systems is desirable when applying them to target detection and estimation. This problem can be circumvented by applying the YOLO v2 network (see Paragraph 3.6.2). Let us illustrate now how this network can be employed to solve the target detection and estimation problem taken into consideration in this paragraph. The q-the element of the collected dataset

$$\mathscr{D}_o \triangleq \{(\mathbf{r}_q, \mathbf{b}_q, t_q)\,; q = 0, 1, ..., \hat{N}_t - 1\} \tag{3.223}$$

consists of the following three components:

1. The noisy observation $\mathbf{r}_q = \mathbf{J}_q$, where $\mathbf{J}_q = [J_{l,m}^{(q)}]$ is a *range-azimuth matrix* having size $N_0 \times \bar{N}_0$ and computed on the basis of the measurements acquired in the q-th trial. The element on the l-th row and the m-th column of $\mathbf{J}_q$ is defined

as

$$J_{l,m}^{(q)} \triangleq \frac{1}{N_0 \bar{N}_0} | \sum_{v=0}^{\bar{N}_0-1} \sum_{n=0}^{N_0-1} S_{l,m} |, \tag{3.224}$$

where

$$S_{l,m} = \hat{r}_{v,n}^{(ZP)} \exp\left(-j2\pi n \, f_l T_s\right) \exp\left(-j2\pi v \frac{d}{\lambda} s_m\right) \tag{3.225}$$

with $l = 0, 1, ..., N_0 - 1$ and $m = 0, 1, ..., \bar{N}_0 - 1$; here, $\hat{r}_{v,n}^{(ZP)}$ is the n-tn element of the N_0-dimensional vector $\mathbf{r}_v^{(ZP)}$that results from zero padding N-dimensional vector of time domain signal samples acquired over the v-th virtual antenna, T_s the sampling period, $f_l \triangleq l/(N_0 T_s)$ is the center frequency of the l-th frequency bin and $s_m \triangleq 2\,(m - \bar{N}_0/2)/\bar{N}_0$ is the m-th *normalized spatial frequency*. Note that the matrix $\mathbf{J}_q$ can be computed through a $N_0 \times \bar{N}_0$-order 2D FFT, and that the range and azimuth associated with $J_{l,m}^{(q)}$ (3.224) are (see eqs. (3.29) and (3.39), respectively)

$$\bar{R}_l = f_l \frac{c}{2\mu} \tag{3.226}$$

and

$$\bar{\phi}_m = \arcsin s_m, \tag{3.227}$$

respectively.

2. The vector

$$\mathbf{b}_q = [l_q, m_q, w_q, h_q] \tag{3.228}$$

describing the bounding box associated with the detected target; here, the couple of integers (l_q, m_q) identifies the frequency bin and the normalised spatial frequency, respectively, corresponding to the center of the box and w_q (h_q) represents the width (height) of the box itself.

3. The label t_q; this equal to 1 (-1) if a target is detected (absent).

In our experiment, we have selected $N_0 = 8192$ and $\bar{N}_0 = 128$ in the computation of the elements of the matrix $\mathbf{J}_q$. However, since $R_m = 1.0$ m ($R_M = 3.5$ m) and $\phi_m = -55°$ ($\phi_M = 55°$) have been assumed for the minimum (maximum) range, the $N_0 \times \bar{N}_0$ matrix $\mathbf{J}$ has been resized to an $\bar{N}_l \times \bar{N}_m$ matrix, where $\bar{N}_l = b_M - b_m + 1 = 106$ (the values of the parameters b_m and b_M are expressed by eqs. (3.212) and (3.213), respectively), $\bar{N}_m = d_M - d_m + 1 = 106$, with

$$d_m = \left\lfloor \frac{\bar{N}_0}{2} \left(s_m + 1\right) \right\rfloor = 11 \tag{3.229}$$

and

$$b_m = 42. \tag{3.230}$$

In addition, a square shape with $w_q = h_q = 12$ has been always assumed for the bounding box; its parameters l_q and m_q have been computed as

$$l_q = \arg \min_{b_m \le \bar{l} \le b_M} \left| \hat{R}_q - R_{\bar{l}} \right| \tag{3.231}$$

and as

$$m_q = \arg \min_{d_m \leq \tilde{m} \leq d_M} \left| \hat{\phi}_q - \phi_{\tilde{m}} \right| \tag{3.232}$$

where $\hat{R}_q$ ($\hat{\phi}_q$) is expressed by eq. (3.208) (eq. (3.209)) and $R_{\tilde{l}}$ ($\phi_{\tilde{m}}$) by eq. (3.226) (eq. (3.227)) with $l = \tilde{l}$ ($m = \tilde{m}$). The size of the dataset $\mathcal{D}_o$ (3.223) is $\hat{N}_t = 1438$; 80% of its elements are used for training and the remaining part for testing; consequently, the sizes of the training set and the test set are $N_t = 1150$ and $\bar{N}_t = 288$, respectively. *Data augmentation* has been performed on the training and test set in order to reduce network overfitting, since their sizes are not so large; this procedure consists in randomly flipping and scaling the input image and the associated box.

The architecture of the employed network is summarized in Table 3.6. It consists of a cascade of 22 layers and is fed by a normalized version of the resized range-azimuth matrix generated through the procedure illustrated above and having size $\bar{N}_l \times \bar{N}_m = 106 \times 106$. Each of its first two convolutional layers has stride $S = 2$ and is followed by a max pooling layer for dimensionality reduction. The use of a *batch normalization* (BN) layer after each convolutional layer allows to avoid overfitting, since the dataset size is not so large; consequently, other forms of regularization (as dropout) are not required. The activation function at the end of each convolutional layer is ReLu (see eq. (3.219)). The filter depth in the last convolutional layer must be proportional to $N_A \cdot (N_{P_A} + K)$, where N_A is number of anchor boxes, N_{P_A} is the number of predictions per each anchor and K is the number of classes (see refs. [165] and [166]). Since, in our test, $N_A = 1$, $N_{P_A} = 5$ and $K = 1$ (if the background is ignored), the selected filter depth is equal to 6. A *transform* layer and an *output* layer are also included in the architecture of the adopted network. The former layer improves network stability in predicting the possible locations for the bounding box, whereas the latter one refines the estimate of the bounding box location.

If the N_K candidate boxes $\{\mathbf{b}_q[k] = [l_q[k], m_q[k], w_q[k], h_q[k]]^T;\ k = 0, 1, ..., N_K - 1\}$ (all labelled by $t_q = 1$) are identified by network, the index $\hat{k}_q$ of the bounding box

$$\hat{\mathbf{b}}_q = \left[l_q[\hat{k}_q], m_q[\hat{k}_q], w_q[\hat{k}_q], h_q[\hat{k}_q] \right] \tag{3.233}$$

best fitting the ground truth box is evaluated as

$$\hat{k}_q = \arg \max_{\tilde{k} \in \{0,1,...,N_K-1\}} I_{\tilde{k}}^{(q)}, \tag{3.234}$$

where

$$I_{\tilde{k}}^{(q)} = \frac{A_{BG}^{(q)} \cap A_{BP}^{(q)}}{A_{BG}^{(q)} \cup A_{BP}^{(q)}} \tag{3.235}$$

is the *intersection over union* (IOU) associated with the k-th candidate box; here, $A_{BG}^{(q)}$ ($A_{BP}^{(q)}$) represents the surface of the ground truth (predicted) bounding box referring to the q-th observation. In our experiment, a target is detected if $I_{\tilde{k}}^{(q)} > I_{th}$, where $I_{th} = 0.1$ is a properly selected threshold. Once the predicted bounding box $\hat{\mathbf{b}}_q$ (3.233) is known, the estimate of the target range (azimuth angle) is evaluated by setting $l = l_q[\hat{k}_q]$ ($m = m_q[\hat{k}_q]$) in eq. (3.226) (eq. (3.227)); note that the values selected for the parameters l and m identify the center of the predicted bounding box.

Layers	Filters	Size	Stride	Output
Convolutional + BN + ReLu	16	5×5	2	$52 \times 52 \times 16$
Max pooling	-	2×2	2	$26 \times 26 \times 16$
Convolutional + BN + ReLu	32	5×5	2	$12 \times 12 \times 32$
Max pooling	-	2×2	2	$6 \times 6 \times 32$
Convolutional + BN + ReLu	64	3×3	1	$6 \times 6 \times 64$
Max pooling	-	2×2	2	$3 \times 3 \times 64$
Convolutional + BN + ReLu	128	3×3	1	$3 \times 3 \times 128$
Convolutional + BN + ReLu	256	3×3	1	$3 \times 3 \times 256$
Convolutional + BN + ReLu	512	3×3	1	$3 \times 3 \times 512$
Convolutional	6	1×1	1	$1 \times 1 \times 6$

Table 3.6: *Architecture of the CNN employed for target detection and estimation.*

The training procedure of the adopted network has been carried out through the SGD algorithm; a batch size $N_S = 10$, a learning rate $\gamma^{(i)} = 10^{-3}$ and a number of epochs $N_E = 25$ have been assumed (see eq. (3.73)). The testing procedure has evidenced that this network is able to predict the bounding boxes characterized by $I_{\tilde{k}}^{(q)} > 0.1$ over 98% of the test set. A realization of range-azimuth map associated with the matrix $\mathbf{J}$ and of the associated ground truth and the predicted bounding boxes around the detected target is illustrated in Fig. 3.27, where the position of the two corner reflectors placed on the border of the area of interest is also identified (see Fig. 3.26). These results deserve the following comments:

a) The network is able to detect the target on the basis of the value of range and azimuth obtained through the pico-flexx camera.

b) In the considered case, the IOU between the ground truth bounding box (red line) and the predicted one (green line) is quite large, being equal to 0.73. Consequently, the estimate of the position of the target (green circle) is very accurate and certainly much better than the one used as reference (red cross) (note that $|\hat{R}_q - R_{l_q[\hat{k}_q]}| = 0.001$ m and $|\hat{\phi}_q - \phi_{m_q[\hat{k}_q]}| = 0.9^\circ$ in this case).

The values of the achieved accuracy (evaluated in terms of the RMSEs $\hat{\varepsilon}_R$ (3.220) and $\hat{\varepsilon}_\phi$ (3.221)), the detection score (3.222), and the computational time required for training and testing are listed in Table 3.5. From these results it is easily inferred that:

a) The YOLO network outperforms our (deterministic) FFT-based method in target detection and estimation.

b) The value of the YOLO detection score A_c (3.222) is really high and better than that provided by the feed-forward NN.

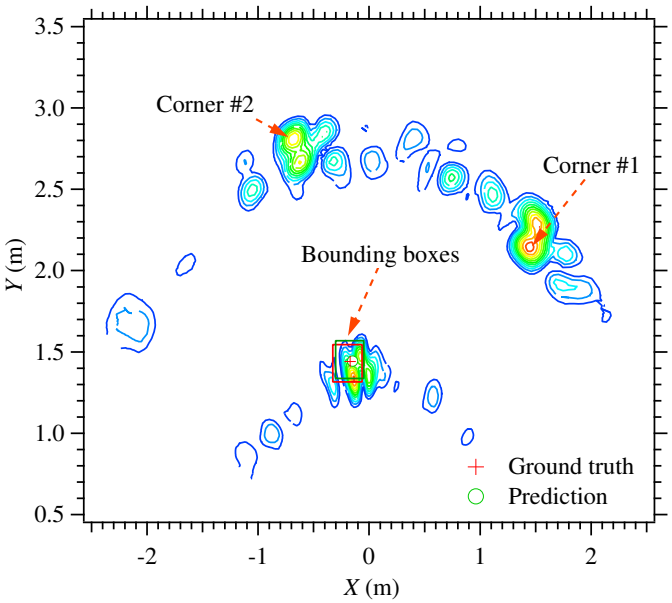


Figure 3.27: *Range-azimuth map referring to the scenario illustrated in Fig. 3.26. The ground truth bounding box and the position of the target are identified by a red square and a red cross, respectively. The prediction of the network, together with the estimated bounding box, are identified by a green circle and a green square, respectively.*

c) The YOLO RMSE $\hat{\varepsilon}_R$ ($\hat{\varepsilon}_\phi$) is smaller than (close to) the one characterizing the feed-forward NN.

These results lead to the conclusion that the YOLO network is more robust than the feed-forward NN. Note also that, even if the complexity of this network is higher than those of the other two methods, the time it employs for computing its prediction is not too long, being in the order of few milliseconds.

Since the YOLO v2 network tries to solve also a binary classification problem, other two important parameters for evaluating its performance are its *precision*

$$P = \frac{T_P}{T_P + T_N} \tag{3.236}$$

and its *recall*

$$R = \frac{T_P}{T_P + F_N}, \tag{3.237}$$

where T_P (T_N) represents the overall number of *true positives* (*true negatives*), i.e. the number of targets (false targets) classified correctly, and F_N is the overall number of false targets classified as targets. The *precision versus recall* plot evaluated in the considered experiment is shown in Fig. 3.28. These results lead to the conclusion that, in this case, the precision remains high for large values of the recall and drops steeply only when the recall exceeds 0.9. The area under the curve shown in Fig. 3.28 represents the so called *mean average precision* (mAP); in this case, we have found

that $mAP = 93\%$ (note that the value of this parameter is expressed as a percentage since the precision P (3.236) and the recall R (3.237) are defined in the range $[0\,,\,1]$).

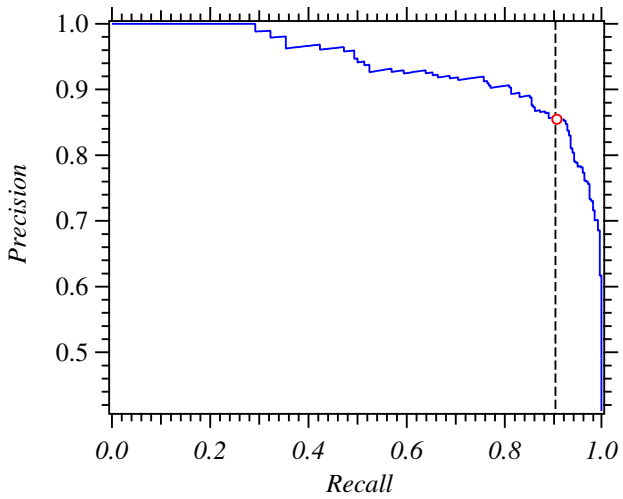


Figure 3.28: *Representation of the precision versus recall plot referring to the YOLO v2 network employed in the second application. Note that, if the recall exceeds the threshold identified by the vertical dashed line, the precision decreases steeply.*

3.8 Conclusions

Thanks to recent developments of electronic technology and advances in signal processing algorithms, colocated MIMO radar systems have reached a stage of maturity that allows their adoption in a number of applications. Existing algorithms developed for target detection and estimation in radar systems equipped with antenna arrays do not always provide satisfactory performance in such applications, because of the complexity of colocated MIMO radar devices and of the propagation scenario in which they operate. This motivates the adoption of machine learning and deep learning techniques, since these are able to extract relevant information from the available data in the absence of an accurate mathematical description of the behaviour of radar devices and of the mechanisms of electromagnetic propagation. Even if important steps have been made in this field in the last years, significant research efforts are still required to make the adoption of these techniques in commercial systems a reality. In this chapter, after providing essential information about MIMO radars and the deterministic algorithms they employ for target detection and estimation, we have shown how some learning methods can be exploited to solve simple classification and regression problems in FMCW radar systems operating in a 2D propagation scenario and in the presence of point targets. This allows readers to become familiar with some basic concepts and tools originating from the fields of radar systems and learning methods. Then, various applications of learning methods to specific technical problems have been illustrated and relevant trends in research on MIMO radars have been identified. Finally, the application of machine learning and deep learning methods to two specific problems, namely human

activity classification and range azimuth estimation, has been investigated. Our numerical results, based on experimental datasets acquired through a colocated MIMO device operating at 77 GHz, allow readers to grasp how such methods can be exploited to solve real world problems. A pervasive use of such methods should be expected in the near future, as understanding of the learning methods described in this chapter is becoming deeper and deeper, and MIMO technology is continuously evolving.

4. Vital Signs Monitoring

4.1 Introduction

Monitoring human vital signs, like heart and respiration rates, represents a routine practice to detect patient deterioration. Changes in vital signs can reveal the existence of serious medical problems; for this reason, early identification of these changes can improve survival rates in several conditions [186]. Vital signs monitoring is often accomplished by means of wearable health devices [187]; this is due to the fact that these devices enable continuous monitoring during daily activities. However, in various situations, like in the case of infected patients, or of patients suffering from mental illness or affected by severe burns or injuries, the use of wearable sensors is not possible or recommended. In such cases, the use of non-contact monitoring devices, like radar systems, can help healthcare professionals by providing critical information about patient state [188]. The application of radar devices to this field and, in particular, to the estimation of the heart and respiration rates, has become an active research area in recent years [189–192]. Actually, the first experimental results in this field date back to 1975, when the use of short-range radar technology was proposed to non-invasively acquire respiratory information by comparing a microwave signal with its echo reflected from the chest of a patient [193, 194]. In the following years, the possibility of employing radar systems for the wireless detection of the physiological movements due to both heartbeat and respiration has been shown [195–198]. This has motivated the investigation of the use of this technology in a number of medical applications, including adult and neonatal sleep monitoring [199–201], disaster medicine (e.g., in the detection of human vital signs under rubbles after earthquakes [202]) and lung cancer radiotherapy [203].

In the last two decades, few review articles about radar-based monitoring of vital signs have been published [198, 204–209]; however, they have a limited scope since they concern the use of specific technologies. In fact, on the one hand, [198, 204–206] and [210] focus on *continuous wave* (CW) Doppler radars, *ultra-wide band* (UWB) and radars equipped with a single transmit/receive antenna (i.e., *single-input single-output*, SISO, radars), respectively. On the other hand, [207] takes into consideration radars equipped with antenna arrays (i.e., *multiple-input multiple-output*, MIMO, radars) and illustrates the advantages they offer with respect to their SISO counterparts. This has motivated the writing of this chapter, that aims at offering a tutorial overview of radar-based monitoring of vital signs and at providing some essential tips for its use in a research laboratory. More specifically, in the remaining part of the chapter, essential information about radar-based monitoring, cardiovascular and respiration physiology,

and the modelling of chest displacement is provided. Then, we focus on the available radar technologies, and describe various radar architectures and signal processing methods developed for radar-based estimation of vital signs; in this description, all the available options are taken into consideration, and their pros and cons are illustrated. This is followed by various technical considerations, formulated in the light of the experience acquired in our experimental activities on radar-based monitoring, and by the analysis of some numerical results based on the measurements we acquired through different radar devices; our considerations concern the essential technical requirements that radar devices employed in this field should have and some essential guidelines to be followed in conducting experimental campaigns. The study of all this allows us to lay the foundations for understanding specific applications of radar-based monitoring and relevant research trends in this field.

This chapter is organized as follows. In Section 4.2 the basic principles, challenges and objectives of radar-based monitoring of vital signs are illustrated. In Section 4.3 the physiological fundamentals of human cardiovascular and respiration activities are provided and simple mathematical models describing the dynamics of chest displacement due to these activities are described. Section 4.4 is devoted to the four radar technologies employed in radar-based monitoring of vital signs, and to SISO and MIMO radar architectures; for each architecture, simple mathematical models are provided for the samples of the baseband signal received in the presence of a single point target. An overview of the most important deterministic and learning-based signal processing techniques employed for vital signs monitoring is offered in Section 4.5. In Section 4.6 we illustrate some basic guidelines to be followed in conducting experimental activities in the considered field, analyse the problem of extracting hear rate from radar measurements and comment on the assessment of estimation accuracy. The applications of the considered radar-based techniques for vital signs monitoring to heart and breath rate estimation, heart sound monitoring and heart rate variability estimation are discussed in Section 4.7, whereas current research trends on radar-based monitoring of vital signs are illustrated in Section 4.8. Finally, some conclusions are offered in Section 4.9.

4.2 Radars for Vital Signs Monitoring: Basic Principles, Objectives and Challenges

Radar-based monitoring of vital signs is based on the idea that the chest-wall of human bodies reflects the electromagnetic waves generated by a radar placed in front of it and that its quasi periodic vibrations, resulting from respiration and heartbeat, *modulate such waves*. Therefore, in principle, essential information about vital signs, i.e. heart and breath rates, can be extracted from the reflected electromagnetic waves [211–213] and a fundamental objective, namely contactless monitoring of vital signs, can be achieved. Note also that, compared with traditional methods, such as pneumotachography and electrocardiography, radar systems make continuous and timely *breath rate* (BR) and *heart rate* (HR) monitoring possible without entailing an additional work load for nurses. For these reasons, radars represent a favourable option for hospital monitoring, especially in the case of severe burn or infectious disease patients, sudden infant death syndrome monitoring, sleep apnoea monitoring, elderly home healthcare

and psychology studies. Moreover, radar signals can be processed to extract more refined medical information and, in particular, to detect anomalous alterations in the sequence of heart beats.

Even if the usefulness of radar systems in vital signs monitoring is now globally recognized and a wide literature about such systems is available, various challenges concerning signal processing techniques for vital signs extraction are still open in this research field; here, we limit to mention:

a) *The estimation of heart rate* – As illustrated in Subsection 4.3.2, the vibrations due to heartbeat are significantly weaker than those originating from respiration. For this reason, the contribution of the first phenomenon to the radar signal may be hidden by that related to the second (and much stronger) one. This makes the task of estimating HR much harder than that estimating BR; additional details about this issue can be found in Subsections 4.7.1 and 4.6.3.

b) *The identification of anomalous alterations of heart beats* – This challenge concerns the possibility of estimating HR *variability* (HRV) and detecting *heart sounds*; these issues are discussed in Subsections 4.7.3 and 4.8.2, respectively.

c) *The simultaneous observation of the vital signs of multiple people* – In principle, MIMO radar systems can be exploited to detect and estimate the vital signs of multiple people located in a restricted area (e.g., in the same room; see Subsection 4.5.2 for further details). However, experimental results supporting the feasibility of this idea and involving a significant number of people are still missing in the technical literature (e.g., see [207, Table I, Sect. IV]).

Finally, it is worth mentioning that the experimental results available in the technical literature about radar-based monitoring of vital signs concern heterogeneous radar technologies and that there is not a broad consensus on the best technology to be adopted in real world systems. For this reason, it is important to analyse the pros and cons of each option available on the market; this issue is discussed in Section 4.4.

4.3 Physiological Fundamentals and Mathematical Modelling

In this section we first provide readers with the physiological fundamentals of heart and lung functions. Then, we concentrate on the movements of the chest surface in human beings and illustrate some mathematical models describing them.

4.3.1 Basics of cardiovascular and respiration physiology

The human heart is made of two separated systems, called *left* and *right sides*. Each side consists of two chambers, namely an *atrium* and a *ventricle*, which are separated and connected by an *atrioventricular valve*. The main function of the left side is to pump oxygenated blood through the aorta and the other arteries to peripheral tissues and organs. The right side, instead, is in charge of pumping deoxygenated blood through pulmonary arteries to lungs. Each side is connected to arteries through the so called *semilunar valves*.

The *cardiac cycle* consists of a rhythmic sequence of contractions (*systoles*) and relaxations (*diastoles*) of the heart; these events occur simultaneously in the left and

right sides. During each cardiac cycle, *sounds* are generated by the action of the heart muscle and the vibrations of the cardiac valves. In the case of an healthy adult heart, two heart sounds are detected. The first one is caused by the contraction of the ventricular muscle during systoles and the closing of atrioventricular valves. The second heart sound, instead, is due to the closure of aortic and pulmonary valves. Hemodynamic properties and HR can be determined by recording heart sounds. As a matter of fact, such sounds can help the physician in the diagnosis of potential cardiovascular diseases [214,215]. The physiologic HR at rest is 60 *beats per min* (bpm) up to 100 bpm; values below this range (above it) characterise the so called *bradycardia* (*tachycardia*).

A *breathing cycle* consists of two consecutive phases, known as *inspiration* and *expiration*. In the first phase, thanks to the contraction of the diaphragm and of the intercostal muscles, the thoracic volume increases; this results in a sub-atmospheric pressure that allows air to flow through the airways into the lungs. Then, air oxygen is absorbed into the blood and carbon dioxide is transferred from the blood to the inhaled air through the alveolar-capillary membrane. Expiration, instead, is caused by the elastic recoil of the lungs and relaxing muscles. In this phase, the deoxygenated air can flow out of the lungs by increased pressure [216, 217]. The physiologic BR at rest is 12 up to 25 acts per minute, whereas values below this range (above it) characterise the so called *bradypnea* (*tachypnoea*).

4.3.2 Modelling of chest displacement

Let us focus now on the problem of modelling the chest displacement of an arbitrary patient. As illustrated in the previous section, his/her inspiration (expiration) phase produces an expansion (compression) of the thoracic wall. Moreover, the vibrations due to his/her heart beat overlap with the thoracic breathing movement. In principle, the time evolution of both the thoracic breathing movement and the cardiac vibrations of the considered patient can be measured by a radar system placed in front of his/her chest at a fixed distance. In the absence of random and large-scale body movements, the displacement $\Delta R\,(t)$ of the chest surface measured by the radar system at the instant t can be modelled as [218]

$$\Delta R\,(t) \triangleq R\,(t) - R_0 = \delta_b\,(t) + \delta_h\,(t)\,, \tag{4.1}$$

where $R(t)$ (R_0) is the radar-chest distance at time t (distance in the absence of respiration) and $\delta_b(t)$ ($\delta_h(t)$) represents the breath (heart) contribution to $\Delta R(t)$. It is important to stress that:

a) the displacement $\delta_b\,(t)$ ($\delta_h\,(t)$) is usually assumed to be *periodic* with period T_{BR} (T_{HR}), with $T_{\mathrm{BR}} > T_{\mathrm{HR}}$;

b) the contribution of $\delta_h\,(t)$ to the displacement $\Delta R\,(t)$ is usually small with respect to that due to $\delta_b\,(t)$; in fact, if $\delta_{b,M}$ ($\delta_{h,M}$) denotes the maximum absolute value of $\delta_b(t)$ ($\delta_h(t)$), it is known that $1 \leq \delta_{b,M} \leq 5$ cm and $1 \leq \delta_{h,M} \leq 9$ mm [219,220].

A simple mathematical model describing the breath displacement $\delta_b(t)$ within each period has been proposed in [221]. According to this model, the displacement in the

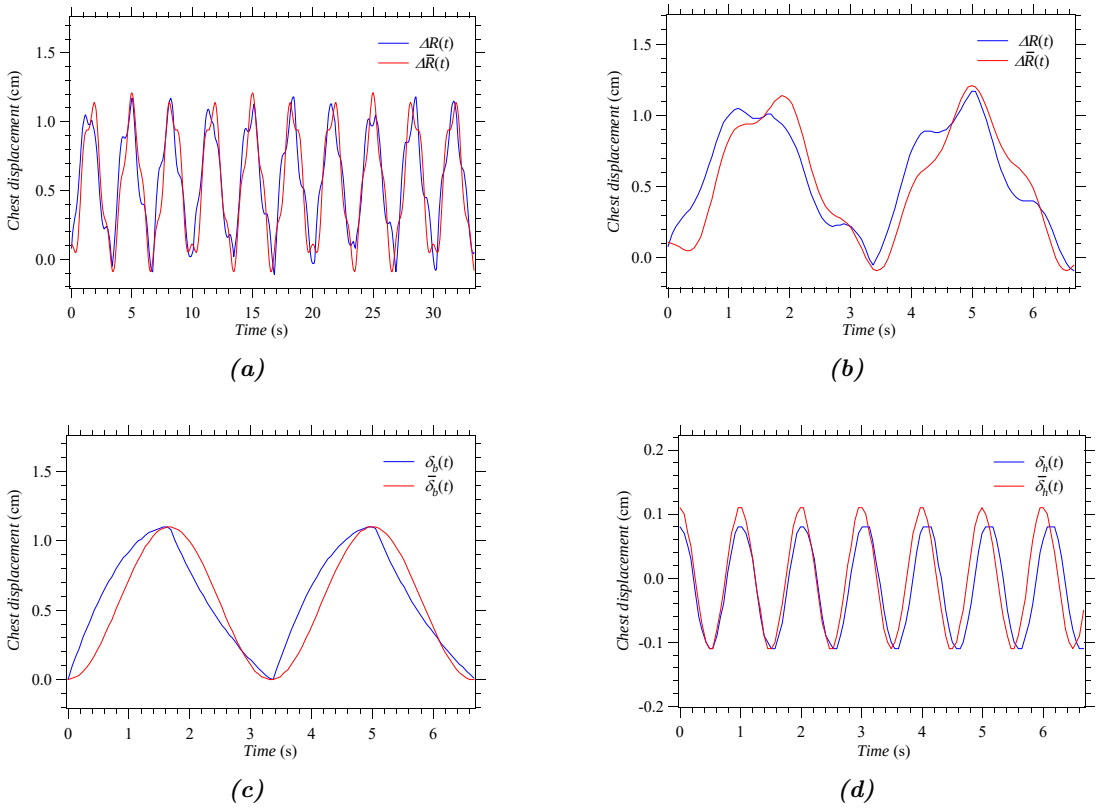


Figure 4.1: *Representation of: a) a computer generated chest displacement (multiple periods of breathing activity are considered); b) a computer generated chest displacement (a couple of consecutive periods of breathing activity are considered); c) the contributions of breathing activity to the chest displacement shown in sub-figure b); d) the contributions of heart activity to the chest displacement shown in sub-figure b). In all these figures, a blue (red) line is used to identify the signals generated on the basis of (4.1) ((4.6)).*

inspiration phase is described by the parabolic profile

$$\delta_b(t) = -\frac{\delta_{b,M}}{T_i\,T_e}(t-t_0)^2 + \frac{\delta_{b,M}T_{BR}}{T_i\,T_e}(t-t_0), \tag{4.2}$$

with $t \in [t_0, t_0+T_i)$, whereas that in the expiration phase by the exponential profile

$$\delta_b(t) = \frac{\delta_{b,M}\exp\left(-\frac{T_e}{\tau}\right)}{1-\exp\left(-T_e/\tau\right)}\left[\exp\left(-\frac{(t-t_0)-T_{BR}}{\tau}\right)-1\right], \tag{4.3}$$

with $t \in [t_0+T_i, t_0+T_{\rm BR})$; here, t_0 is the initial instant of the considered breathing period, T_i (T_e) is the duration of the inspiration (expiration) phase, τ is the time constant of the expiratory profile; note that $T_{\rm BR} = T_i + T_e$. The displacement due to

Table 4.1: *Minimum and maximum values of the parameters appearing in (4.1)-(4.5). The values selected for generating the simulation results shown in Fig. 4.1 are also listed.*

Params	T_{BR} (s)	T_{HR} (s)	T_i (s)	T_e (s)	$\delta_{b,\mathrm{M}}$ (cm)	$\delta_{h,\mathrm{M}}$ (cm)	τ (s)	a_1 (cm)	a_2 (s)	a_3 (s)
Minimum	1.5	0.3	0	0	1	0.1	0	0	0	0
Maximum	6	1.2	T_{BR}	T_{BR}	5	0.9	T_{BR}	1	1	1
Simulation	3	1	$0.5T_{\mathrm{BR}}$	$0.5T_{\mathrm{BR}}$	1.1	0.11	$0.5T_{\mathrm{BR}}$	0.2	0.5	0.8

the cardiac activity, instead, can be modelled as [218]

$$\begin{aligned}\delta_h(t) =&\delta_{h,M}\cos\left(w_1(t-t_1)+\gamma_h(t)\right)\\ &\cdot\exp\left(-\frac{\left((t-t_1)-a_2\right)^2}{a_3}\right),\end{aligned} \tag{4.4}$$

with

$$\gamma_h(t) = a_1\sin\left(w_2(t-t_1)\right) \tag{4.5}$$

and $t\in[t_1, t_1+T_{\mathrm{HR}})$; here, t_1 is the initial instant of the considered heart beat, $\{a_k;\ k=1,\,2,\,3\}$ and $\{w_l;\ l=1 \text{ and } 2\}$ are tunable and fixed parameters, respectively. Note that the function $\gamma_h(t)$ is periodic, but can be easily modified to account for *heart rate variability* (HRV), i.e. for the changes in the time interval between consecutive beats[1].

The minimum and maximum values of all the parameters appearing in (4.2)–(4.5) are listed in the second row and in the third row, respectively, of Table 4.1; note that, given T_{HR} and T_{BR}, w_1 and w_2 can be computed as $w_1 = 2\pi/T_{\mathrm{HR}}$ and $w_2 = 2\pi/T_{\mathrm{BR}}$, respectively, and that both parameters are expressed in rad/s.

The chest displacement resulting from the mathematical model expressed by (4.1)–(4.4), given the values of its parameters[2] listed in the fourth row of Table 4.1, is exemplified by Fig. 4.1, where the contributions due to breathing and cardiac activities are also shown for a couple of consecutive periods of breathing activity (see sub-figures b)-d)). Note that, in generating these numerical results, $t_0 = (k_b-1)T_{\mathrm{BR}}$ (with $k_b = 0, 1, ..., N_b-1$ and $N_b = \lfloor T_F/T_{\mathrm{BR}}\rfloor$), $t_1 = (k_h-1)T_{\mathrm{HR}})$ (with $k_h = 0, 1, ..., N_h-1$ and $N_h = \lfloor T_F/T_{\mathrm{HR}}\rfloor$) have been selected, where T_F is the duration of the whole observation interval; moreover, the values adopted for the model parameters have been selected on the basis of the results obtained in our experimental campaign.

Although the models illustrated above are accurate, a simpler representation of the heart and breathing profile has been adopted by various researchers [212, 222, 223]. More specifically, if T_i is assumed to be equal T_e, the approximate model

$$\Delta R(t) \approx \Delta\bar{R}(t) = \bar{\delta}_b(t) + \bar{\delta}_h(t), \tag{4.6}$$

[1] Additional details about HRV are provided in Subsection 4.8.2.

[2] These values have been selected on the basis of the data acquired in our measurement campaigns.

with

$$\bar{\delta}_b(t) \triangleq \frac{\delta_{b,M}}{2}\left[1-\cos(w_b t)\right] \tag{4.7}$$

and

$$\bar{\delta}_h(t) \triangleq \delta_{h,M}\cos(w_h t), \tag{4.8}$$

can be employed in place of that expressed by (4.1); here, $w_b = 2\pi/T_{\mathrm{BR}}$ ($w_h = 2\pi/T_{\mathrm{HR}}$) represents the HR (BR). An example of chest displacement generated according to (4.6) is provided in Fig. 4.1, where the contributions due to breathing and cardiac activities are also shown. From this figure it can be easily inferred that the results obtained on the basis of the models (4.1) and (4.6) are not so different; however, we should not forget that the model (4.6) is unable to account for the presence of both HRV and all the frequency components[3] observed in the spectrum of the received signal.

Finally, it is worth mentioning that various sensors, such as pressure belts, fiber Bragg gratings and inertial sensors, can be exploited to monitor chest surface motion (some examples of commercial wearable sensors are described in Subsection 4.6.2); however, all these sensors require to be worn by the patient under test. In a measurement campaign for radar-based monitoring, one of these sensors can be used as *reference*; this allows to separate the cardiac activity from the dominant breathing dynamics. In fact, as already mentioned above, the contribution of heart beats to surface chest motion is relatively small with respect to the that due to respiration. Moreover, the spectral components of heart motion may overlap with the respiratory harmonics; this makes separating the former contribution from the latter one really challenging.

4.4 Radar Systems: Technologies and Architectures

In this section, after providing a classification of the radar technologies employed for vital signs monitoring, some radar system architectures other than MIMO FMCW, which is already discussed in previous chapters[4], are illustrated.

4.4.1 Radar technologies and classification

Radar systems can be divided into two categories on the basis of the mechanism according to which the waveform they radiate is generated; more specifically, the first category is made of the *continuous wave* (CW) radars, whereas the second one of the so called *pulsed radars*. In a CW radar, the radiated signal is transmitted continuously, whereas, in a pulsed radar, it is sent over short periods of time. In both cases the transmitted signal can be modulated or unmodulated; for this reason, radar systems can be also classified on the basis of the modulating waveform. In the technical literature on vital sign estimation, the use of the following types of radar systems has been investigated: a) CW Doppler radar; b) *frequency-modulated continuous wave* (FMCW) radar; c) *stepped-frequency continuous wave* (SFCW) radar; d) *impulse radio ultra-wideband* (IR-UWB) radar. In the remaining part of this subsection, a brief description of each type is provided.

[3]The spectral contribution due to respiration is represented by a few relevant harmonics, as evidenced by our numerical results shown in Subsection 4.6.3.

[4]A small difference in notation is present, but the core concepts are identical.

Continuous wave Doppler radars radiate a continuous wave radio signal, characterized by a known stable frequency, and are commonly employed for their hardware simplicity. In these radar systems, the chest displacement due to heart and breathing activities results in a variation of the phase of the received signal. Such a variation is inversely proportional to the wavelength of the signal; therefore, reducing the wavelength of the transmitted wave (i.e., increasing its frequency) results in larger changes in the observed phase and, consequently, allows to detect smaller displacements. One of the main limitations of these radar systems is represented by the fact that they are unable to measure the frontal distance, i.e. the range between the radar and any subject detected by it.

Frequency-modulated continuous-wave radars and *stepped-frequency continuous-wave radars* radiate wideband frequency modulated signals. The main difference between these systems is represented by the fact that, in the former case, the frequency of the transmitted wave evolves over time in a linear manner, whereas, in the latter one, it changes in a stepwise manner. However, in both systems, the propagation delay is extracted from the phase variations observed in the received signal.

Impulse radio ultra-wideband radars radiate wideband frequency modulated signals. In these systems, the chest distance is estimated by assessing the delay experienced by sub-nanosecond pulses, being this delay proportional to the distance between the radar and any detected subject.

The above mentioned radar systems can be also classified on the basis of their wavelength or their maximum measurable range. In fact, in the first case, they are divided into *micro-wave radars*, characterized by a wavelength of few centimetres, and *mm-wave radars*, if their central frequency is equal to 77 GHz or, in general, greater than 30 GHz [224]. In the second case, instead, they can be divided in (e.g., see [50, p. 24, Table 1]): a) *short range radars*, that are able to measure a maximum range of about 30 m; b) *medium range radars*, that are characterized by a maximum range of about 100 m; c) *long range radars*, that achieve the largest maximum range (of the order 250 m).

Each of the considered radar systems is endowed with a single antenna or an antenna array at its TX and/or RX sides. A SISO radar system employs a single antenna at both its sides; note that most of the CW and IR-UWB radars considered in technical literature on the monitoring of human vital signs are of SISO type. Multiple-input multiple-output radar systems, instead, employ antenna arrays in their transmission and reception; various FMCW and SFCW radars of this type are already available on the market and their use in vital signs monitoring is currently being investigated. It is also important to keep in mind that a SISO radar can estimate the range and/or the distance of a single/multiple targets, whereas a MIMO radar makes the estimation of its/their angular coordinates possible.

Multiple-input multiple-output radar systems can be divided into *statistical* radars [9, 10], and *colocated* radars [11], [8] on the basis of the distance between their transmit and receive arrays. In practice, the transmit and receive antennas of statistical radar systems are widely separated. On the contrary, the transmit antennas of colocated radar systems are close to the receive ones and, in particular, are usually placed on the same shield. The last feature allows to develop compact devices; this explains why all the MIMO radars currently being considered for vital signs monitoring are of this type.

In a MIMO radar system, the signals radiated by distinct TX antennas are orthogonal. The simplest strategy to synthesize orthogonal waveforms is represented by *time division multiplexing* (TDM) [12]. Adopting this strategy means that distinct transmit antennas are activated over disjoint time intervals, so that the signals they radiate do not overlap in the time domain.

In selecting a radar system of a specific type, the following relevant features must be taken into considerations:

- Its *maximum detection distance* - This depends on the power radiated by the radar device, on the gain of its antennas and on the signal processing accomplished at both its TX and RX sides. If commercial radar devices operating at the same power level are considered, FMCW and IR- UWB devices are usually found to achieve a higher value of maximum detection distance than their CW and SFCW counterparts. Note also that colocated MIMO radars benefit from the availability of antenna arrays. In fact, increasing the overall number of radiating elements in their TX and RX arrays, results in a larger overall gain and, consequently, in an increase of the maximum detection distance, independently of the adopted modulation format.

- The *accuracy* it can achieve in range estimation - The highest level of accuracy is attained by MIMO radars (of FMCW, SFCW or IR-UWB types), thanks to the fact that the *signal-to-noise ratio* (SNR) level at their RX side is higher that of their SISO counterparts [11].

- Its *ability to detect multiple subjects* characterized by different DoAs - This ability is offered by MIMO radars only[5], since their RX array allows to separate signals received from different directions [12].

- Its *overall complexity* - In general, SISO radars are significantly simpler than their MIMO counterparts. In the category of SISO radar devices, the CW type is undoubtedly the one having the simplest architecture. In the category of MIMO radar devices, instead, MIMO FMCW radars usually have the largest complexity.

- Its *cost* - In general, the cheapest radar devices are the ones of CW type. For a given radar type, the cost of commercial MIMO devices is at least twice that of their SISO counterparts. The cost gap between SISO and MIMO devices mainly depends on the size of the employed antenna arrays and, in particular, increases when the overall number of the antenna elements (i.e., the overall number of RF chips) gets larger.

4.4.2 Architecture of single-input single-output radar systems

In this subsection, a brief description of the architecture of the radar systems employed for vital signs monitoring is provided. All the considered systems are equipped with single TX and RX antennas. Moreover, in illustrating their baseband processing at the

[5]In principle, a SISO radar can be used to detect the vital signs of multiple subjects, provided that their ranges are different (e.g., see [225]). However, it is unable to estimate their angular coordinates.

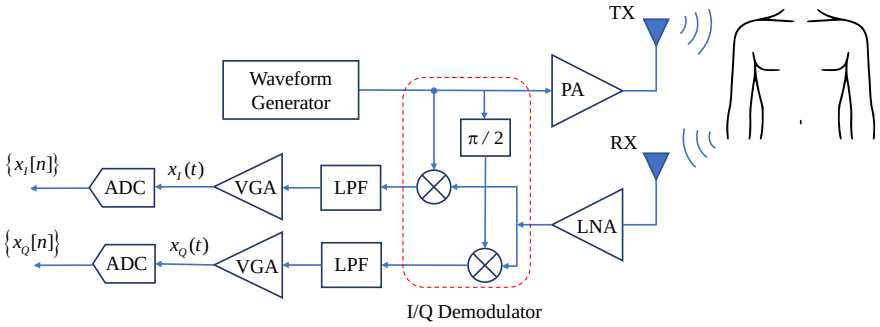


Figure 4.2: *Block diagram of a CW radar.*

receive side, it is always assumed that the chest of the monitored subject, placed in front of the radar system, can be represented as a single point target for simplicity[6].

Continuous wave radars

Let us focus first on CW radar systems. The architecture[7] of a radar system of this type is represented in Fig. 4.2. In its transmitter, the RF signal produced by a waveform generator (and characterised by the carrier frequency f_0) feeds a *power amplifier* (PA), whose response is applied to the TX antenna; this antenna is placed in front of the chest of a human being. The received signal is amplified by a *low noise amplifier* (LNA), whose response undergoes frequency downconversion to extract its in-phase and quadrature components, denoted $x_I(t)$ and $x_Q(t)$, respectively; this task is accomplished by a couple of mixers, each followed by the cascade of a *low-pass filter* (LPF) with a *variable gain amplifier* (VGA). The output of each VGA is sampled by an *analog-to-digital converter* (ADC), operating at the frequency $f_s = 1/T_s$, where T_s is the sampling period. The n-th sample of $x_I(t)$ ($x_Q(t)$) is denoted $x_I[n] \triangleq x_I(nT_s)$ ($x_Q[n] \triangleq x_Q(nT_s)$); note that the mathematical expression of these samples depends on the transmitted waveform. Further mathematical details are provided in the following for each of the three types of CW radars introduced in Subsection 4.4.1.[8]

Continuous wave Doppler radar In this case, the waveform generator appearing in Fig. 4.2 consists in a *local oscillator* (LO) generating a tone at the frequency f_c. The n-th sample of $x_I(t)$ and $x_Q(t)$ can be expressed as (e.g., see [232, Sect. II, eqs. (1)-(2)])

$$x_I[n] = a\cos(\psi[n]) + w_I[n] \tag{4.9}$$

[6] As a matter of fact, the chest of the monitored subject is usually much larger than the resolution of the employed radar sensor (and, for this reason, should be represented as a *cloud* of point targets) and absorbs a significant fraction of the incident power at its skin surface [226]. All this is usually neglected in the technical literature.

[7] Note that a different architecture is adopted by CW *self-injection-locked* radars and CW radars employing *super-regenerative oscillators*; further details about this topic can be found in [227–229], and [230, 231], respectively

[8] Additional details about FMCW radars are not provided here, since the main features of these systems have been illustrated in the previous chapters.

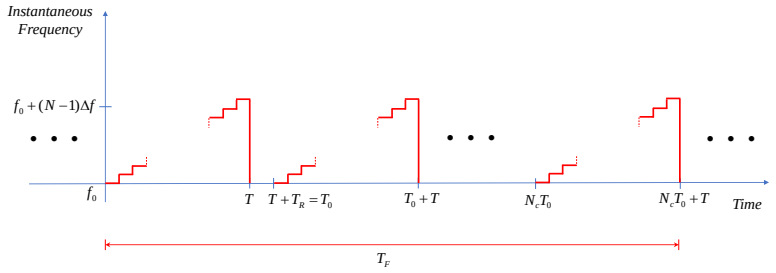


Figure 4.3: *Representation of the instantaneous frequency of the RF signal transmitted by an SFCW radar system.*

and

$$x_Q[n] = a\sin(\psi[n]) + w_Q[n], \tag{4.10}$$

respectively, with $n = 0, 1, ..., N-1$; here, n is the *fast time* index, N represents the overall number of samples acquired in the considered observation interval, a represents the amplitude of the useful signal component, $w_I[n]$ ($w_Q[n]$) is the contribution of the *additive white Gaussian noise* (AWGN) affecting the in-phase (quadrature) component,

$$\psi[n] \triangleq \psi_0 + \Delta\psi[n], \tag{4.11}$$

$$\psi_0 = 4\pi\frac{R_0}{\lambda} \tag{4.12}$$

is a constant phase shift[9] due to the (fixed) distance R_0 between the chest of the considered subject and the radar, and

$$\Delta\psi[n] = \frac{4\pi}{\lambda}\Delta R[n] \tag{4.13}$$

is the phase variation due to the chest displacement $\Delta R(t)$ (see (4.1)); here, $\Delta R[n] \triangleq \Delta R(t = nT_s)$, $\lambda = c/f_0$ is the wavelength of the radiated signal, c is the speed of light and T_s is the sampling period. From the mathematical results illustrated above it can be easily inferred that the chest displacement can be assessed by estimating the phase variations over consecutive samples of the complex sequence $\{x[n]; n = 0, 1, ..., N-1\}$, where

$$x[n] \triangleq x_I[n] + jx_Q[n] = a\exp(j\psi[n]) + w[n], \tag{4.14}$$

and

$$w[n] \triangleq w_I[n] + jw_Q[n] \tag{4.15}$$

is the noise contribution to $x[n]$ (4.14).

[9]The phase shift ψ_0 is called *direct current* (DC) offset. In vital sign monitoring, this quantity depends on the distance between the employed radar and the chest wall in front of it; however, this term may be influenced by other factors, such as the reflections from stationary targets or from other parts of the human body, and the noise of electronic components.

Stepped frequency continuous wave radar The transmitter of an SFCW radar is similar to that of an FMCW radar, the only difference being represented by the fact that the ramp generator of the last system is replaced by a staircase waveform generator. Therefore, the instantaneous frequency of the signal generated by the VCO employed in an SFCW radar changes in a step-wise manner within each radiated frequency sweep. The time evolution of the instantaneous frequency of the signal generated by the VCO over a single frame is shown in Fig. 4.3. In this figure, T_0, T and T_R represent the *frequency sweep duration*, the *sampling time* and the *reset time*, respectively, whereas N and Δf represents the overall number and the width of each frequency step, respectively. Note that, if N_c denotes the overall number of frequency sweeps forming a single *frame*, each frame lasts $T_F = N_c T_0$ s.

If we assume that the sampling interval T_s is equal to the duration of each frequency step (i.e., that the *sampling frequency* $f_s \triangleq 1/T_s$ is equal to Δf), a single complex sample is acquired at the RX side within each single frequency step. Moreover, in this case, the k-th sample of the in-phase and quadrature components available in the n-th frequency sweep interval can be expressed as (e.g., see [233, Sect. II.B, eq. (13)])

$$x_I[k,n] = a\cos(2\pi k\Delta f\tau_n + \psi[n]) + w_I[k,n] \tag{4.16}$$

and

$$x_Q[k,n] = -a\sin(2\pi k\Delta f\tau_n + \psi[n]) + w_Q[k,n], \tag{4.17}$$

respectively;

In the considered radar system, the complex sequence $\{x[k,n];\, k = 0,\, 1,\, ...,\, N-1\}$, where

$$\begin{aligned} x\,[k,n] &\triangleq x_I\,[k,n] + jx_Q[k,n] \\ &= a\,\exp\left(-j\left(2\pi k\Delta f\tau_n + \psi\,[n]\right)\right) + w\,[k,n] \end{aligned} \tag{4.18}$$

for any n, is processed to generate an estimate of the *normalised delay*

$$F_n \triangleq \Delta f\tau_n. \tag{4.19}$$

Since the last quantity can be also interpreted as the normalised frequency of a complex exponential sequence, the frequency estimation algorithms developed for FMCW radar systems can be also employed in SFCW systems. Finally, it is worth stressing that this similarity can be considered as a form of *time-frequency duality* [234]; from this viewpoint, an SFCW radar system can be seen as the dual of an FMCW radar system, as evidenced in [235].

Impulse radio ultra-wideband radar

The architecture of an IR-UWB radar is represented in Fig. 4.4. At the TX side, a tunable Gaussian pulse generator, triggered by a square wave generator, is employed to generate the (baseband) output signal

$$s\,(t) = \sum_n p\,(t - nT_0)\,; \tag{4.20}$$

here, $p\,(t - nT_0)$ represents the n-th transmitted pulse and T_0 is the *pulse repetition interval* (PRI). Note that, if T denotes the duration of each pulse[10], the interval T_0

[10] In practice, T is on the order of hundreds of ps.

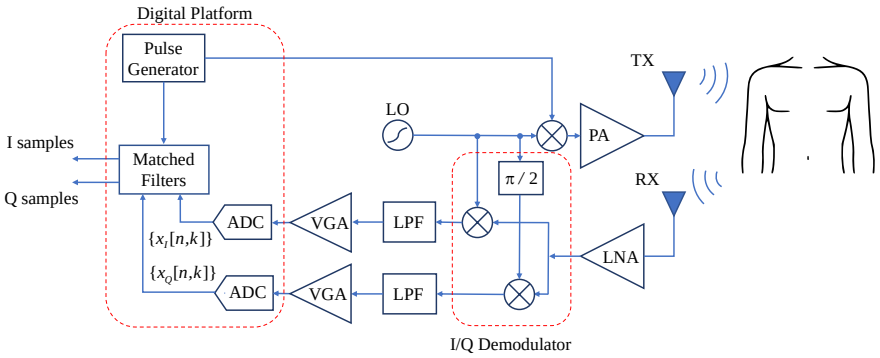


Figure 4.4: *Block diagram of an IR-UWB radar.*

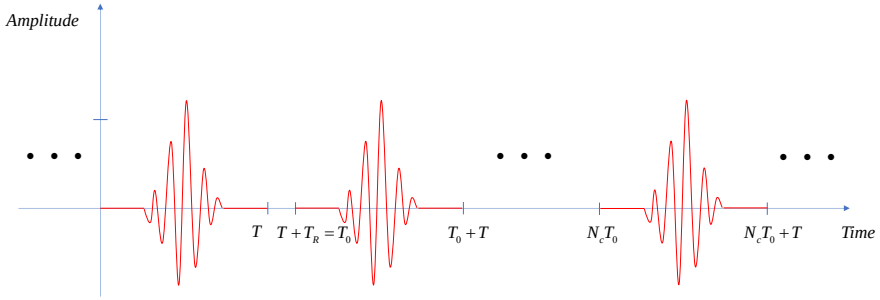


Figure 4.5: *Example of the baseband signal $s(t)$ (4.20) produced by the pulse generator of the IR-UWB radar system represented in Fig. 4.4.*

is can be expressed as $T_0 = T + T_R$, where T_R is the so called *reset time*. The time evolution of $s(t)$ is exemplified by Fig. 4.5, where N_c consecutive pulses are represented; such pulses form a *transmission frame*, whose duration is $T_F = N_c T_0$ s.

The signal $s(t)$ (4.20) undergoes frequency upconversion (accomplished by means of a mixer and a LO operating at the frequency[11] f_0) and power amplification before its transmission.

Since the pulses forming $s(t)$ (4.20) are not overlapped, we can focus our attention on the echo generated by the chest in response to the k-th pulse. At the RX side, the RF signal conveying this pulse is amplified by a LNA and downconverted[12] to extract

[11] This frequency usually belongs to the *industrial, scientific and medical* (ISM) band.
[12] The downconversion scheme is the same as that illustrated in Fig. 4.2.

Table 4.2: *Received signal model and range formulas for the considered radar systems.*

Radar type	Waveform	Received signal model	Target range
Continuous wave	Sine wave	$x[n] = a\exp(j\psi[n])$	unavailable
	Linear frequency mod.	$x[k,n] = a\exp(j(2\pi k f_n T_s + \psi[n]))$	$R_n = f_n \cdot c/(2\mu)$
	Stepped frequency mod.	$x[k,n] = a\exp(-j(2\pi k \Delta f \tau_n + \psi[n]))$	$R_n = \tau_n \cdot c/2$
Impulse-radio	Gaussian pulse	$x[k,n] = a\,p[k,n]\exp(j\psi[n])$	$R_n = \tau_n \cdot c/2$

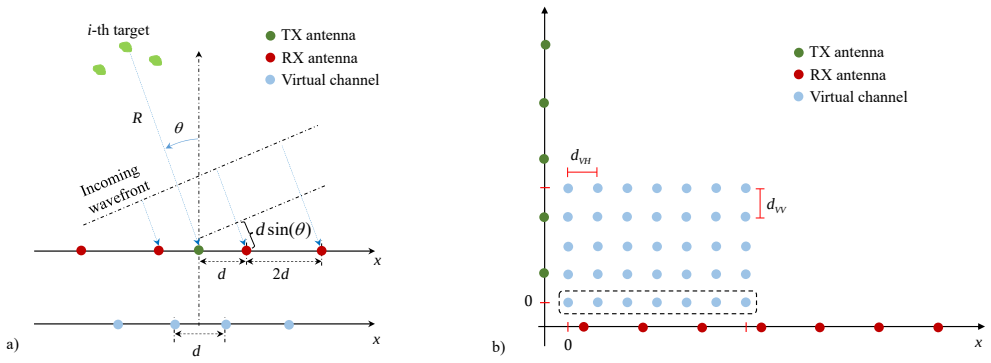


Figure 4.6: *Representation of: a) an ULA; b) an URA. In both cases the physical array and the associated virtual array are considered.*

its in-phase and quadrature components. Then, these components undergo analog-to-digital conversion at the frequency $f_s = 1/T_s$, where T_s is the sampling period. If τ_n denotes the delay experienced by the considered pulse, the k-th sample of the in-phase and quadrature components associated with it can be expressed as (e.g., see [236, Sect. II, eq. (4)])

$$x_I\,[k,n] = a\,p\,[k,n]\cos\left(\psi[n]\right) + w_I\,[k,n] \tag{4.21}$$

and

$$x_Q[k,n] = a\,p\,[k,n]\sin\left(\psi[n]]\right) + w_Q\,[k,n]\,, \tag{4.22}$$

respectively, with $k = 0,\ 1,\ ...,\ N-1$ and $n = 0,\ 1,\ ...,\ N_c - 1$; here, a represents the amplitude of the useful signal component,

$$p[k,n] \triangleq p\left(kT_s - \tau_n - nT_0\right), \tag{4.23}$$

$w_I[k,n]$ and $w_Q[k,n]$ represent noise contributions (they correspond to the samples $w_I[n]$ and $w_Q[n]$ contained in the RHS of (4.9) and (4.13), respectively), and N is the overall number of samples acquired over each PRI (briefly, of *fast time samples*). Each of the sequences $\{x_I[k,n]\}$ and $\{x_Q[k,n]\}$ is sent to a digital *matched filter* (MF), that is to a digital filter whose impulse response is the sequence $\{p\left(T - kT_s\right);\ k = 0,\ 1,\ ...,\ N-1\}$; the responses of the MFs fed by the samples of the in-phase and quadrature components are denoted $\{\hat{x}_I[k,n];\ k = 0,\ 1,\ ...,\ N-1\}$ and $\{\hat{x}_Q[k,n];\ k = 0,\ 1,\ ...,\ N-1\}$, respectively. If the chest displacement $\Delta R(t)$ is approximately constant within a single PRI, i.e.

$$\Delta R\left(kT_s + nT_0\right) \cong \Delta R\left(nT_0\right) \tag{4.24}$$

for $k = 0,\ 1,\ ...,\ N-1$ and any n, matched filtering allows to maximise the output SNR, i.e. the ratio between the energy of the useful component available at the MF output and the average power of the noise component affecting it. The target delay τ_n is estimated by identifying the position of the main peak appearing in the MF response.

Finally, it is useful to point out that:

a) Eqs. (4.21)–(4.22) can be condensed in the complex signal model

$$\begin{aligned} x\,[k,n] &\triangleq x_I\,[k,n] + jx_Q[k,n] \\ &= a\,p\,[k,n]\exp\left(j\psi[n]\right) + w\,[k,n]\,, \end{aligned} \tag{4.25}$$

where $w[k,n] \triangleq w_I[k,n] + jw_Q[k,n]$ is the noise component.

b) In IR-UWB radar systems, the presence of *clutter*, i.e. of reflections from other stationary targets, may affect the estimation of vital signs. The presence of clutter in the received signal can be modelled as DC offset or small amplitude variations in the slow time dimension in (4.21) and (4.22).

The received signal models and range formulas provided above for the different types of radar systems are summarized in Table 4.2.

4.5 Signal Processing Algorithms for Vital Signs Monitoring

In this section, we first describe the most important deterministic and *learning-based* (LB) processing methods that can be employed to extract vital signs from the measurements provided by the SISO radar systems described in the previous section. Then, we provide essential information about the processing accomplished in colocated FMCW and SFCW MIMO radars for estimating the vital signs of multiple people. Finally, we illustrate some numerical results generated by applying some of the considered methods to a synthetically generated dataset.

4.5.1 Deterministic detection and estimation algorithms for single-input single-output radars

The majority of the radar-based methods for vital signs monitoring appeared in the technical literature have the following features: a) they are *deterministic* (i.e., model-based), since their derivation is based on our prior knowledge about the structure of radar echoes; b) they extract vital signs from the phase of the received signal. As far as the last point is concerned, it is useful to focus first on a SISO CW Doppler radar for simplicity and to reconsider the phase expression (see (4.11))

$$\psi[n] \triangleq \psi_0 + \Delta\psi[n] \tag{4.26}$$

provided in Paragraph 4.4.2 for the n-th received signal sample (where n denotes the *fast time* index). As already illustrated in that paragraph, the constant ψ_0 appearing in the last equation represents the so called DC offset, whereas (see (4.13))

$$\Delta\psi[n] = \frac{4\pi}{\lambda}\Delta R[n] \tag{4.27}$$

is a time-varying term related to the body movement induced by breath and cardiac activities and, consequently, conveys the information we are interested in. This explains why the first step accomplished by the deterministic methods developed for CW Doppler radars consists in extracting the phase of the received signal sequence $\{x[n]\}$, as shown in Fig. 4.7, where a block diagram is represented to describe the overall processing they accomplish. As shown in this figure, phase extraction is carried out by the first block, that generates the N_r-dimensional vector

$$\hat{\boldsymbol{\psi}} \triangleq [\hat{\psi}[0], \hat{\psi}[1], ..., \hat{\psi}[N_r - 1]]^T \tag{4.28}$$

on the basis of the sample sequence $\{x[n]\}$ (whose elements are collected in the vector $\mathbf{x}$ appearing in Fig. 4.7); here, N_r is the overall number of measurements and $\hat{\psi}[n]$ represents an estimate of $\psi[n]$ (4.26), i.e. of the phase of the complex sample $x[n]$ (which is expressed by (4.14) in the case of a single point target). Given the vector $\hat{\boldsymbol{\psi}}$ (4.28), estimates of the BR and HR can be evaluated by applying the so called *periodogram method* [94], i.e. by identifying the dominant frequency components in the amplitude spectrum of $\{\hat{\psi}[n]\}$. In fact, it is known that: a) the highest peak in the above mentioned spectrum is found at the breath frequency in normal respiration conditions; b) the HR is higher than the BR (at least more than two times higher). For these reasons, an estimate $\hat{f}_b$ of the BR f_b (expressed in acts/s) can be computed as

$$\hat{f}_b = \hat{b}\, f_r, \tag{4.29}$$

where

$$\hat{b} = \arg \max_{\tilde{b} \in \{0,1,\ldots,N_0/2\}} \left| Y_{\tilde{b}} \right|, \tag{4.30}$$

$$Y_{\tilde{b}} \triangleq \frac{1}{N_r} \sum_{n=0}^{N_r-1} \hat{\psi}[n] \exp\left(-j2\pi n\, \tilde{b}\, / N_0\right), \tag{4.31}$$

$$N_0 \triangleq M\, N_r, \tag{4.32}$$

M is the *oversampling factor* and

$$f_r \triangleq \frac{1}{N_0 T_s} \tag{4.33}$$

for the CW Doppler radar system described in Subsection 4.4.2. Note that: a) $Y_{\tilde{b}}$ (4.31) represents the $\tilde{b}$-th element of an order N_0 *Discrete Fourier Transform* (DFT) of $\hat{\boldsymbol{\psi}}$ (4.28) and can be efficiently computed by adopting a *Fast Fourier Transform* (FFT) algorithm of the same order, as shown in Fig. 4.7; b) the strategy expressed by (4.30) aims at identifying the dominant spectral component in the spectrum of the sequence $\{\hat{\psi}[n]\}$; c) for a given M, the adoption of a larger N_r (i.e., of a longer observation time) allows to achieve a better spectral resolution.

A similar procedure can be employed for estimating the HR f_h. However, in this case, FFT processing is preceded by *bandpass filtering* (BPF) to cancel all the spectral components whose frequency falls outside the interval in which the heart frequency is expected (see Fig. 4.7). Then, an estimate $\hat{f}_h$ of f_h can be evaluated as

$$\hat{f}_h = \hat{h}\, \bar{f}_r, \tag{4.34}$$

where

$$\bar{f}_r \triangleq \frac{1}{\bar{N}_0 T_s}, \tag{4.35}$$

$$\hat{h} = \arg \max_{\tilde{h} \in \left\{0,1,\ldots,\bar{N}_0/2\right\}} \left| Z_{\tilde{h}} \right|, \tag{4.36}$$

$Z_{\tilde{h}}$ (4.31) represents the $\tilde{h}$-th element of an order

$$\bar{N}_0 \triangleq \bar{M}\, N_r, \tag{4.37}$$

DFT of the BPF output vector (whose size is equal to N_r) and $\bar{M}$ denotes the adopted oversampling factor. It is worth pointing out that:

a) Since some prior knowledge about the minimum and maximum BRs to be detected is usually available, the search interval in the RHS of (4.30) can be restricted to reduce the overall computational cost of the search for the maximum over the set $\{|Y_{\tilde{b}}|\}$. Similar considerations hold for (4.36), since $\{|Z_{\tilde{h}}|\}$ takes on significant values in a restricted frequency range because of the employed bandpass filtering.

b) The topology and order of the bandpass filter employed in HR estimation depend on the required selectivity; for instance, a fourth order digital bi-quad filter and a fifth-order low pass Butterworth filter have been adopted in [237] and [238], respectively. The main problem in filter design is represented by the selection of its passband, since the spectral components due to BR (HR) should not fall inside (out of) it. In general, the lower limit f_L and upper limit f_U of the filter passband should be selected in a way that $f_L \geq 2/T_{\mathrm{BR}}$ and $f_U \leq 3/T_{\mathrm{HR}}$, where T_{BR} (T_{HR}) is the breath (heart) period (see Subsection 4.3.1).

c) Generally speaking, the evaluation of an order N_0 ($\bar{N}_0$) FFT in BR (HR) estimation leads to partitioning the frequency interval in N_0 ($\bar{N}_0$) *frequency bins*, all having the same size. The quantity $\hat{b}$ (4.30) ($\hat{h}$ (4.36)) represents an estimate of the index of frequency bin inside which the fundamental frequency of respiration (heart) signal falls, whereas f_r (4.33) ($\bar{f}_r$ (4.35)) is the *bin size*.

d) Since the phase vector $\hat{\boldsymbol{\psi}}$ is real, the FFT output vectors $\mathbf{Y}$ and $\mathbf{Z}$ are Hermitian symmetric; this explains why only a portion of their elements is involved in the search required by (4.30) and (4.36), respectively.

Let us focus now on the extraction of phase information in CW Doppler radars [239–241]. This is usually accomplished by means of the *arctangent demodulation* (AD) technique. This means that the n-th element of the vector $\hat{\boldsymbol{\psi}}$ (4.28) is evaluated as[13]

$$\hat{\psi}[n] = \arctan\left(\frac{x_Q[n]}{x_I[n]}\right), \tag{4.38}$$

for $n = 0, 1, ..., N_r - 1$. If $x_I[n]$ or $x_Q[n]$ are equal to zero, the last equation is replaced by

$$\hat{\psi}[n] = \arcsin\left(x_Q[n]\right) \tag{4.39}$$

and

$$\hat{\psi}[n] = \arccos\left(x_I[n]\right) \tag{4.40}$$

respectively. It is important to point out that:

a) The contribution of the DC offset and of the possible strong low frequency components to the received signal phase can be mitigated by exploiting the method illustrated in [242] and based on *polynomial fitting*. This method allows to capture the slow temporal variations of that phase and to subtract them from it. Other techniques for DC offset removal have been proposed in [243] and [244], and are based on a *least-square* (LS) approach and Kalman filtering.

[13]The arctan($\cdot$) operator can be also replaced by the arctan2($\cdot$) operator in order to extend the codomain from $(-\pi/2, \pi/2)$ to $(-\pi, \pi)$.

b) The estimated phase sequence $\{\hat{\psi}[n]\}$ may exhibit some discontinuities. In fact, a discontinuity appears in the extracted phase whenever the condition[14] (see eq. (4.27))

$$\Delta\hat{\psi}_{n+1,n} \triangleq |\hat{\psi}[n+1] - \hat{\psi}[n]| = \frac{4\pi}{\lambda}\Delta R_{n+1,n} \geq 2\pi \tag{4.41}$$

is met; here, $\Delta R_{n+1,n} \triangleq |\hat{R}[n+1] - \hat{R}[n]|$. This means that, whenever the range variation $\Delta R_{n+1,n}$ observed over two consecutive sampling epochs exceeds half a wavelength, the value of the estimated phase becomes *ambiguous*. Note that this problem becomes more relevant as the frequency f_0 of the radiated signal increases (for instance, if $f_0 = 77$ GHz, any displacement exceeding 2 mm produces a phase ambiguity).

c) The phase sequence $\{\hat{\psi}[n]\}$ always undergoes a transformation known as *unwrapping*; unwrapping aims at ensuring that the variation between two consecutive elements of this sequence does not exceed π. In practice, this result is achieved by adding a multiple of 2π to some of the elements of the sequence $\{\hat{\psi}[n]\}$. However, the use of unwrapping may introduce errors in the presence of abrupt phase variations. In this case, the extended *differentiate and cross-multiply* (DACM) algorithm proposed in [245] or its modified version [246] should be employed to achieve precise phase unwrapping.

Finally, it is worth mentioning that an alternative to the arctangent method is represented by the *complex signal demodulation* (CSD) technique. This technique is based on the idea that the received signal can be seen as a frequency modulated waveform (see (4.6) and (4.27)). Therefore, an approximate model, based on the first order Bessel functions, can be derived for it [247]. This allows to separate the contribution due to the periodic movement of the chest from the one associated with the position of the body (i.e., with the term ψ_0; see (4.12)). In particular, Fourier analysis can be directly applied to the sequence $\{x[n]\}$ (see (4.14)) to estimate both BR and HR.

The architecture of the deterministic methods developed for FMCW, SFWC and IR-UWB radars can be also represented through the block diagram in Fig. 4.7. However, in these cases, the following changes are made:

a) The parameter T_s appearing in the RHSs of (4.33) and (4.35) is replaced by the PRI T_p in the case of IR-UWB radar system and by the chirp (i.e., by the frequency sweep) duration T_0 in the case of an FMCW (SFCW) radar system.

b) The algorithm for extracting phase estimation (i.e., for generating the vector $\hat{\boldsymbol{\psi}}$ (4.28)) is more complicated.

As far as the last point is concerned, further details are provided in the following two sub-subsections.

[14] In the derivation of the following result, the arctan2$(\cdot)$ operator is assumed in place of the simpler arctan$(\cdot)$.

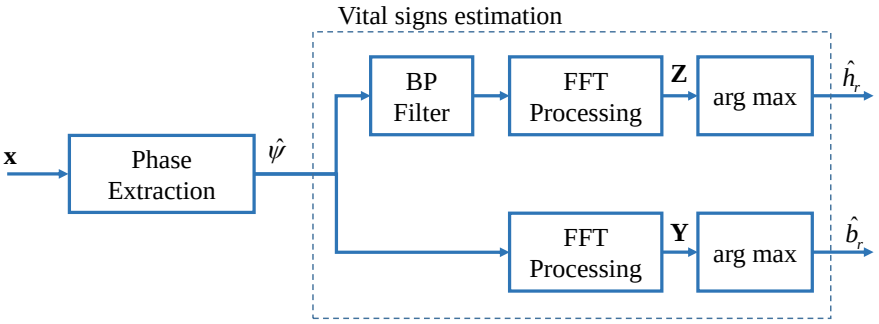


Figure 4.7: *Representation of the signal processing chain characterising various deterministic methods for vital sign estimation.*

Frequency modulated and stepped frequency continuous wave radars

Let us concentrate first on the simplest method that can be adopted in an FMCW radar system for the extraction of the above mentioned phase information. This method processes the received signal samples acquired over a single transmission frame (made of N_c consecutive chirps; see Subsection 4.4.2) and consists of the following steps:

1. The $N_0' \times N_c$ complex matrix $\mathbf{X} = [X[l,n]]$ is computed; here,

$$X[l,n] \triangleq \frac{1}{N} \sum_{k=0}^{N-1} x[k,n] \exp\left(-j2\pi nl/N_0'\right), \tag{4.42}$$

   with $l = 0, 1, ..., N_0' - 1$ and $n = 0, 1, ..., N_c - 1$. Note that $X[l,n]$ represents the l-th coefficient of an order N_0' DFT of the signal samples acquired over the n-th chirp of the considered transmission frame.

2. The index of the frequency bin

$$\hat{l} \triangleq \arg \max_{\tilde{l} \in \{0,1,...,N_0'-1\}} \left| X[\tilde{l},n] \right|, \tag{4.43}$$

   associated with the target (i.e., with the chest of the subject under test) is identified.

3. Phase extraction is accomplished through the AD method (followed by phase unwrapping), but does not involve the time domain samples of the received signal (i.e., the samples $\{x[k,n]\}$). In fact, the samples $x_I[n]$ and $x_Q[n]$ are replaced by $\Re\{X[\hat{l},n]\}$ and $\Im\{X[\hat{l},n]\}$, respectively, in (4.38)–(4.40) (here, $\Re\{x\}$ and $\Im\{x\}$ denote the real part and the imaginary part of x, respectively).

It is important to point out that:

a) The procedure described above generates, as a by product, an estimate of the range of the target, i.e. of the distance between the radar and the chest wall of the subject under test. In fact, given $\hat{l}$ (4.43), such an an estimate is given by

$$\hat{R}_0 = \hat{l} K_m \tag{4.44}$$

where, in the case of a FMCW radar system,

$$K_m \triangleq \frac{c}{2\mu N_0' T_s} \tag{4.45}$$

is the *bin-to-meters* conversion factor. Based on (1.4) and (1.5) (with $N = N_r$), the last formula can be easily put in the form

$$K_m \triangleq \frac{cN_r}{2N_0' B}, \tag{4.46}$$

where B is the radar bandwidth. The last formula and (4.44) lead to the conclusion that a larger bandwidth leads to a better granularity in range estimation, i.e. results in a better range resolution (see also Subsection 4.6.1).

b) Some methods, similar to the ones described above for CW Doppler radars, are available for the compensation of a DC offset or of large movements; in fact, they are based on filtering techniques or on polynomial fitting [206, 242, 248].

c) If the distance between the chest wall and the radar is known with a certain accuracy (e.g., when the radar is employed for monitoring the vital signs of a patient in a bed), a procedure for reducing the *region of interest* (ROI) can be implemented to simplify the estimation of $\hat{l}$. In particular, $\hat{l}$ can be searched in the ordered set $\{l_m, l_m + 1, ..., l_M\}$, made of $l_M - l_m + 1$ elements; here, $l_m = \lfloor R_m/K_m \rfloor$, $l_M = \lceil R_M/K_m \rceil$ and R_m (R_M) is the minimum (maximum) expected range.

Impulse radio ultra-wideband radars

The AD method described in Subsection 4.5.1 can be also employed for phase extraction in IR-UWB radar systems. In this case, however, the samples $x_I[n]$ and $x_Q[n]$ appearing in the RHS of (4.38) are replaced by $x_I[\hat{k}, n]$ and $x_Q[\hat{k}, n]$ (see (4.21) and (4.22)), respectively, with $n = 0, 1, ..., N_c - 1$; here, $\hat{k}$ denotes the value of the fast time index k corresponding to the main peak detected at the MF response[15]. Note also that phase estimation is followed by phase unwrapping and that, given $\hat{k}$, an estimate $\hat{\tau}$ of the delay characterizing the detected target can easily computed; this allows to evaluate the estimate $\hat{R}_0 = \hat{\tau}\, c/2$ of the distance R_0 between the radar and the chest of subject under test (see (4.12)).

Finally, it is worth mentioning that:

a) The clutter contribution to the samples $\{x[k, n]\}$ can be removed by applying an high-pass filter to this sequence. However, if the clutter produces small fluctuations along the slow-time axis, algorithms based on polynomial fitting can be employed for its removal [248].

b) The SNR of the received signal can be also improved by resorting to *noise reduction* techniques. An example of such techniques is provided in [249, Sect.

[15] It is assumed that $\hat{k}$ does not change in the considered frame interval for simplicity.

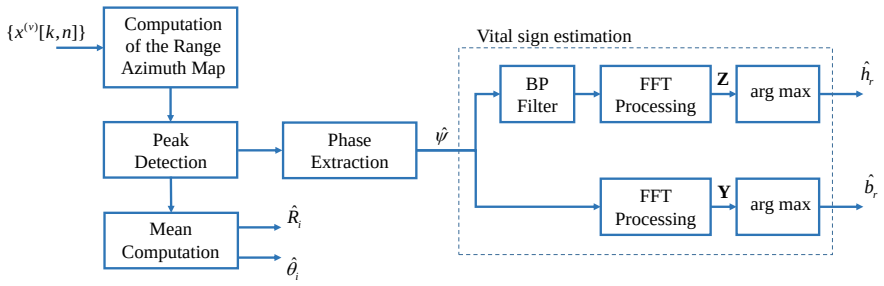


Figure 4.8: *Representation of the signal processing described in Subsection 4.5.2 and employed for estimating the position (range and azimuth) and the vital signs (HR and BR) of multiple people by means of a colocated MIMO radar.*

3.2], where the *empirical mode decomposition* (EMD) is exploited for noise mitigation. In this case, the sequence $\{x[k,n]\}$ is decomposed into a superposition of intrinsic sub-signals, defined at precise instantaneous frequencies and called *intrinsic mode functions* (IMFs). This method allows to retain only the most important frequency components of the input signal and to filter out the oscillations associated with noise components.

Some alternatives to the FFT-based estimation method described above are also available in the technical literature. These are based on the evaluation of a *fourth order cumulant* [250] or of a *continuous wavelet transform* (CWT); note that the use of the CWT in place of the FFT allows to analyse how the frequency components of the received signal phase evolve over time [249].

4.5.2 Estimation of vital signs of multiple subjects through multiple-input multiple-output radars

In the last few years, increasing attention is being paid to MIMO radar systems, mainly because they make range and DOA estimation of multiple targets possible. The use of MIMO FMCW, IR-UWB and CW radar systems for the monitoring of vital signs has been investigated in [237, 251–254], [255, 256] and [257] respectively. More specifically, the use of MIMO FMCW radars operating at 77 GHz and at 120 GHz has been studied in [237, 251, 252] and [254], respectively, whereas the simultaneous use of two FMCW radars, each endowed with single ULA, but one operating at 24 GHz and the other one at 77 GHz, has been investigated in [253]. Reference [257], instead, focuses on the use of beamforming in MIMO CW radar systems.

Let us illustrate now some key concepts that are useful to understand the processing accomplished at the RX side of a MIMO radar system. To begin, we focus on a FMCW MIMO radar equipped with N_T TX and N_R RX antennas (and, consequently, with $N_V = N_T\, N_R$ VAs; see (1.6) and (1.7)). Under the assumptions illustrated in Subsection 1.3, the signal samples acquired in a single frame through the N_V VAs can be collected in $N \times N_V \times N_c$ matrix $\mathbf{x} \triangleq \{x^{(v)}[k,n]\}$, where

$$x^{(v)}[k,n] \triangleq \sum_{i=0}^{L-1} A_{i,n}^{(v)} \exp\left(j2\pi k\, F_{i,n}^{(v)}\right) + w^{(v)}[k,n], \tag{4.47}$$

with $k = 0, 1, ..., N-1$, $n = 0,\ 1,\ ...,\ N_c - 1$ and $v = 0,\ 1,\ ...,\ N_V - 1$; here, L is the overall number of targets, and $A_{i,n}^{(v)}$ ($F_{i,n}^{(v)} = f_{i,n}^{(v)} T_s$) is the complex amplitude (the normalised version of the frequency $f_{i,n}^{(v)}$) characterising the i-th target observed on the v-th VA in the n-th chirp interval. A processing method that can be adopted in this case to estimate the range, the DOA and the vital signs of multiple people in the considered radar system is described by the block diagram represented in Fig. 4.8. Such a method consists in the four steps listed below.

1. *Computation of the range-azimuth map* - A size $N_0' \times N_A$ 2D-FFT of the $N \times N_V$ input matrix $\mathbf{x}[n] = [x^{(v)}[k, n]]$ (with $k = 0, 1, ..., N-1$ and $v = 0, 1, ..., N_V - 1$) is computed in the n-th chirp interval (with $n = 0, 1, ..., N_c - 1$); this produces the $N_0' \times N_A$ matrix

$$\mathbf{X}_n = \left[X^{(n)}[l, m] \right], \tag{4.48}$$

with $l = 0, 1, ..., N_0' - 1$ (see Subsection 4.5.1) and $m = -N_A/2,\ -N_A/2 + 1,\ ...,\ N_A/2 - 1$. Then, the n-th *range-azimuth map* is evaluated as

$$\mathbf{J}\,[n] \triangleq \left[\left| X^{(n)}\,[l, m] \right| \right] \tag{4.49}$$

with $n = 0, 1, ..., N_c - 1$. Note that $N_0' = NM$ ($N_A = N_V M_V$), where M (M_V) denotes the oversampling factor adopted in the fast time (VA) domain.

It is worth mentioning that, in the technical literature on radar systems (e.g., see [50, Sec. "Velocity estimation"]), the first step in the signal processing chain at the RX side often consists in the computation of a *range-Doppler map*, in place of the considered range-azimuth map; this aims at jointly estimating the range and velocity of all the targets detectable in the considered propagation environment. However, in vital signs monitoring, the presence of one or more static subjects is usually assumed. For this reason, on the one hand, Doppler estimation is not meaningful; on the other hand, the localization of the monitored subject in a room or a bed is required to generate accurate estimates of his/her vital signs.

2. *Peak detection* - The peaks appearing in the n-th range-azimuth map $\mathbf{J}\,[n]$ are detected by means of a proper method (for instance, the so called *constant false alarm rate*, CFAR, method can be adopted [155]), since each of them reveals the presence of a potential target. Let us assume that, independently of n, $\hat{L}$ peaks are detected in the n-th range-azimuth map $\mathbf{J}\,[n]$ and that the position of the i-th peak (with $i = 0, 1, ..., \hat{L} - 1$) is identified by the couple ($\hat{l}_i\,[n]$, $\hat{m}_i[n]$) (with $n = 0, 1, ..., N_c - 1$). Then, the estimate

$$\hat{R}_i[n] = K_m\, \hat{l}_i[n] \tag{4.50}$$

of the range and the estimate

$$\hat{\theta}_i[n] = \arcsin(\hat{m}_i[n]\, f_r), \tag{4.51}$$

of the azimuth are evaluated for the target associated with the i-th peak; here, K_m is expressed by (4.45) and $f_r = 2/N_A$ (see eq. (1.18) and (4.33)). Given the set $\{(\hat{l}_i\,[n],\, \hat{m}_i[n]);\, i = 0, 1, ..., \hat{L} - 1\}$, the $\hat{L} \times N_c$ matrix

$$\bar{\mathbf{X}} = [\bar{X}[i, n]] \tag{4.52}$$

is generated; here,

$$\bar{X}[i, n] = X^{(n)}[\hat{l}_i\,[n],\, \hat{m}_i[n]] \tag{4.53}$$

with $i = 0, 1, ..., \hat{L} - 1$ and $n = 0, 1, .., N_c - 1$.

3. *Arctangent demodulation and phase unwrapping* - In this step, the i-th row of the matrix $\bar{\mathbf{X}}$ (4.52) undergoes AD followed by phase unwrapping (with $i = 0, 1, ..., \hat{L} - 1$); this produces the phase vector

$$\hat{\boldsymbol{\psi}}_i \triangleq [\psi_i[0], \psi_i[1], ..., \psi_i[N_c - 1]]^T . \tag{4.54}$$

4. *Breath and heart rate estimation* - The phase vector $\hat{\boldsymbol{\psi}}_i$ undergoes FFT processing (with $i = 0, 1, ..., \hat{L} - 1$); this produces the N_m-dimensional spectrum $\bar{\boldsymbol{Y}}_i$ (whose $\tilde{b}$-th element is expressed by (4.31)). Then, an estimate $\hat{b}_r[i]$ ($\hat{h}_r[i]$) of the BR (HR) associated with the i-th target is evaluated on the basis of (4.30) ((4.36)).

At the end of the procedure described above, the range estimate

$$\hat{R}_i = \frac{1}{N_c} \sum_{n=0}^{N_c-1} \hat{R}_i[n] \tag{4.55}$$

and the azimuth estimate

$$\hat{\theta}_i = \frac{1}{N_c} \sum_{n=0}^{N_c-1} \hat{\theta}_i[n] \tag{4.56}$$

can be computed for the i-th target by averaging the corresponding estimates evaluated over each of the N_c chirps forming the considered transmission frame (with $i = 0, 1, ..., \hat{L} - 1$).

Finally, it is important to point out that the procedure illustrated above for a MIMO FMCW radar system can be easily adapted to MIMO IR-UWB and SFCW radars; in the last case, FFTs must be replaced by IFFTs.

4.5.3 Numerical results

In this subsection, we show some numerical results generated by applying various estimation methods illustrated above to a set of synthetically generated data. We first focus on a SISO FMCW radar and a SISO SFCW radar, both placed in front of the chest of a single subject at the distance $d = 0.5$ m. The following assumptions have been made in generating our dataset:

a) The model described by (4.1)–(4.4) is adopted for the chest displacement of the monitored subject (modelled as a single point target for simplicity); its parameters are the same as those listed in the fourth row of Table 4.1 in Subsection 4.3.2).

b) The FMCW and SFCW radars are characterised by the following parameters: a) carrier frequency $f_0 = 77$ GHz (the corresponding wavelenght is $\lambda = c/f_0 = 3.9$ mm); b) bandwidth $B = 2$ GHz; c) number of samples $N = 256$ (in the time domain for each chirp of the FMCW radar, in the frequency domain for each frequency modulated pulse of the SFCW radar); d) chirp (pulse) repetition period $T_0 = 0.06$ s for the FMCW (SFCW) radar; e) number of chirps/frame (pulses/frame) $N_c = 1034$ for the FMCW (SFCW) radar; f) ADC sampling frequency $f_s = 9$ MHz ($f_s = 7.8$ MHz) for the FMCW (SFCW) radar. Moreover, the ramp parameters of the FMCW are: a) reset time $T_R = 0$; b) ramp time $T = N/f_s = 28\mu$s; c) chirp slope $\mu = 70.312$ MHz/μs.

c) In both systems the frame duration is $T_F = N_c T_0 \cong 60$ s; this parameter represents also the duration of our observation interval.

d) The real and imaginary part of the AWGN noise samples $\{w[k,n]\}$ appearing in the RHS of (1.11) and (4.18) have zero mean and variance $\sigma_n^2 = 0.1$.

Our synthetically generated data have been processed by means of the deterministic method described in Subsection 4.5.1: its overall architecture is represented in Fig. 4.7. The AD method followed by phase unwrapping has been used for phase estimation and the following choices have been made: a) DFT orders $N_0' = 512$ and $N_0 = \bar{N}_0 = MN_c = 4 \cdot 1034 = 4136$; b) $K_m = 31$ (bin-to-meters conversion factor; see (4.45)); c) the bandpass filter employed in HR estimation is a fourth-order Butterworth filter and the lower (upper) limit of its passband is $f_L = 50/60 \cong 0.83$ Hz ($f_U = 100/60 \cong 1.7$ Hz), so that its bandwidth is $B_{BP} = f_U - f_L \cong 0.83$ Hz.

The elements of the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) obtained for the FMCW and SFCW radar systems are represented in Fig. 4.9. The amplitude spectrum obtained at the output of the slow-time FFT (IFFT) executed for phase estimation in the FMCW (SFCW) radar system is shown in Fig. 4.10, whereas the amplitude spectrum evaluated after BPF of the phase vector is shown in Fig. 4.11. The results obtained for the considered radar systems are similar and deserve the following comments:

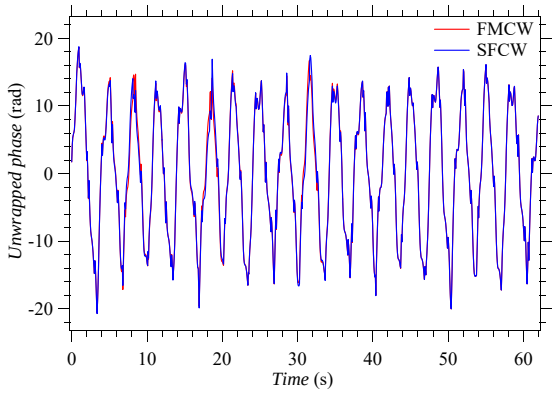


Figure 4.9: *Representation of the elements of the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) resulting from synthetically generated measurements. Both FMCW and SFCW radar systems are considered.*

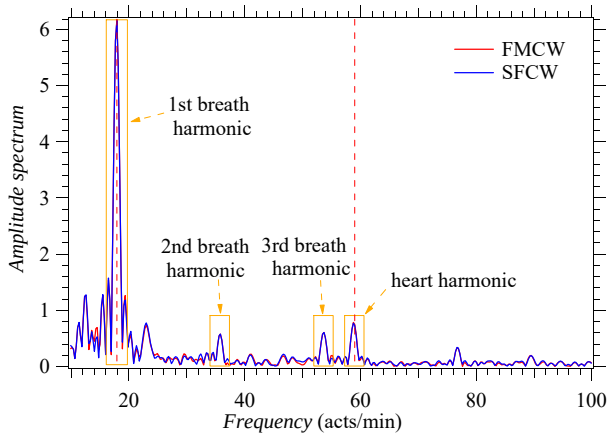


Figure 4.10: *Representation of the amplitude spectrum referring to the the phase vector $\hat{\psi}$ shown in Fig. 4.9. The dashed vertical lines identify the exact values of BR and HR. Both FMCW and SFCW radar systems are considered.*

a) The phase signal does provide information about the dynamics of chest displacement; in fact, comparing Fig. 4.9 with Fig. 4.1 leads to the conclusion that its evolution is similar to that characterising the chest displacement in the considered scenario.

b) The peaks appearing in the amplitude spectrum shown in Fig. 4.10 are associated not only with the BR and the HR, but also with the harmonics of the BR. Note, in particular, that the third harmonic falls inside the passband of the filter and that its presence may lead to a wrong estimate of the HR.

c) Based on the available spectra, the estimates $\hat{b}_r = 17.9$ acts/min and $\hat{h}_r = 58.8$ acts/min evaluated for the BR and the HR, respectively are very accurate since the exact values of these rates are $b_r = 18$ acts/min and $h_r = 60$ acts/min, respectively.

Let us focus now on a MIMO FMCW radar system placed in front of the chest of three static subjects (the i-th subject is denoted S_i, with $i = 1$, 2 and 3), whose range (R), azimuth (θ), HR (h_r) and BR (b_r) are listed in Table 4.3. In this case, the following assumptions have been made in synthetically generating our dataset:

a) The chest displacement characterising each person is modelled by (4.1)–(4.4); moreover, the values of all the model parameters are the same as those employed in Subsections 4.3.2 and 4.5.1.

b) The values of the parameters f_0, B, N, T_0, N_c, f_s, T_R, μ of the employed MIMO radar are the same as those listed for the SISO FMCW radar considered above. Moreover, the MIMO radar is endowed with a virtual ULA like the one shown in Fig. 4.6-b); the overall number of VAs forming this array is $N_V = 16$, wheras the distance between adjacent virtual elements is $d = \lambda/4$.

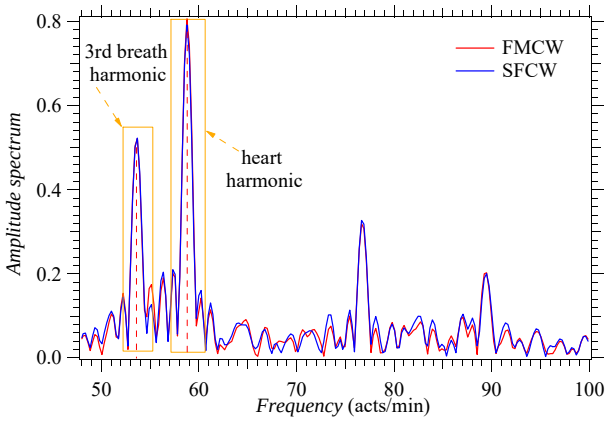


Figure 4.11: *Representation of the amplitude spectrum of the sequence generated through bandpass filtering of the phase vector $\hat{\boldsymbol{\psi}}$ shown in Fig. 4.9. The dashed vertical lines identify the exact values of BR and the third harmonic of HR. Both FMCW and SFCW radar systems are considered.*

Our data have been processed by the deterministic algorithm illustrated in Subsection 4.5.2; $N_0' = 2$ and $N_A = 4$ have been selected in the computation of the 2D-FFT output $\mathbf{X}_n$ (4.48). The range-azimuth map $\mathbf{J}[0]$ generated on the basis of the measurements acquired through the MIMO radar in the first chirp interval is represented in Fig. 4.12. The peaks detected by the CFAR algorithm are indicated by small circles, whereas black crosses are used to identify the centroids[16] of each cluster of adjacent peaks. The elements of the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) and of its amplitude spectrum in correspondence of the centroid obtained for the first (closest to the radar) target are shown in Fig. 4.13-a) and 4.13-b), respectively, whereas the estimates of the positions of the three subjects together with the estimates of their BR and HR are listed in Table 4.3. From these results the following conclusions can be easily inferred:

a) The range-azimuth map allows to detect all the subjects and estimate their position.

b) The phase vector associated with each centroid provides important information about the dynamic of the chest displacement of the subject associated with it.

c) The position and vital signs estimated for each subject are reasonably accurate.

4.5.4 Detection and estimation algorithms exploiting learning-based methods

Various deterministic algorithms may fail in complicated, highly dynamic and time varying scenarios [233]. In such cases, LB methods and, in particular, *machine learning* (ML) and *deep learning* (DL) techniques [100], may be extremely useful, since they are able to: a) learn the regularities characterizing the raw data acquired by radar systems; b) automatically extract information from them. Note that, on the one hand, ML

[16] Note that each centroid represents the position estimated for one of the subjects.

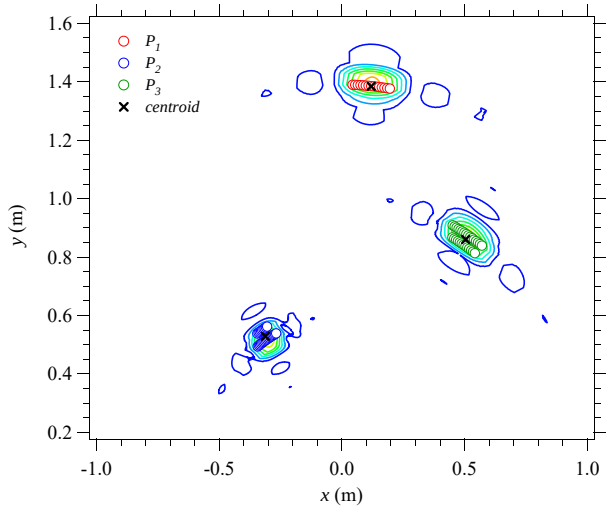
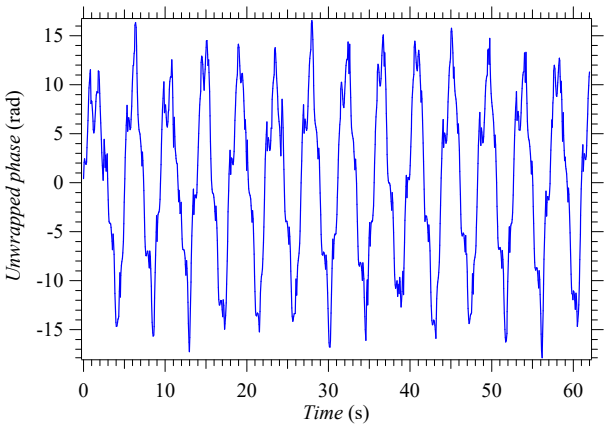


Figure 4.12: *Contour plot of the range-azimuth map generated on the basis of our synthetically generated data (a MIMO radar system is considered). The peaks P_1, P_2 and P_3 detected by the CFAR algorithm are identified by small circles, whereas black crosses indicate the position of the centroid of each cluster of peaks.*

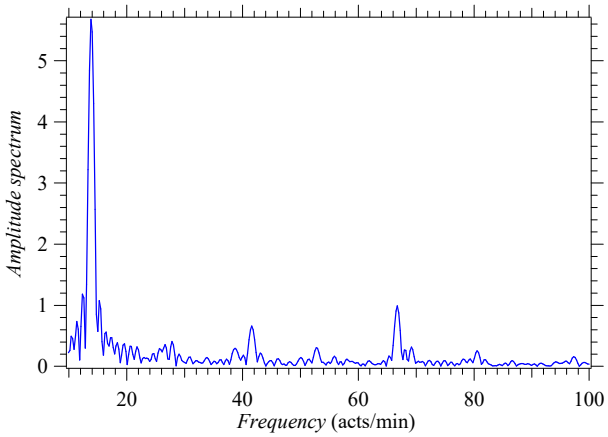
Table 4.3: *Exact values and corresponding estimates of the range (R), azimuth (θ), breath rate (b_r) and heart rate (h_r) of three distinct subjects.*

Person	R	$\hat{R}$	θ	$\hat{\theta}$	b_r	$\hat{b}_r$	h_r	$\hat{h}_r$
	(m)		(°)		acts/min		acts/min	
S_1	0.60	0.61	-30	-30.5	14.0	13.8	70	66.8
S_2	1.00	1.01	30	30.3	17	16.9	66.0	63.1
S_3	1.40	1.39	5	4.9	20.0	19.8	57.0	55.6

techniques exploit a *customized set of features*, manually extracted from the available raw data by means of different processing techniques. On the other hand, DL techniques can learn specific data patterns and extract useful information directly from the same data through the use of *neural networks*; this approach requires limited expertise on signal analysis in radar systems. Each class of methods includes *supervised* and *unsupervised* learning techniques. Generally speaking, supervised techniques, mainly used for solving classification or regression problems, are based on the idea of exploiting a *labelled dataset* for learning from them through a specific procedure, called *training*. Training aims at the identification of the model, i.e. at estimating of the *probability density function* (also called *predictive distribution*) on the basis of which the available dataset has been generated. Given the model learnt during training (i.e., the predictor), the label associated with a completely new observation or the value of a continuous variable related to it can be predicted in a way that a specific *loss* (known as *generalisation loss*) is minimised; such tasks are known as *classification* and *regression*,



(a)



(b)

Figure 4.13: *Representation of: a) the elements of a phase vector generated on the basis of the measurements acquired through a MIMO FMCW radar during a single chirp interval; b) its amplitude spectrum. The target closest to the radar is considered.*

respectively. Well known ML methods for classification are the K-*nearest neighbours* (K-NN), the *support vector machine* (SVM) and the *ensemble classifier* [75].

Unsupervised methods, instead, do not require a labelled dataset and learn some specific properties of the mechanism on which the generation of the considered dataset is based. They can be exploited to solve specific technical problems, like *data clustering*, *dimensionality reduction* and *feature selection*. In practice, data clustering aims at partitioning the available dataset in a number of groups such that data points in the same group are dissimilar from the data points belonging to all the other groups. Dimensionality reduction is employed to generate a reduced dimensionality representation of the observations, whereas feature selection consists in deriving a vector valued function that produces a useful and lower-dimensional representation of the available feature.

In general, the dataset employed to train a specific LB method contains N_t couples,

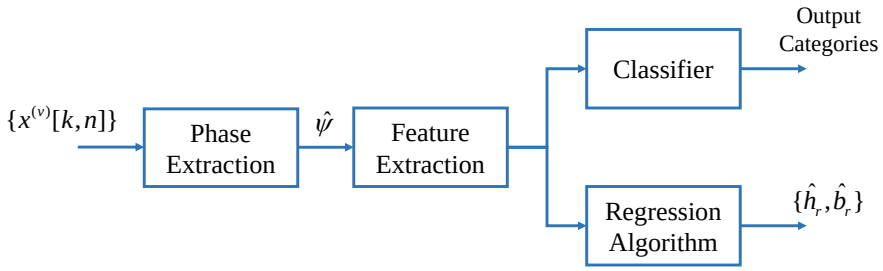


Figure 4.14: *Block diagram representing the signal processing chain of a radar-based system employing a LB method for vital signs monitoring.*

each of which consists of a D_r-dimensional real vector of *features* and a D_t-dimensional real vector of *labels*. The size N_t of the dataset can be reasonably small in the case of ML (say, between a few dozen and a few hundred), but is substantially larger in the case of DL (say, at least one thousand). In the case of radar-based monitoring of vital signs, a vector of customized features can be obtained from the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) through a procedure called *feature extraction* and can be processed for *classification* purposes. For instance, different breathing diseases could be recognised on the basis of several features, such as the breathing frequency, the chest displacement and the variability of the breathing frequency in short and long observation time. Generally speaking, for all the radar topologies, the extracted features can be grouped in three different classes, namely *time*, *short-term* and *time-frequency domain* features. Time features are represented by various characteristics of the evolution of the elements of the phase vector $\hat{\boldsymbol{\psi}}$, such as their average (or maximum) peak amplitude, the variability of their amplitude and the number of peaks. Short-time features (time-frequency domain features[17]), instead, allow to monitor how the energy (the spectral content) of the elements of the phase vector $\hat{\boldsymbol{\psi}}$ evolves over time. The above mentioned features can be employed for both classification and regression, as exemplified by Fig. 4.14, where the overall signal processing chain of a radar system employing a LB method is illustrated.

An alternative to ML methods is provided by DL methods. In the last case, feature extraction is automatically accomplished by a neural network and, if classification is required, a *softmax layer* is employed to evaluate the probability that a certain observation is associated with one of the given classes.

A limited literature on the use of ML and DL methods in radar-based monitoring of vital signs is available. The use of ML methods in CW Doppler radar systems for vital signs monitoring has been investigated in [138, 258–260], while an unsupervised method for DC offset and clutter suppression using FMCW radars has been investigated in [261]. Moreover, an HR estimation technique based on a *convolutional neural network* (CNN) has been developed in [262] for an IR-UWB radar system. In particular, supervised methods have been exploited for classifying breathing disorders, for heart sound classification and for removing the high order harmonics from the phase vector $\boldsymbol{\psi}$ (4.28) in [258], [260] and [138], respectively. An *artificial neural network* (ANN) for the reliable detection of heartbeats through a CW Doppler radar has been

[17]Different tools for time-frequency analysis, like the CWT and the *short time Fourier transform* (STFT), can be employed to extract these features.

proposed in [259]. The adopted ANN is composed by a cascade of multiple layers with different number of neurons and its architecture is quite simple; its main drawback is represented by its inability to cope with time series, i.e. to extract features related to the time evolution of the observed signal.

4.6 Some Considerations on Radar Selection and on its Use in Experimental Campaigns

In this section, we illustrate some important lessons that we have learnt from our experimental work conducted on healthy adult volunteers in the laboratories of the Department of Engineering "Enzo Ferrari" and of the Cardiology Division, Department of Biomedical, Metabolic and Neural Sciences at the University of Modena and Reggio Emilia. First, we focus on the essential requirements that radar devices employed for vital signs monitoring should meet. Then, we provide some guidelines for developing an experimental setup and illustrate some numerical results about the estimates of vital signs extracted from our experimental measurements. Finally, we comment on how to assess estimation accuracy in vital signs monitoring.

4.6.1 Fundamental requirements of radar devices

Nowadays, a number of compact radar devices, not explicitly developed for medical applications and operating in the *millimeter wave* (mmwave) spectrum, are available on the market at various prices. A radar system employed for vital signs monitoring needs to satisfy various technical requirements, that are influenced by the environment in which measurements are acquired [263]. These requirements concern: a) the maximum distance of the radar from the body of the subject under test; b) its bandwidth and its operating frequency f_0; c) its displacement resolution; d) its angular resolution. Different values of the parameters mentioned above may have a substantial impact on the achievable accuracy, as illustrated below.

The *maximum distance* at which a radar should operate depends on the considered application. If long range detection is required, an high transmission power and/or highly directive antennas should be employed to guarantee a proper SNR at the RX side [264]. The power radiated by colocated MIMO radars employed in vital signs monitoring is small; however, in this case, *beamforming* techniques can be employed to constructively combine the signals received by multiple antennas, so enhancing the overall SNR. It is also important to keep in mind that any radar system, independently of its antennas and maximum radiated power, is characterized by a *maximum unambiguous range* (denoted R_{max}), that is by a maximum distance beyond which target range is not correctly estimated. In the case of FMCW or SFCW radars, R_{max} can be expressed as[18]

$$R_{max} = N\frac{c}{4B}, \tag{4.57}$$

[18]If a FMCW radar system is considered, this result originates from the fact that the frequency f_n associated with the target (1.18) can be unambiguously estimated if its does not exceed half the sampling frequency $f_s \triangleq 1/T_s$.

where B is the bandwidth of the radiated signal and N is the overall number of samples acquired in a chirp interval (frequency sweep) by an FMCW (SFCW) radar. Then, for a given bandwidth, R_{max} can be raised by increasing N, i.e. by adopting an higher sampling rate at the receive side.

The *bandwidth* of the radiated signal plays a fundamental role in radar-based monitoring of vital signs. In fact, a larger bandwidth results in a better *range resolution*, i.e. in an improved ability to discriminate multiple targets in range as well as to generate a more detailed image of an extended target. Since the bandwidth is usually a fraction of the carrier frequency, this consideration has motivated the adoption of *operating frequencies* that fall in the mmwave spectrum, i.e. that belong to the frequency range[19] (30, 300) GHz. The selection of an higher operating frequency (i.e., of a shorter wavelength λ) has various implications. First of all, it leads to an higher *phase sensitivity*, that is to faster changes in the received signal phase observed in the presence of the small movements to be detected in vital signs monitoring (see (4.11)-(4.13)). This is certainly beneficial, provided that the *phase ambiguity* problem does not arise (i.e., that the observed phase variations remain within the range $(-\pi, \pi)$). The last problem can be circumvented by increasing the sampling rate of the received signal; this has also a beneficial effect on the maximum range at which a target can be detected, as already mentioned above. A change in the operating frequency of the radar has also a significant impact on the penetration depth of the radiated EM waves through human tissues [206]. Readers should keep in mind that: a) microwave signals are partially reflected and partially absorbed by the human skin [265]; b) the attenuation of the reflected EM field increases with its frequency; c) the penetration through human skin, instead, decreases with frequency (for instance, the skin penetration depth is 2.7 mm at 10 GHz and just 0.5 mm at 60 GHz [266, 267]); d) the EM wave reflection due to body tissues becomes stronger as the operating frequency increases; e) the quality of the echo of a mmwave radar is negligibly affected by the thickness of the clothes of the monitored subject [268]. Based on these results we can state that, when a mmwave radar device is employed in vital signs monitoring of a still subject, the phase and amplitude fluctuations observed in its received signal really originate from the chest and skin displacement stimulated by his/her cardio-pulmonary activity.

The *displacement resolution* can be defined as the minimum measurable displacement over two consecutive frames transmitted by the considered radar device (such frames are separated by a time interval lasting T_0 s; see Subsection 4.4.2). Based on (4.41), the displacement $\Delta R_{k,k+1}$ experienced by a point target between the k-th frame and the subsequent (i.e., the $(k+1)$-th) frame can be expressed as

$$\Delta R_{k,k+1} = \frac{\lambda}{4\pi} \Delta\psi_{k,k+1}, \tag{4.58}$$

where $\Delta\psi_{k,k+1}$ is the phase variation observed in the electromagnetic echo. If $\delta_{b,M}$ denotes the maximum chest displacement due to breathing (see Subsection 4.3.2), the inequality

$$\frac{\Delta R_{k,k+1}}{T_0} \geq \frac{2\delta_{b,M}}{T_{BR}} \tag{4.59}$$

[19]The corresponding wavelengths range from 1 millimeter and 10 millimeters

should be satisfied to achieve a sufficient resolution in detecting chest movements. The last inequality can be rewritten as

$$T_0 \leq \frac{T_{BR}}{2\delta_{b,M}} R_{k,k+1} \tag{4.60}$$

or, equivalently, as

$$T_0 \leq \frac{T_{BR}}{2\delta_{b,M}} \frac{\lambda}{4\pi} \Delta\hat{\psi}_{k,k+1}, \tag{4.61}$$

thanks to (4.58). If we assume that $\delta_{b,M}$ is equal to 4λ (where $\lambda \cong 4$ mm, if the employed radar system operates at 77 GHz) and keep in mind that[20] the inequality $\Delta\psi_{k,k+1} < 2\pi$ must be satisfied to avoid any phase ambiguity, from (4.61) it is easily inferred that

$$\frac{1}{T_0} \geq \frac{16}{T_{BR}}, \tag{4.62}$$

so that the selected frame rate must be substantially higher than the breathing frequency. Similar considerations can be formulated for the displacement due to heart activity; the only difference being represented by the fact that T_{BR} is replaced by T_{HR}. Then, achieving sufficient accuracy in the estimation of BR and HR requires the selection of a proper frame rate $1/T_0$. In practice, as already mentioned in Subsection 4.3.1, typical values of BR at rest range from 10 to 25 acts/min (i.e., from 0.2 and 0.4 Hz), whereas those of HR from 50 to 100 acts/min (i.e., from 1 to 1.67 Hz); under stress conditions, the BR and HR increase up to 40 acts/min (0.67 Hz) and 180 acts/min (3 Hz), respectively. Therefore, based on (4.62), the frame rate should be on the order of 25 Hz at rest and 45 Hz under stress. It is also worth mentioning that, whatever the selected frame rate, a sufficiently long observation time is required to generate reasonably accurate estimates of vital signs through spectral analysis. In practice, using a commercial MIMO radar, such estimates should be extracted from the radar measurements acquired on a time interval that covers at least three periods of breathing (and, consequently, about ten periods of heartbeat), i.e. lasting at least 10 s at rest; this unavoidably introduces some latency in vital sign estimation.

Let us focus now on the *angular resolution*. This parameter plays an important role when a MIMO radar is employed to detect the vital signs of multiple people, characterized by different angular coordinates. The achievable angular resolution is strictly related to the number of virtual channels N_V made available by the employed radar system. If this system is endowed with a single ULA characterized by N_V equally spaced channels, the angular resolution $\Delta\theta$ can be evaluated as

$$\Delta\theta = \frac{\lambda}{2d(N_V - 1)} \frac{180^\circ}{\pi}, \tag{4.63}$$

where d is the distance between two adjacent virtual elements; for instance, if $d = \lambda/4$, the resolution is equal to $2/(N_V - 1)$ rad.

In our measurement campaigns performed in an hospital environment (see Subsection 4.6.2), two colocated FMCW radars, namely the IWRxx43 TI radar [82] and

[20] This expression is valid if the arctan2(·) operator is employed, as illustrated in Subsection 4.5.1.

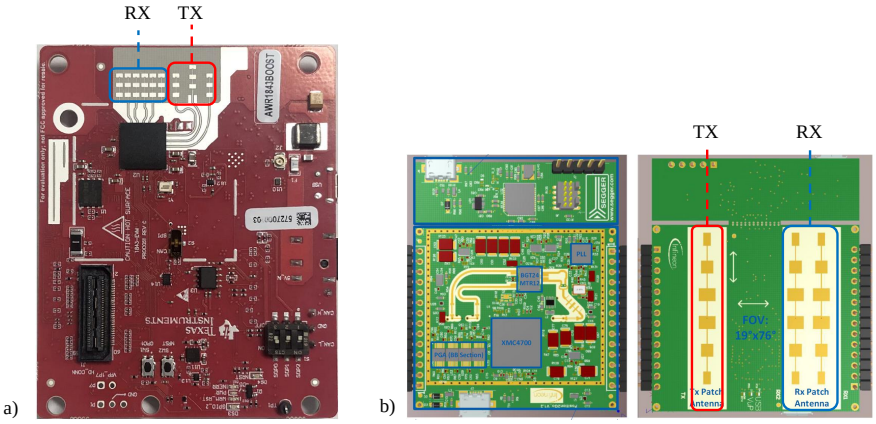


Figure 4.15: *Photos of: a) the IWRxx43 radar front end; b) both sides of the Position2Go module.*

the Position2go (P2G) FMCW radar[21] [269] have been employed. The former device is manufactured by Texas Instrument and operates at the frequency $f_0 = 77$ GHz, whereas the latter one is manufactured by Infineon and operates at the frequency $f_0 = 24$ GHz; the difference in their transmission frequencies allowed us to assess the potential advantages originating from the use of higher frequencies and/or from the exploitation of frequency diversity. The front-end of these devices is shown in Fig. 4.15. It is important to note that in our application $T_F = 60$ s, i.e., the duration of each transmitted frame is equal to that of the whole observation interval and the chirp duration $T_0 = T + T_R$ is characterised by a long reset time, since $T_R > T$.

The TI radar is endowed with an array composed by $N_T = 3$ ($N_R = 4$) TX (RX) antennas (see Fig. 4.15-a)); therefore, it makes available a virtual array consisting of $N_V = 3 \cdot 4 = 12$ (virtual) antenna elements (see (1.8)); these elements form two different horizontal ULAs composed by $N_{HULA_1} = 8$ and $N_{HULA_2} = 4$ virtual channels; therefore, only $N_{VULA} = 2$ aligned channels are available along the vertical direction. The horizontal (i.e., azimuthal) resolution and the vertical (i.e., elevation) resolution are $\Delta\theta \simeq 16.6°$ and $\Delta\phi = 45°$, respectively (see eq. (4.63)). In our work, the following values of the radar parameters have been selected: a) frequency slope $\mu = 86$ MHz/μs; b) overall number of samples per chirp $N = 256$; c) ADC sampling frequency $f_s = 9$ MHz; d) chirp duration $T_0 \approx 60$ msec. These choices entail that: 1) the bandwidth of the radiated signal is $B = \mu T \simeq 2.45$ GHz, since the ramp up time is $T = N/f_s = 18.4$ μs; 2) the maximum unambiguous distance is $R_{max} \simeq 15.67$ m (see (4.57)).

The P2G radar is equipped with an array composed by $N_T = 1$ transmit and $N_R = 2$ receive antennas (see Fig. 4.15-b)). Therefore, a virtual array of $N_V = 1 \cdot 2 = 2$ virtual elements (see (1.8)), forming an horizontal ULA, is available. In our work, the following values of the radar parameters have been selected: a) $\mu = 0.78$ MHz/ms; b) overall number of samples per chirp $N = 256$; c) ADC sampling frequency $f_s = 1$ MHz; d) chirp duration interval $T_0 \approx 60$ msec. In this case, the ramp up time is

[21] Detailed descriptions of these devices are available at `https://www.ti.com/tool/IWR1843BOOST` and at `https://www.infineon.com/cms/en/product/evaluation-boards/demo-position2go/`, respectively.

Figure 4.16: *Photos of both sides of the XM112 UWB radar.*

$T = N/f_s = 256$ μs, the bandwidth of the radiated signal is $B = 200$ MHz and the maximum unambiguous distance is $R_{max} = 192$ m. Therefore, this radar device can be employed for long-range applications, but achieves a lower range resolution than the TI IWR1843 radar.

A short measurement campaign was also conducted in an office environment (see Subsection 4.6.2). In that case, a XM112 *pulsed coherent* (PCR) UWB radar has been employed (see Fig. 4.16). This device, manufactured by *Acconeer*[22], operates at the frequency $f_0 = 60$ GHz and is of SISO type. Its PRI T_0 is equal to 76.9 psec, its transmission frame consists of $N_c = 621$ pulses and its range resolution is equal to 0.5 mm. In our acquiring our measurements from the XM112 UWB radar, a minimum (maximum) range $R_m = 50$ cm ($R_M = 80$ cm) have been assumed, since the distance of the monitored subject from the radar itself always belonged to the interval $[R_m, R_M]$.

4.6.2 Data acquisition

The typical differences between the datasets analysed in the technical literature about radar-based monitoring of vital signs concern (see Table 4.6 in Section 4.7):

a) the *overall number of subjects* involved in the data acquisition procedure;

b) the *heterogeneity of the acquired dataset* and, in particular, the *positions* of the involved subjects with respect to the employed radar (e.g., sitting, lying down, etc.) and their *physical conditions* (e.g., in a rest condition, under strain, sleeping, etc.).

The *overall number of subjects* ranges from few units to a few dozens. Readers should keep in mind that the analysis of deterministic and ML methods do not usually require a large number of measurements, whereas that of DL techniques needs large and heterogeneous datasets. For instance, the measurements analysed in [270] have been

[22]See https://www.acconeer.com/company/ for further details.

acquired from ten people (more precisely, six males and four females), whose distance from the employed radar system ranged from 20 to 80 cm. On the contrary, a much larger dataset was needed to train and test an LSTM network in [271]. In the last case, the overall acquisition time of the whole dataset is equal to 18900 s (corresponding to approximately five hours); moreover, the acquired measurements refer to 30 different subjects.

The *heterogeneity of the acquired dataset* can be improved by observing the considered subjects in different positions (e.g., in front of the radar with the chest facing towards the antennas, or on its left/right side), but always at a fixed distance from the employed radar systems, [213]. Subjects can also be placed at different distances from radar systems, but, in this case, their angular coordinates should not change (e.g., see [270, Sect. III]).

The measurements analysed in most of the technical literature refer to people breathing at rest. However, especially in last years, contactless systems based on radar technology have been employed for HR and BR monitoring of subjects in different breathing conditions. For instance, the measurements of the dataset employed in [272] have been acquired in apnea, during the Valsalva maneuver[23] and in two different positions (tilt-up and tilt-down).

The *selection of proper reference sensors* represents another important technical issue to be considered before starting a measurement campaign. In fact, the data collected from reference sensors are always required for the validation of deterministic algorithms [273], [248], or for training LB methods [262]. In the measurement campaigns described in the technical literature, various medical instruments, like electrocardiographs and wearable sensors, have been used. Wearable sensors include elastic bands with built-in electrodes [274], *micro-electro-mechanical-systems* (MEMS), pulse-oximeters or Bragg-gratings sensors [275]. Most of the commercially available wearable sensors are easy to use, since they provide excellent user interfaces and *application programming interfaces* (APIs) for Python or MATLAB programming environments. However, an important issue to be taken into consideration before selecting a specific reference sensor is the possibility of accurately synchronizing the timing of its measurements with that of the data provided by the employed radar device.

A long measurement campaign is being conducted at the Cardiology Division, Department of Biomedical, Metabolic and Neural Sciences at the University of Modena and Reggio Emilia (Hospital of Modena) on healthy adult volunteers; the experimental setup adopted in this case is shown in Fig. 4.17. The two FMCW radar sensors described in the previous subsection are mounted on a wooden bar, which, in turn, is put on a tripod. Moreover, their antenna arrays are oriented towards the chest of each subject, always lying down a bed, whose upper part is slightly tilted. The typical vertical (horizontal) distance d_h (d_v) between the radars (the tripod) and the subject under test is approximately equal to 75 cm (50 cm); our tripod allows to move up or down the radars, so that the distance between them and the monitored subject can be modified. In our measurements, different positions are being considered for the subject under test. In practice, he/she is sitting or lying down on the bed with different tilting angles; in the first case, his/her movements are very small and this makes the estima-

[23]The Valsava maneuver is performed by moderately forceful attempted exhalation against a closed airway; this can be practically implemented by expiring against a closed glottis.

Figure 4.17: *Photo of the setup employed in our measurement campaigns conducted in an hospital; both the IWRxx43 and the Position2Go radars are used.*

tion of HR and BR potentially very accurate, whereas, in the second one, the tension of his/her muscles may affect the estimation of vital sign parameters. In our setup, the reference sensor is the *Shimmer3* device manufactured by Shimmer [276]. This device, controlled by means of a simple user interface, is able to send data over a Bluetooth connection and is equipped with five electrodes (right hand side mid-axillary, left hand side mid-axillary, right leg, left leg, V1; see [276, p. 6, Fig. 3-2]), that need to be placed in specific positions of the chest of the subject under test. In our acquisitions, the reference *electrocardiogram* (ECG) signal is read from the voltage difference between the left leg and right hand side mid-axillary (LL-RA) electrodes. The breathing signal, instead, is generated by measuring the impedance between these two electrodes; in practice, these electrodes are used to inject a weak high frequency (10 kHz) alternating current in the chest tissues and the voltage variations due to chest displacement are measured. Then, an estimate of the chest impedance variations is obtained by computing the ratio between the measured voltage and the injected current; finally, breathing dynamics are inferred from such variations (further details about this method can be found in [277]).

Thanks to the availability of a proper API, Shimmer3 measurements can be easily synchronized in time with those acquired through the IWRxx43 radar or the P2G radar, so that data acquisition can be accomplished in real time. In fact, the Shimmer3 device offers the possibility of sharing the time reference (namely, the CPU timestamp) with our radar devices; this has allowed us to synchronise the reference and the probing signals with great accuracy. As already mentioned above, in our experiment, each data acquisition refers to an observation interval lasting $T_F = 60$ s. However, since the radars and the reference sensor are characterized by different frame rates, the samples of each of the two signals (namely, the ECG and breathing signals) provided by the

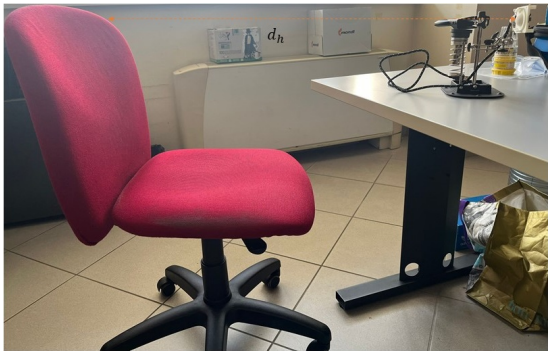


Figure 4.18: *Photo of the setup employed in our measurement campaign conducted in an university laboratory.*

Shimmer3 sensor are stored in a sequence of vectors, each having size $N_{c,\mathrm{S}} = 15317$, whereas the received signal samples provided by the IWRxx43 (P2G) radar are stored in a 3D matrix of size $N \times N_V \times N_{c,\mathrm{TI}} = 256 \times 4 \times 997$ ($N \times N_V \times N_{c,\mathrm{P2G}} = 256 \times 2 \times 933$). It is also worth mentioning that, in general, if the construction of a dataset requires that acquisitions originating from the given reference sensor and those coming from the employed radar device have the same length, a subsampling or upsampling procedure can be adopted for the reference or the radar signals.

Independently of the nature of the algorithm to be tested on the acquired measurements, it is highly recommended to build up a reasonably large dataset, since the chest dynamics of distinct subjects can exhibit very different characteristics. Moreover, in our measurement campaigns, in order to guarantee a sufficient variability in the generated dataset, the subjects under test are being observed in different conditions. In practice, the following conditions are considered in our data acquisitions: a) breathing normally, at rest; b) during inspiratory (expiratory) apnoea (for as long as he/she could, ending the acquisition with normal breathing); c) hyperventilating (for as long as he/she could); d) during the Valsalva maneuver; e) after making an effort. In the last case, the subject under test is required to move up a step and down from it for at least one minute).

As already mentioned in the previous subsection, an additional measurement campaign has been also conducted in a different environment and, in particular, in one of the research laboratories of the Department of Engineering "Enzo Ferrari". The experimental set-up adopted in this case is shown in Fig. 4.18. An XM112 IR-UWB radar has been mounted on a small tripod, which, in turn, has been positioned on an office desk in front the chest of the monitored subject (sitting on chair at an horizontal distance d_h approximately equal to 70 cm). Our measurements have been acquired from five healthy subjects; the resulting dataset consist of $\bar{N} = 45$ waveforms, each referring to an observation interval that lasts $T_{\mathrm{F}} = 60$ sec and represented by $N_{c,\mathrm{AC}} = 5926$ complex samples. The *Shimmer3* device has been employed again as a reference sensor and its output has been synchronized with that of that of the XM112 device. The data collected by the reference sensor have been stored in a vector of size $N_{c,\mathrm{S}} = 15317$, assuming an average frame rate equal to 100 samples per second.

4.6.3 Numerical results

In this subsection the estimation methods described in Subsection 4.5.1 are applied to the measurements acquired through our three radar devices described in Subsection 4.6.1. Our objective is estimating the BR and HR in the following two scenarios: 1) a single subject lying down on a bed at a distance $R_{ref} = d_v = 0.75$ cm from the radar sensor (see Fig. 4.17); 2) a single subject sitting on chair at an horizontal d_h=70 cm cm from the radar sensor (see Fig. 4.18). In our study, emphasis is put on the extraction of HR from radar measurements, since this represents a substantially more challenging task than BR estimation.

Let us focus now on the first scenario. In this case, the signal samples acquired through the four (two) VAs of the IWRxx43 (P2G) radar in the k-th chirp interval, with $k = 0, 1, ..., N_{c,\mathrm{x}} - 1$ and x = TI (P2G) are processed by a *beamforming algorithm* [8]; this allows to constructively combine the echoes impinging on the RX array along a specific direction and to produce a single N-th dimensional column vector $\mathbf{x}_k$; this vector feeds the range estimation and bin selection block shown in Fig. 4.8. The processing accomplished by the last block is based on (4.42) (a tape meter is used to compute the reference range R_{ref}). Moreover, this block produces the couple $(\hat{l}\,[k],\, \hat{\psi}[k])$, consisting of the bin index $\hat{l}\,[k]$ and the phase estimate $\hat{\psi}[k]$, for any k. It is worth mentioning that: a) that an estimate $\hat{R}[k]$ of the target *range* is evaluated as

$$\hat{R}[k] = \hat{l}\,[k]\, K_m, \tag{4.64}$$

where

$$\hat{l}\,[k] = \arg \max_{l_m \leq \bar{l} \leq l_M} |X[l, k]|^2, \tag{4.65}$$

$X[l, k]$ is defined by (4.42), $l_m = 21$, $l_M = 33$, $K_m = 32.6$ m ($l_m = 2$, $l_M = 3$, $K_m = 2.6$ m) for x = TI (x = P2G); b) $N_0 = 512$ and $R_m = 0.65$ m, $R_M = 1$ m have been selected for both radar systems. Note that $\hat{R}[k]$ (4.64) does not necessary coincide with the reference range R_{ref}, since:

a) Some errors are unavoidably introduced in the range measurement procedure. In fact, it not easy to identify exactly the point of the chest on which the beam radiated by the employed radar is focused, since the antennas of both radar devices are not highly directive.

b) The range estimate $\hat{R}[k_1]$ computed in the k_1-th chirp interval may be slightly different[24] from the estimate $\hat{R}[k_2]$ obtained in the k_2-th chirp interval, with $k_1 \neq k_2$.

The elements of the unwrapped phase vector

$$\hat{\boldsymbol{\psi}} \triangleq [\hat{\psi}[0], \hat{\psi}[1], ..., \hat{\psi}[N_{c,\mathrm{x}} - 1]]^T, \tag{4.66}$$

computed on the basis of the measurements that have been acquired through the TI (P2G) radar in a single observation interval, are represented by the red (green) line in

[24]Note that, in principle, information about chest displacement are contained in the sequence $\{\hat{R}[0], \hat{R}[1], ..., \hat{R}[N_{c,\mathrm{x}} - 1]\}$, but, in general, its elements are too noisy for a reliable detection of vital signs.

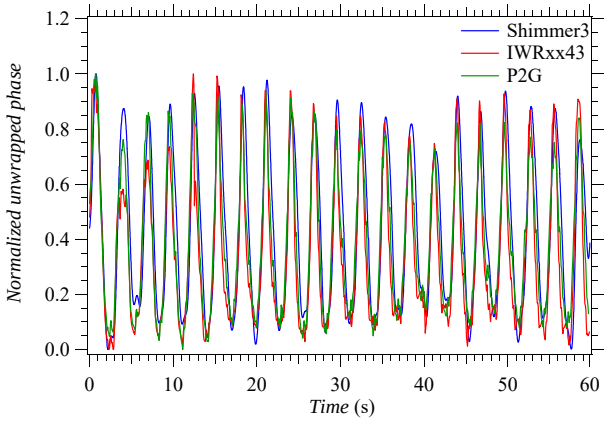


Figure 4.19: *Representation of the normalized unwrapped phase extracted from our IWRxx43 radar measurements (red line) and P2G radar measurements (green line). The breath signal acquired through our reference sensor is also shown (blue line).*

Fig. 4.19 (note that in order to ease the interpretation of these results, the elements of the two phase vectors have been normalised[25]); in this case, AD followed by DC offset removal and phase unwrapping is employed. The results illustrated in Fig. 4.19 deserve the following comments: a) the two radar signals represent the dynamics of the chest due to both breathing and cardiac activities; b) the phase trajectories computed on the basis of the measurements provided by the two radars overlap, even if these devices operate at different frequencies, and exhibit similar evolutions as the signal provided by our reference sensor (Shimmer3; see the blue line appearing in Fig. 4.19). The amplitude spectrum generated by applying an FFT of order $N_0' = N_{c,\mathrm{S}} = 15317$ to each of the signals represented in Fig. 4.19 is shown in Fig. 4.20. These results evidence that: a) the spectral peak associated with breathing (in correspondence of 21 acts/min) is much larger than the one related to heart beat (visible at approximately 61 acts/min); b) the second order harmonic of breathing is clearly visible at approximately $21 \times 2 = 42$ acts/min, whereas its third harmonic is expected at $21 \times 3 = 63$ acts/min.

The heart signal can be extracted from the unwrapped phase through band-pass filtering [212]. In this case, a fourth order band-pass Butterworth is used to select the spectral components whose frequencies belong to the interval [0.91 Hz, 3 Hz]; its output is shown, for both radars, in Fig. 4.21 (the ECG signal generated by the reference sensor is also represented). Note that, in the considered observation interval, the peak-to-peak period of the radar waveforms is comparable to the *normal to normal* (NN) peak interval characterizing the ECG signal[26]. However, the heart and the ECG signal are not perfectly aligned; this suggests that the heart frequency, i.e. the distance between two consecutive peaks, is not completely stable during the observation time. This is confirmed by the amplitude spectrum of the bandpass filter output generated

[25] A *min-max* normalization has been applied to the vector $\hat{\boldsymbol{\psi}}$ (4.66); the k-th element $\bar{\psi}[k]$ of the normalised phase vector is evaluated as $\bar{\psi}[k] = (\hat{\psi}[k] - \min(\hat{\boldsymbol{\psi}}))/(\max(\hat{\boldsymbol{\psi}}) - \min(\hat{\boldsymbol{\psi}}))$, with $k = 0, 1, \ldots, N_{c,\mathrm{x}} - 1$.

[26] In an ECG signal, NN intervals are represented by the time intervals between adjacent peaks resulting from sinus node depolarizations

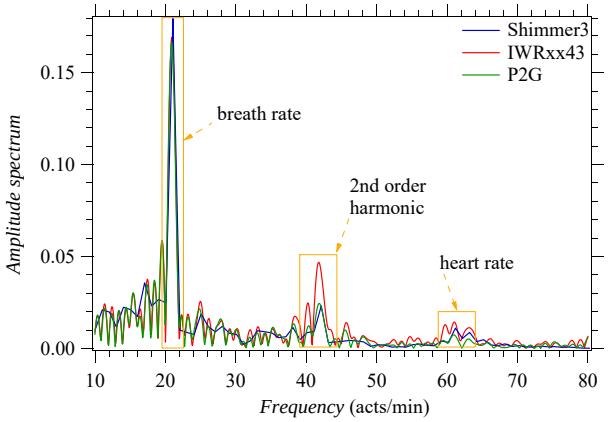


Figure 4.20: *Representation of the amplitude spectrum of: a) the breath signal acquired through our reference sensor (blue line); b) the normalized unwrapped phase extracted from our IWRxx43 radar measurements (red line) and P2G radar measurements (green line).*

in response to the two radar signals; the resulting two spectra are represented in Fig. 4.22, where the amplitude spectrum of the response of the same band-pass filter to the breathing signal acquired by the reference sensor is also shown. Note that, in all these amplitude spectra, *three spectral peaks* are visible between 60 and 70 acts/min, i.e. in the frequency range in which the spectral contribution due to heartbeat is expected. This is due to the fact that: a) the third order harmonic of the breathing signal is close to the fundamental frequency of the heart beat, so that the spectral contribution of the former signal may partially overlap with that of the latter one; b) as suggested by the time-domain signals shown in Fig. 4.19, the HR changes during our observation interval (lasting 60 s) and this entails some spectral broadening. In principle, the last phenomenon can be mitigated by reducing the duration of the observation interval over which spectra are computed; however, an excessive shortening of this interval may lead to an inaccurate estimation of the frequency components due to heart activity.

Let us take into consideration now the second scenario. The signal processing chain employed in this case is represented by Fig. 4.7. The elements of a normalised unwrapped phase vector $\hat{\boldsymbol{\psi}}$ (4.66), referring to a single observation interval and available after DC offset removal, are shown, together with the breath signal generated by our reference sensor, in Fig. 4.23. From these results it is easily inferred that: a) the phase evolution follows the dynamic of the chest of the monitored subject; b) the dynamic of heart manifests itself as small fluctuations in the observed phase. The amplitude spectra generated by applying an FFT of order $N_0' = N_{c,\mathrm{S}} = 15317$ to each of the signals appearing in Fig. 4.23 are shown in Fig. 4.24. The estimate of the BR, $\hat{b}_r = 13.8$ acts/min, is easily found by identifying the position of the main spectral peak; the amplitude of this peak is higher than that of the peak originating from heart, which is visible at the frequency $\hat{h}_r = 78$ acts/min. Note also that, similarly to what has been observed in the first scenario, in the amplitude spectrum *three spectral peaks* are visible between 70 and 80 acts/min, i.e. in the frequency range in which the spectral contribution due to heartbeat is expected. This is mainly due to the fact that the HR changes during the observation interval. In our signal processing chain, a fourth order

band-pass Butterworth filter has been employed to extract the heart signal from the unwrapped phase. This filter allows us to select the spectral components belonging to the interval $[0.91, 3]$ Hz; its output is shown in Fig. 4.25, where the ECG signal generated by our reference sensor is also represented. Note that, in the considered observation interval, the peak-to-peak period of the radar waveform is comparable to the NN peak interval characterizing the ECG signal.

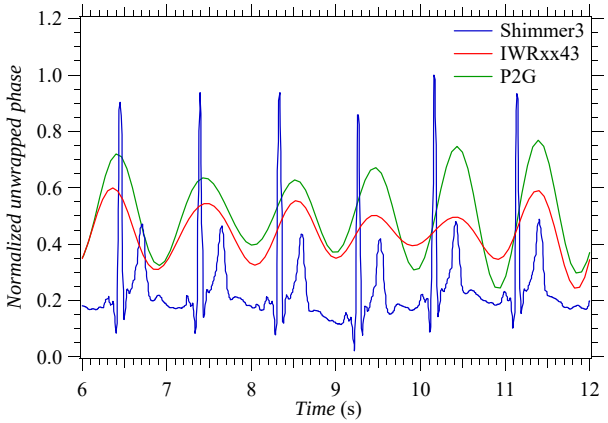


Figure 4.21: *Representation of the heart signal (normalized unwrapped phase) extracted from our IWRxx43 radar measurements (red line) and P2G radar measurements (green line). The ECG signal acquired through our reference sensor is also shown (blue line).*

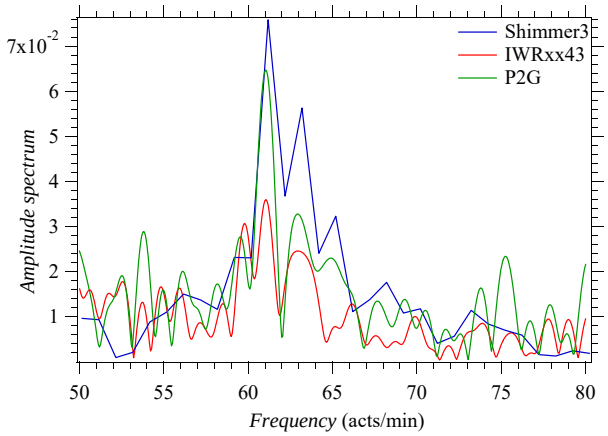


Figure 4.22: *Representation of the amplitude spectrum of: a) the heart signal acquired through our reference sensor (blue line); b) the normalized unwrapped phase extracted from our IWRxx43 radar measurements (red line) and P2G radar measurements (green line).*

Finally, it is worth pointing out that the use of a band-pass filter represents a conceptually simple solution to the problem of extracting the HR components from the phase signal provided by a radar device. Actually, this filter needs to be adapted to the specific conditions of the patient under test. If his/her BR increases unexpectedly (e.g., in the case of hyperventilation), the spectral components associated with the breathing

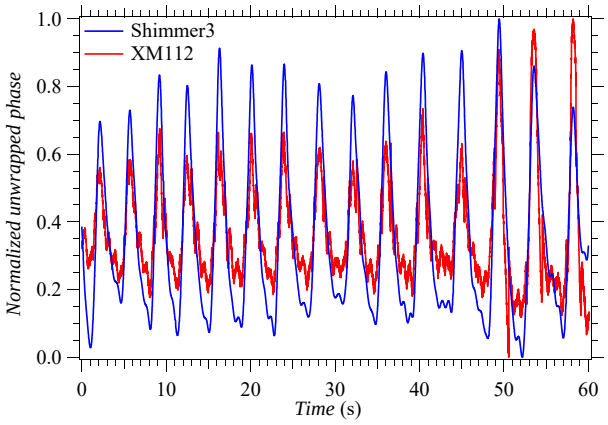


Figure 4.23: *Representation of the normalized unwrapped phase extracted from our XM112 radar measurements (red line). The breath signal acquired through our reference sensor is also shwon (blue line).*

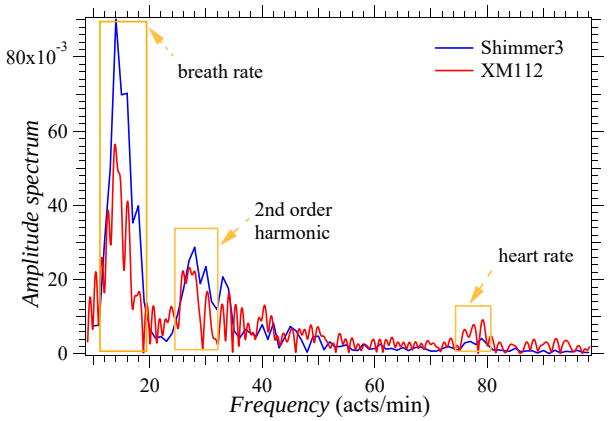


Figure 4.24: *Representation of the amplitude spectrum of: a) the breath signal acquired through our reference sensor (blue line); b) the unwrapped phase extracted from our XM112 radar measurements (red line).*

activity may not be canceled by this filter and may overlap with the components due to heart beat.

4.6.4 Estimation accuracy

Let us suppose that an $\bar{N}$-dimensional set $\mathscr{D} \triangleq \{(\hat{b}_r\,,\,\hat{h}_r\,,\,b_r\,,\,h_r);\, r = 0, 1, ..., \bar{N}-1\}$, where $\hat{b}_r$ and $\hat{h}_r$ denote the estimates of the BR b_r and of HR h_r, respectively, is available after that all our radar-based measurements have been processed. Then, the accuracy achieved in vital sign estimation can be assessed by evaluating the *mean absolute error* (MAE)

$$\hat{\varepsilon}_{m,x} \triangleq \frac{1}{\bar{N}} \sum_{i=0}^{\bar{N}-1} |\hat{x}_r[i] - x_r[i]|, \tag{4.67}$$

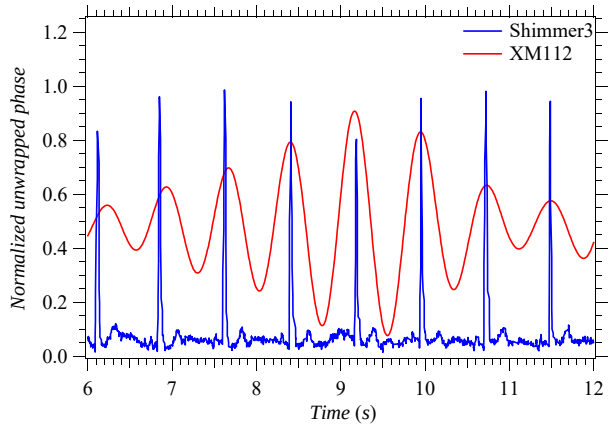


Figure 4.25: *Representation of the heart signal (normalized unwrapped phase) extracted from our XM112 radar measurements (red line). The ECG signal acquired through our reference sensor is also shown (blue line).*

the *peak absolute error* (PAE)

$$\hat{\varepsilon}_x \triangleq \arg \max_{0 \leq i \leq \bar{N}-1} |\hat{x}_r[i] - x_r[i]|, \tag{4.68}$$

and the *root mean square error* (RMSE)

$$\bar{\varepsilon}_x \triangleq \sqrt{\frac{\sum_{i=0}^{\bar{N}-1} (\hat{x}_r[i] - x_r[i])^2}{\bar{N}}}, \tag{4.69}$$

where $x = b$ ($x = h$) if BR (HR) is considered. Other relevant parameters are represented by the *correlation of variation* (CV)

$$\bar{\gamma}_x \triangleq \frac{\bar{\varepsilon}_x}{\frac{1}{\bar{N}} \sum_{i=0}^{\bar{N}-1} \hat{x}_r[i]} \tag{4.70}$$

and the *Pearson coefficient*

$$\rho_x \triangleq \frac{\sum_{i=0}^{\bar{N}-1} (\hat{x}_r[i] - \bar{x}_r)}{\sqrt{\sum_{i=0}^{\bar{N}-1} (\hat{x}_r[i] - \bar{x}_r)^2}} \frac{\sum_{i=0}^{\bar{N}-1} (x_r[i] - \bar{x}'_r)}{\sqrt{\sum_{i=0}^{\bar{N}-1} (x_r[i] - \bar{x}'_r)^2}}, \tag{4.71}$$

where

$$\bar{x}_r = \frac{1}{\bar{N}} \sum_{i=0}^{\bar{N}-1} \hat{x}_r[i], \tag{4.72}$$

$$\bar{x}'_r = \frac{1}{\bar{N}} \sum_{i=0}^{\bar{N}-1} x_r[i] \tag{4.73}$$

$x = b$ ($x = h$) if BR (HR) is considered and $x_r[i]$ represents the BR (HR) provided by the adopted reference sensor. Note that value of the parameter ρ_x (4.71) falls in the interval $[-1\,,\,1]$; a positive (negative) unitary value is found when the two available

Table 4.4: *Root mean square error ($\bar{\varepsilon}_x$), peak absolute error ($\hat{\varepsilon}_x$), mean absolute error ($\hat{\varepsilon}_{m,x}$), correlation of variation ($\bar{\gamma}_x$) and Pearson coefficient (ρ_x) referring to the BR (HR) estimates $\{\hat{b}_r\}$ ($\{\hat{h}_r\}$) computed on the basis of the data set acquired in our first scenario (the* measurement unit*, m.u., is specified for each parameter).*

Errors	m.u.	TI Radar	P2G Radar
$(\bar{\varepsilon}_b\,,\,\bar{\varepsilon}_h)$	acts/ min	(0.75 , 1.90)	(1.15 , 1.70)
$(\hat{\varepsilon}_b\,,\,\hat{\varepsilon}_h)$	acts/min	(2.80 , 4.80)	(5.50 , 4.00)
$(\hat{\varepsilon}_{m,b}\,,\,\hat{\varepsilon}_{m,h})$	acts/min	(0.45 , 1.60)	(0.60 , 1.40)
$(\bar{\gamma}_b\,,\,\bar{\gamma}_h)$	%	(4.40 , 3.24)	(6.80 , 2.90)
$(\bar{\rho}_b\,,\,\bar{\rho}_h)$	%	(90 , 82)	(84 , 85)

datasets (namely, the dataset generated through the employed radar device and that acquired from the reference sensor) exhibit a positive (negative) correlation, whereas a null value means that they are completely uncorrelated.

Typical values of the MAE, the RMSE, the CV and the Pearson coefficient evaluated in radar-based monitoring of vital signs can be found in [211, 248, 278]. In those manuscripts, a reasonable estimation accuracy is achieved if: a) the MAE and RMSE for BR (HR) estimation are in the order of some acts (beats) per minute; b) the CV is close to 5% for both BR and HR estimation; c) the Pearson coefficient is greater than 70%. It is also important to remember that estimation accuracy can be improved through the development of a proper measurement setup; for instance, a small laser device can be employed to verify that the employed radar device is really pointing toward the center of the chest of the patient under test. Another relevant technical issue to be taken into consideration is represented, as already mentioned in the previous subsection, by the availability of an accurate synchronization between the radar and the reference sensor; unluckily, if the radar and reference devices have independent local clocks, full synchronization cannot be achieved and an accurate evaluation of the parameters defined above is not possible.

Let us analyse now some results obtained on the basis of a limited set of the measurements acquired in the first scenario described in Subsection 4.6.2 (and involving the IWRxx43 and P2G FMCW radars, and the Shimmer3 reference sensor). We assume that:

a) All our measurements have been acquired from 13 young healthy subjects lying down on a bed and each of them refers to an observation interval lasting 60 s.

b) The HR and BR estimates have been computed by resorting to the processing chain described in Fig. 4.7 and in Subsection 4.5.1.

c) The size of the dataset $\mathscr{D}_{\rm TI}$ ($\mathscr{D}_{\rm P2G}$) referring to the IWRxx43 (P2G) radar is $\bar{N} = 45$ (on the average, three distinct measurements have been acquired from each subject).

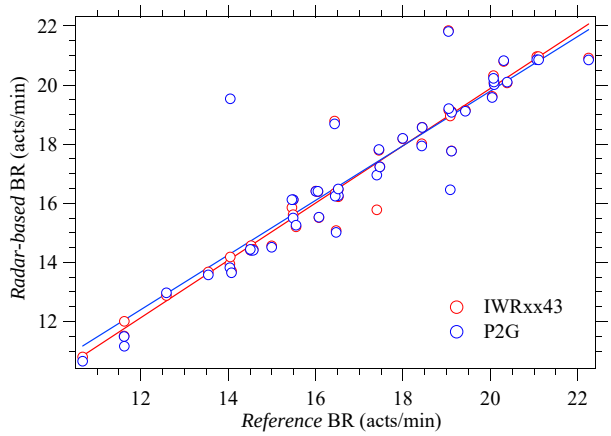


Figure 4.26: *Representation of* 45 *couples* $\{(b_r, \hat{b}_r)\}$*; the BRs estimated through a IWRxx43 (P2G) radar are identified by a red (blue) marker.*

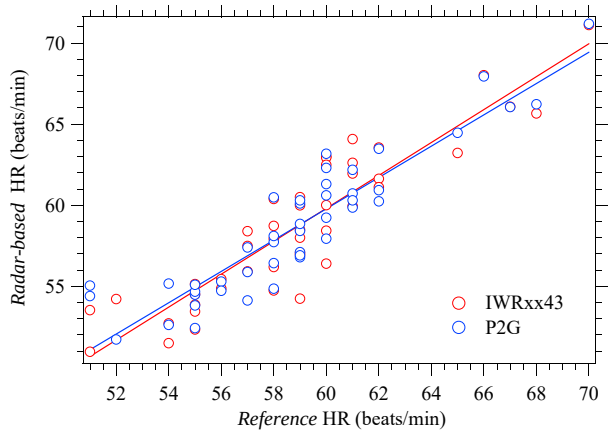


Figure 4.27: *Representation of* 45 *couples* $\{(h_r, \hat{h}_r)\}$*; the HRs estimated through a IWRxx43 (P2G) radar are identified by a red (blue) marker.*

The values of MAE, PAE, RMSE and CV evaluated on the basis of the available datasets are listed in Table 4.4, whereas the pairs $\{(b_r\,,\,\hat{b}_r)\}$ ($\{(h_r\,,\,\hat{h}_r)\}$) are represented on a Cartesian plane in Fig. 4.26 (Fig. 4.27) for both the IWRxx43 and P2G FMCW radars. Moreover, in Figs. 4.26 and 4.27, the lines generated through a linear fitting of the available pairs are also shown. These results deserve the following comments:

a) The estimates of BR and HR evaluated through both our radar devices are reasonably accurate;

b) Radar based-estimates are highly correlated with the measurements provided by our reference sensor.

c) The IWRxx43 radar achieves a better accuracy on BR estimation than the P2G radar; on the other hand, the latter device outperforms the former one in HR estimation.

d) The accuracy of the HR estimates is worse than those of BR estimates, since the latter rely on stronger and cleaner spectral information (see fig. 4.22 and 4.24). Note also that the RMSEs and peak errors of HR are in the order of few acts/min.

As far as the last issue is concerned, it is important to keep in mind that: a) the HR frequency may overlap with (or be very close to) the third order harmonic of breath; b) the displacement due to heart is very small and may be not fully detected by our radar devices, if they are not accurately oriented towards the chest of the subject under test.

Let us analyse now some results obtained on the basis of all the measurements acquired in the second scenario described in Subsection 4.6.2 (and involving the XM112 UWB radas and the Shimmer3 reference sensor). We assume that the HR and BR estimates have been computed through the processing chain described in Fig. 4.7 and in Subsection 4.5. The pairs $\{(b_r, \hat{b}_r)\}$ and $\{(h_r, \hat{h}_r)\}$ obtained in this case are represented in Fig. 4.28, whereas the corresponding values of the parameters defined in this subsection are listed in Table 4.5. Comments similar to those expressed for the first scenario also apply to these results. In fact, the estimation errors estimated in the last case are comparable with those listed in Table 4.4 and referring to the two FMCW radars we employed.

Table 4.5: *Root mean square error ($\bar{\varepsilon}_x$), peak absolute error ($\hat{\varepsilon}_x$), mean absolute error ($\hat{\varepsilon}_{m,x}$), correlation of variation ($\bar{\gamma}_x$) and Pearson coefficient (ρ_x) referring to the BR (HR) estimates $\{\hat{b}_r\}$ ($\{\hat{h}_r\}$) computed on the basis of our data set acquired in our second scenario (the* measurement unit*, m.u., is specified for each parameter).*

Errors	m.u.	XM112
$(\bar{\varepsilon}_b\,,\,\bar{\varepsilon}_h)$	acts/ min	(0.87 , 1.32)
$(\hat{\varepsilon}_b\,,\,\hat{\varepsilon}_h)$	acts/min	(4.1 , 3.9)
$(\hat{\varepsilon}_{m,b}\,,\,\hat{\varepsilon}_{m,h})$	acts/min	(1.32 , 1.75)
$(\bar{\gamma}_b\,,\,\bar{\gamma}_h)$	%	(10.1 , 2.38)
$(\bar{\rho}_b\,,\,\bar{\rho}_h)$	%	(78.8 , 97.5)

4.7 Applications of the Radar Technology to Vital Signs Monitoring

In this section, we propose a synopsis of the technical literature concerning the following specific issues: a) the monitoring of HR and BR; b) the experimental setups adopted in vital sign monitoring; c) the monitoring of heart sounds.

4.7.1 Heart rate and breath rate monitoring

The use of radar systems for monitoring HR and BR has been investigated by several research groups, whose work has allowed to assess the performance achieved by different

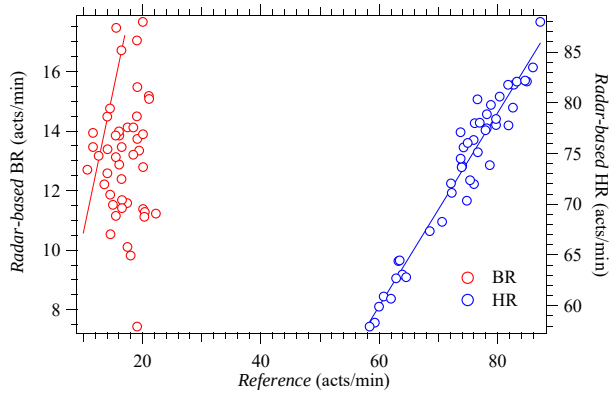


Figure 4.28: *Representation of* 45 *couples* $\{(b_r, \hat{b}_r)\}$ *and* $\{(h_r, \hat{h}_r)\}$ *estimated on the basis of the measurements acquired through a XM112 UWB radar.*

radar prototypes operating at distinct frequencies and radiating heterogeneous signals. As far as the use of the radar technologies described in Subsection 4.4.1 is concerned, it is worth mentioning that:

a) *Continuous wave Doppler radars* operating at 2.4 GHz, at 5.8 GHz and at 24 GHz have been employed to measure respiration, heartbeat, or motion activity in [140, 188, 200, 279], in [280] and in [211, 259, 281], respectively. In [282], instead, a 24 GHz CW radar sensor is employed for cuffless blood pressure measurement, whereas, in [283], a dual radar system operating at 5.8 and 120 GHz is adopted to measure respiration, heartbeat, HRV, blood pressure, and other vital parameters.

b) *Frequency modulated continuous wave* radars operating in the C and X bands, at 24 GHz, at 60 GHz, at 77 GHz, and at 122 GHz have been used to simultaneously estimate of the the vital parameters of multiple people in [213, 284], [285], in [278, 286], in [222, 248] and in [287], respectively.

c) *Stepped frequency continuous wave* radars operating at frequencies lower than 3 GHz and in the X band are have been exploited for the estimation of the vital signs parameters of single or multiple people in [164, 288] and in [289, 290], respectively.

d) *Impulse radio* UWB radars operating in the C, X and E bands and having large bandwidths ($2 - 3$ GHz) have been employed to measure vital parameters with high accuracy in [212, 249, 262, 291–293].

Model-based (MB) and LB methods have been employed for the processing of the measurements provided by the above mentioned radar systems; moreover, such measurements have been acquired over population of different sizes. Essential information about the adopted processing methods and the size of the involved population are summarised in Table 4.6; note that, in this table, the size N_p of the population (*low*, L,

medium, M, or *high*, H) on which it has been tested have been specified[27]. Some details about the most important processing methods and the main results achieved through their use are provided below.

Continuous wave Doppler radars - Three novel deterministic techniques for estimating vital signs have been developed in [279], where a radar system operating at 2.4 GHz has been tested on a (single) human subject located at a fixed distance. These techniques are based on: a) the FFT processing of the time-domain phase signal estimated through AD (see the block diagram represented in Fig. 4.7 and in Subsection 4.5.1); b) the computation of the autocorrelation of the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) to estimate the period of the time-domain phase signal; c) the FFT processing of the above mentioned autocorrelation. The last two techniques are shown to achieve a better accuracy than the first one. Other interesting results are offered by [188], in which a deterministic method based on the block diagram appearing in Fig. 4.7 is adopted to process the measurements acquired through a mmwave Doppler radar. In that manuscript it is shown that the HR estimated on the basis of an ECG is highly correlated with that estimated on the basis of the measurements acquired through the devised radar system.

Some potential benefits originating from the use of ML methods are illustrated in [140], where the problem of recognizing and classifying breathing disorders of various patients during their sleep is investigated. Five ML techniques, namely, SVM, *linear discriminant analysis* (LDA), K-NN, *decision tree* (DT) and *ensemble learning techniques*, trained on a manually selected set of features, are tested; the obtained results evidence that all the considered techniques are able to achieve an high accuracy in the classification of breathing disorders.

Frequency modulated continuous wave radars - In [286] the BR and HR of a single person have been measured through an FMCW radar and an ECG; moreover, a video camera has been also used to inspect the chest dynamics of the patient under test. A dataset composed by the measurements acquired from six different people has been used for validating the adopted signal processing methods. Two different methods have been tested, one based on the FFT (similar to the one described by the block diagram shown in Fig. 4.7), the other one based on the computation of the autocorrelation of the phase vector $\hat{\boldsymbol{\psi}}$ (4.28). The obtained results have evidenced that: a) both BR and HR can be accurately extracted from radar measurements; b) the estimation of HR in the presence of the respiration harmonics can be challenging. Methods similar to the ones employed in [286] have been successfully exploited in [287] and in [278], for simultaneously estimating the vital signs of multiple people and for identifying the vital parameters of ten subjects that experience different sleep scenarios, respectively. In [285], a system for measuring blood pressure is described; results on human subjects reveal that radar-based arterial pulse detection is very promising for future applications in blood pressure detection and monitoring.

Stepped frequency continuous wave radars - An FFT-based estimation method, similar to the one described in Subsection 4.5.1, has been employed in [164,288] to estimate the BR and HR of multiple subjects in a room; these subjects have been sitting on a chair or lying down in a bed, and have been characterized by different orientations. In [164], radar-based estimates have been compared with the HR and BR estimates

[27] In this table, the population size is *low*, *medium* or *high* if $1 < N_p \leq 10$, $10 < N_p \leq 20$ or $N_p > 20$, respectively.

Table 4.6: *Classification of the references cited in Subsections 4.7.1 and 4.7.2 on the basis of: a) the size of the acquired dataset (L, M and H correspond to $1 < N_p \leq 10$, $10 < N_p \leq 20$ and $N_p > 20$, respectively, where N_p denotes the overall number of acquisitions); b) the position of the subject/subjects under test; c) the employed radar technology; d) the category of the adopted signal processing method.*

Ref. no.	Dataset size	Subject position			Radar technology				Freq.	Signal proc.	
		sitting	lying down	other	CW	FMCW	SFCW	UWB	GHz	MB	LB
[279]	H	✓			✓				2.4	✓	
[280]	H	✓			✓				5.8	✓	
[188]	L			✓	✓				2.4	✓	
[140]	M		✓		✓				2.4		✓
[259]	M	✓			✓				24		✓
[281]	M		✓		✓				24	✓	
[211]	L		✓		✓				24	✓	
[200]	L			✓	✓				2.4	✓	
[282]	L			✓	✓				2.4	✓	
[283]	L	✓			✓				5.8,120	✓	
[285]	M			✓		✓			60	✓	
[286]	L		✓			✓			24	✓	
[287]	L	✓	✓			✓			122	✓	
[248]	L		✓			✓			77	✓	
[213]	L	✓				✓			5.8	✓	
[278]	M	✓				✓			24	✓	
[284]	L			✓		✓			10	✓	
[288]	M		✓				✓		0.3-1.3	✓	
[164]	M	✓					✓		2.4	✓	
[289]	L						✓		9.0	✓	
[290]	L						✓		5.8	✓	
[222]	L	✓				✓		✓	77	✓	
[249, 291]	L	✓						✓	6.8	✓	
[292]	L	✓		✓				✓	4.3	✓	
[293]	H		✓					✓	94	✓	
[212]	H	✓						✓	7.25		✓
[262]	H	✓						✓	79		✓

provided by a contact reference sensor. The employed radar system has been shown to achieve the best performance when the chest of the subject under test is in front of it; however, even if this condition is not met, radar-based estimates are still sufficiently accurate. In [288], the performance of radar-based monitoring in scenarios in which obstacles, characterised by different shapes and materials (like walls), are placed between the radar and the subject under test are evaluated. Finally, an SFCW radar system has been employed in [289, 290] to simultaneously estimate the vital signs of multiple people. It is also worth mentioning that the estimation method devised in [290] consists in applying a CWT to the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) in order to separate the breath contribution from the one due to heart.

Impulse radio UWB radars - An FFT-based estimator (similar to the one described by the block diagram of Fig. 4.7) and an estimator based on autocorrelation of the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) have been employed in [293] and [292], respectively, where the raw data acquired through an IR-UWB radar system have been processed to extract the vital signs of different subjects. The accuracy achieved by an IR-UWB radar system in vital sign estimation has been compared with that of an FMCW radar system in [222]; the latter system is shown to outperform the former one thanks to its ability to perform clutter suppression. Note that, as shown in [291], noise and clutter affecting the measurement acquired through an IR-UWB radar can be mitigated by means of Kalman filtering.

4.7.2 Radar setups in real world scenarios

Radar systems can be potentially exploited for remote monitoring of vital signs as conveniently and as easily as wearable devices in heterogeneous scenarios. Readers should not forget, however, that essential requirements for their adoption in real world applications are represented by their ease of use and accuracy. Various results about the accuracy of radar systems for vital sign monitoring in real world (and often challenging) scenarios can be found in [211, 213, 248, 278, 284] and [294]; note that [213], [248] and [278] concern FMCW radar systems, whereas [211, 284] and [294] CW Doppler and IR-UWB radars, respectively. More specifically, the use of radar systems in an hospital for measuring the vital signs of patients has been investigated in [211,213,248], whereas more challenging scenarios have been taken into consideration in [278,284,294]. The use of a IR-UWB radar for through the wall BR and HR estimation has been proposed for a single subject and multiple subjects in [295] and [296], respectively. A brief description of the experimental setup adopted in the measurement campaigns conducted by the authors of the manuscripts cited above and of the main results extracted from their experimental data is provided below.

a) The use of a MIMO FMCW radar in an *hospital bedroom* has been investigated in [248]. The radar device has been positioned on the ceiling of a room, in front of the bed on which a static subject was lying down; moreover, the monitored subject was facing up the radar during the data acquisition process, that lasted 40 min. A good correlation between the BR and HR acquired through a reference sensor and their estimates provided by the radar system has been found. Note that the scenario considered in this case is that of a typical hospital room in

which vital signs monitoring concerns patients at rest; for this reason, this task is accomplished in the absence of random movements of their chests.

b) An FMCW radar endowed with a custom array of antennas has been employed to estimate the vital signs of *a person sitting on a chair* in [213]. The analysed measurements have been acquired on five subjects (three males and two females), not suffering from any cardiac or respiratory pathology, and whose age ranged from 25 to 63 years. These subjects have been positioned in front of the radar with different orientations (their chest, their left, their back and right side facing the radar antenna). The employed system has been able to accurately detect BR and HR, regardless of chests patient orientation towards the radar antenna.

c) A custom-designed CW radar system placed *under a bed mattress* has been employed in [211] to continuously measure the vital parameters of various patients without restricting their movements. The developed system has been able to detect the activity of each patient (i.e., entering the bed, getting out of it, or moving inside it) and to continuously measure his/her vital signs in different positions.

d) An FMCW radar has been employed for the estimation of the vital signs of *a person experiencing different sleep conditions* in [278]. In this case, the radar device has been positioned on the ceiling of a room in front of the bed on which a static patient was lying down in different positions or was performing simple activities; this allowed to emulate real-life sleep conditions. The measurements have been acquired on eleven patients (whose age ranged from 25 to 55 years); a good correlation has been found between the radar-based estimates of vital signs and those provided by a reference device. These results, together with those illustrated in [211], have evidenced that radar-based systems can be very useful in various healthcare applications (e.g., in the study of sleep apnoea, in the monitoring of bedridden patients and, more in general, in the monitoring of hospitalized patients).

e) The use of a *dual-frequency*[28] CW microwave radar for vital sign estimation *inside an ambulance* has been studied in [284]. The measurements have been acquired on eight healthy male subjects whose age ranged from 21 to 24 years. Each of the subjects was lying down on a stretcher contained inside an isolator; therefore, his respiratory and cardiac activities were monitored from outside the isolator. The employed radar system has been able to measure the HR and BR of the monitored subject with good accuracy, in both static and dynamic conditions of the ambulance. Note that the innovative method proposed for vital sign detection could be very useful in other scenarios (e.g., in the monitoring of infectious patients).

f) Various results about the use of an IR-UWB radar system for *monitoring the BR of six neonates*[29] in a *neonatal intensive care unit* have been illustrated in [294]. In all the experiments, the employed radar system has been covered with a white

[28]The considered radar system operated at both 10 GHz and 24 GHz.

[29]In this case, the subjects under test were two males and four females with a median gestational age of 38 weeks and a median birth weight of 3100 g.

plastic cap and has been hung at the end of a specially designed arm placed on the top of a tripod; moreover, it has been placed at a distance of 35 cm from the chest of each subject. Each of the neonates has been placed in a supine position inside an open-air crib, and his/her torso has been covered with a blanket. An high accuracy has been achieved, despite the small movements of the babies under test.

g) Some results about the use of IR-UWB radar systems in vital sign monitoring have been illustrated in [295] and [296]. In particular, in [295], an IR-UWB monostatic radar system operating at 4.3 GHz and having a bandwidth (resolution) of 2.3 GHz (6.5 cm) has been employed for vital signs estimation of a single subject located behind: 1) a gypsum wall, 2) a wooden door, 3) a brick wall; 4) a load bearing concrete wall. Three different methods have been proposed for extracting vital signs from radar measurements, namely: 1) a DFT-based method for BR estimation; 2) a clutter reduction technique based on the *singular value decomposition* (SVD); 3) a STFT for analysing the temporal evolution of the spectral components of the received signal. The numerical results provided in the manuscript cited above confirm that the proposed methods are achieve similar performance in the considered scenario. In [296], instead, a MIMO IR-UWB radar, equipped with an array of 10×10 physical elements and operating at a centre frequency of 2 GHz, is employed to detect the vital signs of three male adults (having different heights and weights), sitting in front of the radar system at a distance that does not exceed 2.5 meters; both line-of-sight conditions and the presence of a wall hiding them are considered. A simple FFT processing has been considered. The obtained numerical results evidence that a MIMO IR-UWB *radar can achieve through-wall detection of multiple subjects, and estimate their BR and HR, provided that their body movements are quite small.*

4.7.3 Heart sounds monitoring

Currently, the standard reference tool for continuous cardiac monitoring is electrocardiography; it requires touch-based wiring of patient skin. In clinical practice, a series of pathological processes would certainly benefit from contactless monitoring; these conditions may include patients with an infections or a sepsis (e.g., infected by SARS CoV-2), or patients with mental disorders that do not allow conventional monitoring according to cardiology settings. In cardiac intensive units or in intermediate care units, continuous monitoring of heartbeat is a common practice, because detection and prevention of critical states of health can lead to timely therapeutic interventions, with the result of a better outcome. In fact, several cardiac conditions could be immediately harmful and potentially fatal (e.g., cardiac arrhythmias, heart attacks, acute heart failure, and stroke), thus requiring prompt detection of cardiac or respiratory anomalies. Recently, in the field of radar-based systems for vital signs monitoring, an effort has been made to take a step forward and, in particular, to detect *heart sounds* [271, 272, 297]. The classification of normal or abnormal heart sounds has been investigated in [297]. In that manuscript, it has been shown that the heart signal recorded by a custom-designed CW radar system is highly correlated with the signal registered by a phonocardiograph, used as reference. Based on this correlation, each period of the recorded radar signal can be

divided in the different temporal phases of the cardiac activity, namely into *systole* and *diastole* (see Subsection 4.3.1); this procedure is called *heart sound segmentation*. After recognizing systole and diastole in the received radar signal, a reliable detection of normal or abnormal heart sounds can be accomplished using an *Long Short Time Memory* (LSTM) network for heart sound segmentation; this has been shown in [271], where an heterogeneous dataset of recorded heart sounds and vital signs acquired through a CW Doppler radar has been used [272]. The measurements of this dataset originate from multiple subjects in different positions of their bodies and in various conditions, such as during breath-holding, during speaking and after post-exercise. The obtained results have evidenced that more than 90% of the recordings were of high quality, and that the correlation between radar and ECG signals was almost perfect. As far as the impact of body position is concerned, it has been found that, in general, heart sounds can be detected in almost all the considered positions; however, the measurability at a certain position changes from subject to subject. This is partly due to the fact body anatomy, and in particular, heart position inside chest, may exhibit some variations from person to person.

Further analysis should be performed to achieve solid results, as these represent only preliminary data about the innovative radar-based detection of heart sounds. This technology could lead to a new way of non-invasive cardiac monitoring and could be revolutionary in the cardiology settings. Currently, during medical visits, cardiologists use a stethoscope for heart evaluation; this tool allows them to check for sounds which may indicate pathological changes in the heart or the heart valves. However, the validity of the assessment strongly depends on the experience of the physician. An objective, operator-independent and automated analysis of heart sounds accomplished through radar technology would be very useful; note also that, in this case, the availability of large datasets for big data analysis would be very useful.

4.8 Current Trends

In this section, the most relevant trends in the ongoing research activities on radar systems for vital signs monitoring are described. In particular, we focus on research activities related to: a) the techniques for the compensation of *random body movements* (RBMs); b) the impact of body orientation on the detection of vital signs and the quality of their estimates; c) the detection of HRV; d) the exploitation of radar-based monitoring for user identity authentication.

4.8.1 Compensation of random body movements and impact of body orientation

The *random body movements* (RBMs) of any person affect the estimate of his/her vital signs provided by radar devices. These movements, in fact, modulate the radar waveform both in its amplitude and frequency, so distorting the received echoes. This may significantly affect the quality of the estimates of vital signs generated by radar systems. Consequently, RBMs may represent an important obstacle to the adoption of radar technology in certain scenarios. It is also worth mentioning that, if a mmwave radar device is employed, the *eye blinking* of the monitored subject may be detected as

a large RBM, resulting in a significant degradation in the quality of the received signal phase; the impact of this disturbance can be substantially mitigated through the use of MIMO radars, since these are able to steer their beam towards his/her chest [207].

Recently, two approaches to motion compensation have been proposed. The first approach is an hardware RBM compensation acting at RF front-end level and, consequently, limiting the risks of saturating radar transceivers in the presence of strong echoes [206]. The second one, instead, consists in accomplishing a digital compensation after demodulation; this can be implemented more easily and lends itself to a more precise control. For this reason, in the remaining part of this subsection, we focus on the last approach only.

One of the easiest strategies to extract the (weak) vital sign components from the radar measurements and removing the distortion due to RBMs is represented by digital filtering. However, this solution is optimal only when the statistical characteristics of the filter input match prior information on which the design of the filter is based; unluckily, such characteristics are normally unknown. For this reason, a more robust solution has been proposed in [298], where an *adaptive noise cancellation* (ANC) recursive algorithm is employed, in combination with polynomial fitting, in a CW Doppler radar system. It is important to point out that: a) polynomial fitting is used to reconstruct the signal components due to RBMs and that must be subtracted from the overall signal provided by the radar receiver; b) this strategy allows to achieve RBM compensation when a single radar system is used and does not require additional sensors. The use of a multi-radar system for cancelling RBMs have been proposed in [205]. In this case, two CW Doppler radar systems have been put on opposite sides with respect to the body of the monitored subject; when his/her body was leaning towards one of the radars, it moved away from the other one [205], so that the distance between each radar and the body changed in an opposite manner. Based on this consideration, it has been shown that RBMs can be cancelled by combining the measurements provided by the two radars. The main drawbacks of a multi-radar approach consists in an increase of system complexity, cost and power consumption and in the need of a larger room for the experimental setup. A different approach to RBM compensation relies on the use of an hybrid system including radar and camera [299]. In this case, the information provided by the camera has been used to compensate for the phase distortion due to body movements. Unluckily, this approach has been shown to work well when body movements are regular and deterministic.

As its can be easily inferred from our previous analysis, RBM compensation should be considered as an open research problem since few solutions are available in the technical literature.

Another open research problem mainly concerns the impact of *body orientation* on the detectability of HR and BR, and on the quality of their estimates. The breathing movement and the heart dynamics can be detected not only if the radar is positioned in front of the chest of the monitored subject, but also when his/her body has a different orientation with respect to the other radar itself, even if some degradation is experienced in the estimation of vital signs [213]. Some interesting methods for correctly estimating BR and HR in the presence of different body orientations have been proposed in [212] and [300]. In particular, the use of an ANN for fusing measurements coming from three different radar sensors, distributed as endpoints of an equilateral triangle, and

for compensating the effect of body orientation is investigated in [212]. The obtained results show that the network is able to achieve a 95% score in the classification of six different body orientations and to compensate for them; moreover, an accurate estimate of HR is obtained for every body orientation. The method developed in [300] is also based on the exploitation of a neural network for the compensation of body orientation in the estimation of BR and HR; however, the employed mm-wave radar, mounted on a robot system, allows to acquire four different sitting poses and 180 minutes of data.

4.8.2 Heart rate variability

The HR estimated by means of a radar system represents a measure of the overall number of heartbeats observed over a given time interval (e.g., over one minute). However, we should not forget that, within a certain observation interval, the temporal distance between two adjacent heartbeats may not remain constant. This phenomenon, known as HRV, is related to heart-brain interactions and is regulated by the *autonomic nervous system* [301,302]. More specifically, HRV reflects beat-to-beat changes in peak RR intervals[30], which depend on the interrelation between sympathetic and vagal tones. In fact, the sinus node, the principal heart's pacemaker, has its own intrinsic activity; however, several external and internal stimuli altering the autonomic balance influence the final HR [303].

Heart rate changes may originate from a variety of conditions such as mental or physical stress, cardiac or noncardiac diseases, or pharmacological or invasive treatments; the respiration-related fluctuation of HR, known as *respiratory sinus arrhythmia*, is probably the most commonly investigated component of HRV.

The autonomic nervous system imbalance with increased sympathetic and decreased vagal tone has been proven to be associated with higher risk of cardiac mortality. Therefore, HRV has become an important and recognized tool in identifying patients at risk of cardiovascular death [304], and can be considered as a indicator for both physiological conditions and pathological processes, such as depression, diabetic neuropathy and heart failure. Moreover, it can be exploited to monitor post-surgery and post-infarction patients in order to assess the risk of ventricular tachyarrhythmias leading to sudden cardiac death.

Nowadays, different methods can be employed to measure HRV; these include a series of simple bedside reflex tests and more advanced computer-based algorithms for detecting spontaneous peak RR interval changes. The accomplished analysis is usually based on long-term (usually 24 hours) Holter ECG recordings or short-term (usually few minutes) ECG recordings, and aims at avoiding any influence from external stimuli that could affect autonomic nervous tone [305]. In general, an accurate analysis of HRV may require a long observation interval in a clinical environment or in home-care scenarios. The assessment of this phenomenon is based on the evaluation of various time-domain and frequency-domain features (see [278, Tab. 2, Par. 2.6]). Time domain features aim at quantifying the variability in inter-beat intervals (also called NN peak intervals[31])

[30]The RR interval represents the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram. It is a function of intrinsic properties of the sinus node as well as autonomic influences.

[31]In the case of a CW radar system, a NN interval can be defined as the temporal distance between

and include the *standard deviation of normal to normal peak intervals* (SDNN)

$$\text{SDNN} \triangleq \sqrt{\frac{\sum_{i=1}^{\bar{N}} \left(\Delta t^{(i)} - \Delta \bar{t}\right)^2}{\bar{N}}} \tag{4.74}$$

and the *root mean square successive difference of intervals* (RMSSD)

$$\text{RMSSD} \triangleq \sqrt{\frac{\sum_{i=2}^{\bar{N}} \left(\Delta t^{(i)} - \Delta t^{(i-1)}\right)^2}{\bar{N}-1}}; \tag{4.75}$$

here, $\bar{N}$ is the total number of beats detected in the heart signal, $\Delta t^{(i)}$ is the duration of the time interval between the $(i+1)$-th detected beat and the previous beat, and

$$\Delta \bar{t} \triangleq \frac{\sum_{i=1}^{\bar{N}} \Delta t^{(i)}}{\bar{N}} \tag{4.76}$$

is the average duration of the interval between two consecutive beats. If an estimate of the probability density function (in the form of a *probability mass function*, pmf) of the NN peak intervals is available, a further meaningful feature is represented by the *triangular index* (TRI), defined as

$$\text{TRI} = \frac{\bar{N}}{\mathscr{P}_{NN,max}}, \tag{4.77}$$

where $\mathscr{P}_{NN,max}$ is the maximum of the above mentioned pmf [301].

Frequency domain features are usually evaluated through FFT processing, and refer to the *low frequency* or to the *high frequency* band. The former band accounts for modulations whose period ranges from 7 s to 25 s, whereas the latter one refers to shorter periods. The presence of a significant component in the HF band is typically a symptom of stress or anxiety.

Radar-based estimation of HRV may represent an appealing and challenging alternative to ECG for monitoring the physical and mental status of patients. However, as far as we know, this topic is addressed by few manuscripts in the technical literature [188,200,278,287,306,307]. More specifically, various results about the use of CW Doppler radars for analysing HRV and drowsiness can be found in [188,306] and [200], respectively. The other manuscripts, instead, involve FMCW radars. In particular, an FMCW radar operating in the K-band has been employed in [278] to monitor HRV in eleven patients of different ages during their sleep. The obtained results evidence that the time and frequency features extracted from the radar signal are correlated with those evaluated on the basis of the measurements acquired from a reference sensor. The effect of the coupling between breathing and heartbeat signals on HRV has been investigated in [287]; in this case, an FMCW radar operating at 122 GHz has been used. Finally, a completely novel approach based on deep learning (and, in particular, on LSTM neural networks) has been proposed in [307] to accurately estimate HRV, by analysing the data collected by a custom-designed, six-port, CW Doppler radar operating at 24 GHz.

---

two consecutive maxima of the phase vector $\hat{\boldsymbol{\psi}}$ available at the output of the bandpass filter appearing in the block diagram of Fig. 4.7.

4.8.3 Vital sign-based authentication

The capability of radars to accurately estimate the vital signs of a person is attractive not only for monitoring the health status of a patient and detect possible diseases, but also for *user identity authentication*. Nowadays, many people are used to log-in in their own smartphones by simply looking at their camera or thanks to their fingerprint. These approaches can be classified as *what you are* methods, since they make use of personal traits (biometrics) that are hard to reproduce or mimic. Within this category, facial recognition represents a less robust user authentication method with respect to other physiological biometrics, such as fingerprints or iris scans [308, 309]. Authentication methods exploiting radar-based identification of vital signs, instead, are gaining attention because, unlike other physiological biometric-based approaches (e.g., ECG), they do not require direct contact between the human body and the sensor. Moreover, an identity authentication system based on the recognition of breath or heart traits may be sufficiently robust and reliable, since, as evidenced by various studies, the *respiratory personality* is unique and is preserved over long periods in adult humans at rest [310, 311]. This can be related to the fact that the physiological structure (e.g., the strength of the diaphragm and intercostal muscles and volume of the thoracic cavity) and the respiratory motions associated with chest movements have specific characteristics for each person.

The use of a CW Doppler radar device for heart-based and breathing-based user authentication has been investigated in [308] and [312–314], respectively. All the proposed methods make use of ML classification algorithms fed by a set of features extracted from the phase vector $\hat{\boldsymbol{\psi}}$ (4.28). More specifically, in [312] three different sets of features have been used. The most relevant features of the first set are the BR $\hat{b}_r$ (see Subsection 4.5.1), the breathing depth, the average speed of exhale and the average speed of inhale. The other two sets of features, instead, aim at monitoring the ratio of inhale and exhale breathing areas, and the breathing mechanism right after and before the apex (full lung volume). Moreover, it has been shown that a K-NN algorithm, trained over a dataset consisting of measurements lasting 60 s and acquired over six different subjects, is able to recognize the breathing pattern of different people with a good classification score. Better classification results can be obtained by means of a SVM classifier, as shown in [313, 314].

Despite these positive results, respiratory-based identity authentication is far from being mature and requires extensive analysis and investigation. In fact, people need to be authenticated under various mental or physical states. Therefore, potential variations occurring in the normal breathing pattern of a person must be taken into account; ignoring them could reduce the identification accuracy, as observed in [314]. For this reason, an heart-based authentication approach has been proposed in [308]. In this case, the heart signal extracted from the phase vector $\hat{\boldsymbol{\psi}}$ (4.28) has been segmented in different periods, each encompassing a small number of cardiac cycles; within each period, a set of eight features has been manually extracted. The results obtained through an SVM classifier trained on the data acquired over 78 different subjects have confirmed that an authentication method based on radar-based recognition of cardiac motion is really feasible.

4.9 Conclusions

Nowadays, a significant body of literature is available in the field of radar-based monitoring of vital signs. This chapter has offered a broad introduction to this field with the aim of explaining some fundamental concepts, technologies, methods and results to a wide audience. We really hope that our overview of the available radar technologies, of the employed signal processing methods and of the specific applications being considered in medicine will stimulate the interest in radar-based monitoring of vital signs. We believe that radar technology is now mature enough for being taken into account in the medical field. However, there is still ample room for the development of accurate and computationally efficient estimation techniques, and for their implementation on commercial hardware platforms. Readers should also keep in mind that most of the results available in the technical literature refer to a very limited human population and, usually, to healthy subjects. In fact, studies about the monitoring of real patients in realistic medical scenarios are still scarce. Despite this, it has become clear that radar systems can represent a viable alternative to wearable sensors or the only possible option in some critical scenarios, where contactless monitoring is absolutely required. Furthermore, the technological improvements and the advances in processing techniques achieved in recent years have made it possible to overcome various limitations. Therefore, thanks to their capability of continuous and contactless detection, radars may revolutionise patient monitoring in hospitals and in other healthcare facilities in the near future.

5. Conclusions, considerations and future work

In this thesis various deterministic and machine learning methods for MIMO radar systems have been investigated. First of all, it has been shown that target detection and estimation method, dubbed RASCA and based on a serial estimation and cancellation approach, can outperform other well known techniques (e.g., MUSIC or other FFT-based techniques) in the presence of closely spaced targets at the price of a limited computational cost. Moreover, the RASCA method, thanks to its structure, lends itself to a parallel implementation. All the described methods have been tested resorting to both synthetically generated data and measurements acquired through commercial radars working in the W band, and their accuracy has been assessed.

Secondly, the use of machine learning and deep learning techniques in the processing of radar data has been explored. Our results have evidenced that these techniques can be exploited to successfully solve simple bidimensional tracking problems or recognize various human activities.

Finally, vital signs monitoring by means of MIMO FMCW radars has been studied. Our results have shown that good accuracy can be achieved in breath and heart rate estimation in well controlled environments (e.g., for people lying in an hospital bed); however, radar-based estimation of vital signs may be seriously affected by the presence of random body movements.

Future research work should focus on:

a) the application of a serial estimation and cancellation approach to other radar systems. In this thesis, only FMCW radars have been taken into considerations.

b) The exploitation of machine learning and deep learning techniques in various automotive applications.

c) The development of new techniques for mitigating the impact of random body movements in the estimation of vital signs.

Finally, it worth stressing that in all our tests were carried commercially available hardware has been employed; this suggests that radar technology is now mature enough to be be applied to a number of real world use cases. Despite this, the problem of the large computational power required to process the large amount of data generated by MIMO radars still remains; note that,moving such data from radar ADCs to processing units greatly increases system complexity (and cost). Unluckily, at the moment, real time execution of the techniques described in this thesis requires a computational power larger than that provided by the embedded systems currently available on the market.

Bibliography

[1] J. Winters, "On the Capacity of Radio Communication Systems with Diversity in a Rayleigh Fading Environment," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 5, pp. 871–878, June 1987.

[2] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas," *Bell Labs Tech. J.*, pp. 41–59, Oct. 1996.

[3] G. Foschini and M. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[4] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[5] G. Vitetta, D. Taylor, G. Colavolpe, F. Pancaldi, and P. Martin, ***Wireless Communications: Algorithmic Techniques***. New York: Wiley, May 2013.

[6] E. Fishler, A. Haimovich, R. Blum, D. Chizhik, L. Cimini, and R. Valenzuela, "MIMO radar: an idea whose time has come," in *Proc. 2004 IEEE Radar Conf.*, 2004, pp. 71–78.

[7] D. W. Bliss and K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging: degrees of freedom and resolution," in *Proc. 37th Asilomar Conf. Signals, Sys. Comput.*, vol. 1, 2003, pp. 54–59.

[8] J. Li and P. Stoica, "MIMO Radar with Colocated Antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, Sep. 2007.

[9] A. M. Haimovich, R. S. Blum, and L. J. Cimini, "MIMO Radar with Widely Separated Antennas," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 116–129, 2008.

[10] E. Fishler, A. Haimovich, R. Blum, R. Cimini, D. Chizhik, and R. Valenzuela, "Performance of MIMO radar systems: advantages of angular diversity," in *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, vol. 1, 2004, pp. 305–309 Vol.1.

[11] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2005.

[12] C. Pfeffer, R. Feger, C. Wagner, and A. Stelzer, "FMCW MIMO Radar System for Frequency-Division Multiple TX-Beamforming," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 12, pp. 4262–4274, 2013.

[13] R. Feger, C. Pfeffer, and A. Stelzer, "A frequency-division MIMO FMCW radar system using delta-sigma-based transmitters," in *Proc. 2014 IEEE MTT-S Int. Microw. Symp. (IMS)*, 2014, pp. 1–4.

[14] D. Schindler, B. Schweizer, C. Knill, J. Hasch, and C. Waldschmidt, "MIMO-OFDM Radar Using a Linear Frequency Modulated Carrier to Reduce Sampling Requirements," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 7, pp. 3511–3520, July 2018.

[15] H. Griffiths, P. Knott, and W. Koch, "Christian Hülsmeyer: Invention and Demonstration of Radar, 1904," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 9, pp. 56–60, 2019.

[16] I. C. Society, ***A Brief History of Communications****: IEEE Communications Society - a Fifty-year Foundation for the Future.* The Society, 2002.

[17] N. Wiener, ***Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications***. Technology Press of the Massachusetts Institute of Technology, 1949.

[18] A. Fenn, D. Temme, W. Delaney, and W. Courtney, "The Development of Phased-Array Radar Technology," *Lincoln Lab. J.*, vol. 12, Jan. 2000.

[19] P. Barton, "Digital beam forming for radar," *IEE Proc. F - Commun., Radar Signal Process.*, vol. 127, no. 4, pp. 266–277, 1980.

[20] S. H. Talisa, K. W. O'Haver, T. M. Comberiate, M. D. Sharp, and O. F. Somerlock, "Benefits of Digital Phased Array Radars," *Proc. IEEE*, vol. 104, no. 3, pp. 530–543, 2016.

[21] J. Li and P. Stoica, ***MIMO Radar Signal Processing***. New York: Wiley, Mar. 2008.

[22] Luzhou Xu, Jian Li, and P. Stoica, "Radar imaging via adaptive MIMO techniques," in *Proc. 14th Eur. Signal Process. Conf.*, 2006, pp. 1–5.

[23] M. Kronauge and H. Rohling, "New chirp sequence radar waveform," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2870–2877, 2014.

[24] T. Spreng, U. Prechtel, B. Schönlinner, V. Ziegler, A. Meusling, and U. Siart, "UWB near-field MIMO radar: Calibration, measurements and image reconstruction," in *Proc. 2013 Eur. Radar Conf. (EuRAD)*, 2013, pp. 33–36.

[25] C. Sturm and W. Wiesbeck, "Waveform Design and Signal Processing Aspects for Fusion of Wireless Communications and Radar Sensing," *Proc. IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.

[26] C. Pfeffer, R. Feger, and A. Stelzer, "A stepped-carrier 77-GHz OFDM MIMO radar system with 4 GHz bandwidth," in *Proc. 2015 Eur. Radar Conf. (EuRAD)*, 2015, pp. 97–100.

[27] A. Bourdoux, U. Ahmad, D. Guermandi, S. Brebels, A. Dewilde, and W. Van Thillo, "PMCW waveform and MIMO technique for a 79 GHz CMOS automotive radar," in *Proc. 2016 IEEE Radar Conf. (RadarConf)*, 2016, pp. 1–5.

[28] D. E. Dudgeon, "Fundamentals of digital array processing," *Proc. IEEE*, vol. 65, no. 6, pp. 898–904, 1977.

[29] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[30] P. Stoica, Zhisong Wang, and Jian Li, "Robust Capon beamforming," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 172–175, 2003.

[31] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[32] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, July 1989.

[33] J. Li and R. T. Compton, "Angle and polarization estimation using ESPRIT with a polarization sensitive array," *IEEE Trans. Antennas Propag.*, vol. 39, no. 9, pp. 1376–1383, 1991.

[34] F. C. Robey, S. Coutts, D. Weikle, J. C. McHarg, and K. Cuomo, "MIMO radar theory and experimental results," in *Conf. Rec. 38th Asilomar Conf. Signals, Sys. Comput.*, vol. 1, Nov. 2004, pp. 300–304.

[35] C. Duo-fang, C. Bai-xiao, and Q. Guo-dong, "Angle estimation using ESPRIT in MIMO radar," in *Electron. Lett.*, vol. 44, no. 12, 2008, pp. 770–771.

[36] D. Ciuonzo, G. Romano, and R. Solimene, "Performance Analysis of Time-Reversal MUSIC," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2650–2662, 2015.

[37] F. Belfiori, W. van Rossum, and P. Hoogeboom, "Application of 2D MUSIC algorithm to range-azimuth FMCW radar data," in *Proc. 9th Eur. Radar Conf.*, 2012, pp. 242–245.

[38] D. Cohen and Y. C. Eldar, "Sub-Nyquist Radar Systems: Temporal, Spectral, and Spatial Compression," *IEEE Signal Process. Mag.*, vol. 35, no. 6, pp. 35–58, Nov. 2018.

[39] L. Zhao, L. Wang, L. Yang, A. M. Zoubir, and G. Bi, "The Race to Improve Radar Imagery: An overview of recent progress in statistical sparsity-based techniques," *IEEE Signal Process. Mag.*, vol. 33, no. 6, pp. 85–102, Nov. 2016.

[40] S. Fortunati, R. Grasso, F. Gini, M. Greco, and K. Lepage, "Single-snapshot DOA estimation by using Compressed Sensing," *EURASIP J. Adv. Signal Process. (JASP)*, Nov. 2014.

[41] J. Bock, H. Schafer, K. Aufinger, R. Stengl, S. Boguth, R. Schreiter, M. Rest, H. Knapp, M. Wurzer, W. Perndl, T. Bottner, and T. F. Meister, "SiGe bipolar technology for automotive radar applications," in *Bipolar/BiCMOS Circuits and Technol., Proc. 2004 Meeting*, 2004, pp. 84–87.

[42] A. Hajimiri, H. Hashemi, A. Natarajan, Xiang Guan, and A. Komijani, "Integrated Phased Array Systems in Silicon," *Proc. IEEE*, vol. 93, no. 9, pp. 1637–1655, 2005.

[43] K. Koh and G. M. Rebeiz, "An X- and Ku-Band 8-Element Phased-Array Receiver in 0.18-μm SiGe BiCMOS Technology," *IEEE J. Solid-State Circuits*, vol. 43, no. 6, pp. 1360–1371, 2008.

[44] H. P. Forstner, H. Knapp, H. Jager, E. Kolmhofer, J. Platz, F. Starzer, M. Treml, A. Schinko, G. Birschkus, J. Bock, K. Aufinger, R. Lachner, T. Meister, H. Schafer, D. Lukashevich, S. Boguth, A. Fischer, F. Reininger, L. Maurer, J. Minichshofer, and D. Steinbuch, "A 77GHz 4-channel automotive radar transceiver in SiGe," in *Proc. 2008 IEEE Radio Freq. Integr. Circuits Symp.*, 2008, pp. 233–236.

[45] D. Freundt and B. Lucas, "Long Range Radar Sensor for High-Volume Driver Assistance Systems Market," in *SAE Techn. Paper*. SAE International, Apr. 2008.

[46] C. M. Schmid, R. Feger, C. Wagner, and A. Stelzer, "Design of a linear nonuniform antenna array for a 77-GHz MIMO FMCW radar," in *Proc. 2009 IEEE MTT-S Int. Microw. Workshop Wireless Sens., Local Positioning, RFID*, 2009, pp. 1–4.

[47] Z. Tong, A. Stelzer, and E. Kolmhofer, "77 GHz center-fed differential microstrip antenna array," in *Proc. 5th Eur. Conf. Antennas Propag. (EUCAP)*, 2011, pp. 583–586.

[48] J. Lee, Y. Li, M. Hung, and S. Huang, "A Fully-Integrated 77-GHz FMCW Radar Transceiver in 65-nm CMOS Technology," *IEEE J. of Solid-State Circuits*, vol. 45, no. 12, pp. 2746–2756, 2010.

[49] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, "Radar and Communication Coexistence: An Overview: A Review of Recent Methods," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 85–99, Sept. 2019.

[50] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Sig. Proc. Mag.*, vol. 34, no. 2, pp. 22–35, March 2017.

[51] J. Gamba, ***Radar Signal Processing for Autonomous Driving***, 1st ed. Singapore: Springer Publishing Company Inc., 2019.

[52] L. L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley Pub. Co, 1991.

[53] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian, "The Rise of Radar for Autonomous Vehicles: Signal Processing Solutions and Future Research Directions," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 20–31, 2019.

[54] S. Sun, A. P. Petropulu, and H. V. Poor, "MIMO Radar for Advanced Driver-Assistance Systems and Autonomous Driving: Advantages and Challenges," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 98–117, Jul. 2020.

[55] T. Strohmer and B. Friedlander, "Compressed sensing for MIMO radar - algorithms and performance," in *Conf. Rec. 43rd Asilomar Conf. Signals, Sys. Comput.*, 2009, pp. 464–468.

[56] Y. Yu, A. P. Petropulu, and H. V. Poor, "MIMO Radar Using Compressive Sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 146–163, 2010.

[57] W. Roberts, P. Stoica, J. Li, T. Yardibi, and F. A. Sadjadi, "Iterative Adaptive Approaches to MIMO Radar Imaging," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 5–20, 2010.

[58] G. Hakobyan and B. Yang, "High-Performance Automotive Radar: A Review of Signal Processing Algorithms and Modulation Schemes," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 32–44, Sept. 2019.

[59] V. C. Chen, F. Li, S. . Ho, and H. Wechsler, "Micro-Doppler effect in radar: phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006.

[60] L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, and S. Godsill, "Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking," *Digital Signal Processing*, vol. 25, pp. 1–16, March 2014.

[61] S. Rao, "MIMO Radar," *Texas Instruments - Application Report SWRA554A*, July 2018.

[62] D. Oh and J. Lee, "Low-Complexity Range-Azimuth FMCW Radar Sensor Using Joint Angle and Delay Estimation Without SVD and EVD," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 4799–4811, Sep. 2015.

[63] J.-J. Lin, Y.-P. Li, W.-C. Hsu, and T.-S. Lee, "Design of an FMCW radar baseband signal processing system for automotive application," *SpringerPlus*, vol. 5, no. 42, pp. 1–16, Jan. 2016.

[64] E. Sirignano, A. Davoli, G. M. Vitetta, and F. Viappiani, "A Comparative Analysis of Deterministic Detection and Estimation Techniques for MIMO SFCW Radars," *IEEE Access*, vol. 7, pp. 129 848–129 861, 2019.

[65] P. Di Viesti, A. Davoli, G. Guerzoni, and G. M. Vitetta, "Novel Methods for Approximate Maximum Likelihood Estimation of Multiple Superimposed Undamped Tones and Their Application to Radar Systems," Jul. 2021. [Online]. Available: https://doi.org/10.36227/techrxiv.15054321

[66] A. A. Gorji, R. Tharmarasa, W. D. Blair, and T. Kirubarajan, "Multiple Unresolved Target Localization and Tracking Using Colocated MIMO Radars," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2498–2517, Jul. 2012.

[67] L. Xu, J. Li, and P. Stoica, "Target Detection and Parameter Estimation for MIMO Radar Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 927–939, Jul. 2008.

[68] S. Zuther, M. Biggel, M. M. Muntzinger, and K. Dietmayer, "Multi-Target Tracking for Merged Measurements of Automotive Narrow-Band Radar Sensors," in *2009 12th International IEEE Conference on Intelligent Transportation Systems*, Oct. 2009, pp. 1–6.

[69] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans Ac., Speech and Sig. Proc.*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.

[70] A. Serbes and K. Qaraqe, "A Fast Method for Estimating Frequencies of Multiple Sinusoidals," *IEEE Signal Process. Lett.*, vol. 27, pp. 386–390, 2020.

[71] S. Ye and E. Aboutanios, "An Algorithm for the Parameter Estimation of Multiple Superimposed Exponentials in Noise," in *Proc. of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 3457–3461.

[72] ——, "Rapid Accurate Frequency Estimation of Multiple Resolved Exponentials in Noise," *Signal Processing*, vol. 132, pp. 29–39, Mar. 2017.

[73] A. Serbes, "Fast and Efficient Estimation of Frequencies," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4054–4066, Jun. 2021.

[74] J. Selva, "Efficient Wideband DOA Estimation Through Function Evaluation Techniques," *IEEE Trans. Sig. Proc.*, vol. 66, no. 12, pp. 3112–3123, June 2018.

[75] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. New York, NY: Springer, 2006.

[76] O. Simeone, "A Very Brief Introduction to Machine Learning With Applications to Communication Systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 648–664, 2018.

[77] N. Shlezinger, R. Fu, and Y. C. Eldar, "Deep Soft Interference Cancellation for MIMO Detection," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 8881–8885.

[78] C. M. Schmid, S. Schuster, R. Feger, and A. Stelzer, "On the Effects of Calibration Errors and Mutual Coupling on the Beam Pattern of an Antenna Array," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 8, pp. 4063–4072, Aug. 2013.

[79] C. M. Schmid, C. Pfeffer, R. Feger, and A. Stelzer, "An FMCW MIMO Radar Calibration and Mutual Coupling Compensation Approach," in *2013 European Radar Conference*, Oct. 2013, pp. 13–16.

[80] R. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[81] T.-T. V. Cao, J. Palmer, and P. E. Berry, "False alarm control of cfar algorithms with experimental bistatic radar data," in *2010 IEEE Radar Conference*, 2010, pp. 156–161.

[82] "Imaging radar using cascaded mmwave sensor reference design - tidep-01012." [Online]. Available: https://www.ti.com/tool/TIDEP-01012

[83] "INRAS RadarLog device." [Online]. Available: http://www.inras.at/en/products/radarlog.html

[84] "CamBoard pico flexx." [Online]. Available: https://pmdtec.com/picofamily/

[85] H. Van Trees, ***Optimum Array Processing – Part IV of Detection, Estimation, and Modulation Therory***. New York: Wiley, May 2002.

[86] F. Engels, P. Heidenreich, A. M. Zoubir, F. K. Jondral, and M. Wintermantel, "Advances in Automotive Radar: A Framework on Computationally Efficient High-Resolution Frequency Estimation," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 36–46, Mar. 2017.

[87] S. Saponara, M. S. Greco, and F. Gini, "Radar-on-Chip/in-Package in Autonomous Driving Vehicles and Intelligent Transport Systems: Opportunities and Challenges," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 71–84, Sept. 2019.

[88] J. Le Kernec, F. Fioranelli, C. Ding, H. Zhao, L. Sun, H. Hong, J. Lorandel, and O. Romain, "Radar Signal Processing for Sensing in Assisted Living: The challenges associated with real-time implementation of emerging algorithms," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 29–41, July 2019.

[89] J. Yu and J. Krolik, "MIMO adaptive beamforming for nonseparable multipath clutter mitigation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 2604–2618, 2014.

[90] N. Shlezinger, R. Fu, and Y. C. Eldar, "Deep Soft Interference Cancellation for MIMO Detection," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 8881–8885.

[91] S. Z. Gurbuz and M. G. Amin, "Radar-Based Human-Motion Recognition With Deep Learning: Promising applications for indoor monitoring," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 16–28, July 2019.

[92] J. Lombacher, M. Hahn, J. Dickmann, and C. Wöhler, "Potential of radar for static object classification using deep learning methods," in *Proc. 2016 IEEE MTT-S Int. Conf. Microw. Intell. Mobility (ICMIM)*, May 2016, pp. 1–4.

[93] C. A. Balanis, ***Antenna theory: analysis and design***. Wiley-Interscience, 2005.

[94] J. Selva, "ML Estimation and Detection of Multiple Frequencies Through Periodogram Estimate Refinement," *IEEE Sig. Proc. Lett.*, vol. 24, no. 3, pp. 249–253, 2017.

[95] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[96] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[97] R. E. Schapire, "A Brief Introduction to Boosting," in *Proc. 16th Int. Joint Conf. Artif. Intell. (IJCAI'99)*, vol. 2, 1999, p. 1401–1406.

[98] J. Fürnkranz, "Round Robin Classification," *J. Mach. Learn. Res.*, vol. 2, pp. 721–747, Sept. 2002.

[99] S. Escalera, O. Pujol, and P. Radeva, "On the Decoding Process in Ternary Error-Correcting Output Codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, pp. 120–34, Jan. 2010.

[100] O. Simeone, ***A Brief Introduction to Machine Learning for Engineers***. Now Foundations and Trends, 2018.

[101] I. Jolliffe, ***Principal Component Analysis***, M. Lovric, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[102] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[103] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–44, May 2015.

[104] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.

[105] M. Møller, "A Scaled Conjugate Gradient Algorithm For Fast Supervised Learning," *Neural Networks*, vol. 6, pp. 525–533, Dec. 1993.

[106] M. Gori, ***Machine Learning: A Constraint-Based Approach***, 1st ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2017.

[107] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar," in *Proc. 2017 IEEE Radar Conf.*, May 2017, pp. 1703–1708.

[108] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.

[109] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training Recurrent Neural Networks," *Proc. 30th Int. Conf. Mach. Learn. (ICML 2013)*, Nov. 2012.

[110] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural comput.*, vol. 9, pp. 1735–80, Dec. 1997.

[111] It.mathworks.com, "Phased Array System Toolbox," 2020. [Online]. Available: https://it.mathworks.com/products/phased-array.html

[112] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. Int. Conf. Learn. Representations*, Dec. 2014.

[113] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *Adv. in Neural Inf. Process. Sys.*, vol. 3, June 2014.

[114] S. Abdulatif, K. Armanious, F. Aziz, U. Schneider, and B. Yang, "Towards Adversarial Denoising of Radar Micro-Doppler Signatures," in *Proc. 2019 Int. Radar Conf.*, 2019, pp. 1–6.

[115] B. Vandersmissen, N. Knudde, A. Jalalvand, I. Couckuyt, A. Bourdoux, W. De Neve, and T. Dhaene, "Indoor Person Identification Using a Low-Power FMCW Radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3941–3952, July 2018.

[116] Youngwook Kim, Sungjae Ha, and Jihoon Kwon, "Human Detection Using Doppler Radar Based on Physical Characteristics of Targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 2, pp. 289–293, Feb. 2015.

[117] Y. Li, Z. Peng, R. Pal, and C. Li, "Potential Active Shooter Detection Based on Radar Micro-Doppler and Range-Doppler Analysis Using Artificial Neural Network," *IEEE Sensors J.*, vol. 19, no. 3, pp. 1052–1063, Feb. 2019.

[118] F. Fioranelli, M. Ritchie, and H. Griffiths, "Classification of Unarmed/Armed Personnel Using the NetRAD Multistatic Radar for Micro-Doppler and Singular Value Decomposition Features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1933–1937, Sept. 2015.

[119] P. Molchanov, J. Astola, K. Egiazarian, and A. Totsky, "Ground moving target classification by using DCT coefficients extracted from micro-Doppler radar signatures and artificial neuron network," in *Proc. 2011 Microw., Radar, Remote Sens. Symp.*, Aug. 2011, pp. 173–176.

[120] R. J. Javier and Y. Kim, "Application of Linear Predictive Coding for Human Activity Classification Based on Micro-Doppler Signatures," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1831–1834, Oct. 2014.

[121] C. Clemente, L. Pallotta, A. De Maio, J. J. Soraghan, and A. Farina, "A novel algorithm for radar classification based on doppler characteristics exploiting orthogonal Pseudo-Zernike polynomials," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 417–430, Jan. 2015.

[122] J. Zabalza, C. Clemente, G. Di Caterina, Jinchang Ren, J. J. Soraghan, and S. Marshall, "Robust PCA micro-doppler classification using SVM on embedded systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 3, pp. 2304–2310, July 2014.

[123] G. E. Smith, K. Woodbridge, and C. J. Baker, "Naive Bayesian radar micro-doppler recognition," in *Proc. 2008 Int. Radar Conf.*, Sept. 2008, pp. 111–116.

[124] S. Björklund, T. Johansson, and H. Petersson, "Evaluation of a micro-Doppler classification method on mm-wave data," in *Proc. 2012 IEEE Radar Conf.*, May 2012, pp. 0934–0939.

[125] S. Abdulatif, Q. Wei, F. Aziz, B. Kleiner, and U. Schneider, "Micro-Doppler Based Human-Robot Classification Using Ensemble and Deep Learning Approaches," *Proc. 2018 IEEE Radar Conf. (RadarConf18)*, pp. 1043–1048, Apr. 2018.

[126] Y. Kim and H. Ling, "Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, 2009.

[127] Y. Kim and T. Moon, "Human Detection and Activity Classification Based on Micro-Doppler Signatures Using Deep Convolutional Neural Networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.

[128] M. S. Seyfioğlu, A. M. Özbayoğlu, and S. Z. Gürbüz, "Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1709–1723, Aug. 2018.

[129] S. Abdulatif, F. Aziz, K. Armanious, B. Kleiner, B. Yang, and U. Schneider, "Person Identification and Body Mass Index: A Deep Learning-Based Study on Micro-Dopplers," in *Proc. 2019 IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[130] G. Malysa, D. Wang, L. Netsch, and M. Ali, "Hidden Markov model-based gesture recognition with FMCW radar," in *Proc. 2016 IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 1017–1021.

[131] S. Wang, J. Song, J. Lien, I. Poupyrev, and O. Hilliges, "Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum," in *Proc. 29th Annu. Symp. User Interface Softw. Technol. (UIST '16)*, 2016, pp. 851–860.

[132] Y. Kim and B. Toomajian, "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network," *IEEE Access*, vol. 4, pp. 7125–7130, 2016.

[133] Z. Peng, C. Li, J.-M. Munoz-Ferreras, and R. Gomez-Garcia, "An FMCW radar sensor for human gesture recognition in the presence of multiple targets," in *Proc. 1st IEEE MTT-S Int. Microw. Bio Conf. (IMBIOC)*, May 2017, pp. 1–3.

[134] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic Continuous Hand Gesture Recognition Using FMCW Radar Sensor," *IEEE Sensors J.*, vol. 18, pp. 3278–3289, 2018.

[135] Z. Peng, J.-M. Munoz-Ferreras, R. Gomez-Garcia, and C. Li, "FMCW radar fall detection based on ISAR processing utilizing the properties of RCS, range, and Doppler," in *Proc. 2016 IEEE MTT-S Int. Microw. Symp. (IMS)*, May 2016, pp. 1–3.

[136] B. Jokanović and M. Amin, "Fall detection using deep learning in range-doppler radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 180–189, Feb. 2018.

[137] H. Hong, L. Zhang, C. Gu, Y. Li, G. Zhou, and X. Zhu, "Noncontact Sleep Stage Estimation Using a CW Doppler Radar," *IEEE Trans. Emerg. Sel. Topics in Circuits and Syst.*, vol. 8, no. 2, pp. 260–270, 2018.

[138] J. Saluja, J. Casanova, and J. Lin, "A Supervised Machine Learning Algorithm for Heart-Rate Detection Using Doppler Motion-Sensing Radar," *IEEE J. Electromagn., RF and Microw. in Medicine and Biol.*, vol. 4, no. 1, pp. 45–51, 2020.

[139] C. Gu, J. Wang, and J. Lien, "Deep Neural Network based Body Movement Cancellation for Doppler Radar Vital Sign Detection," in *Proc. 2019 IEEE MTT-S Int. Wireless Symp. (IWS)*, 2019, pp. 1–3.

[140] H. Zhao, H. Hong, D. Miao, Y. Li, H. Zhang, Y. Zhang, C. Li, and X. Zhu, "A Noncontact Breathing Disorder Recognition System Using 2.4-GHz Digital-IF Doppler Radar," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 1, pp. 208–217, 2019.

[141] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. 2nd Int. Conf. Knowl. Discovery, Data Mining*, 1996, p. 226–231.

[142] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Wheel extraction based on micro doppler distribution using high-resolution radar," in *Proc. 2015 IEEE MTT-S Int. Conf. Microw. Intell. Mobility (ICMIM)*, June 2015.

[143] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probability*, vol. 1, 1967, pp. 281–297.

[144] A. Coluccia, A. Fascista, and G. Ricci, "Robust CFAR Radar Detection Using a K-nearest Neighbors Rule," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 4692–4696.

[145] A. El Gonnouni, M. Martinez-Ramon, J. L. Rojo-Alvarez, G. Camps-Valls, A. R. Figueiras-Vidal, and C. G. Christodoulou, "A Support Vector Machine MUSIC Algorithm," *IEEE Trans. Antennas Propag.*, vol. 60, no. 10, pp. 4901–4910, Oct. 2012.

[146] Y. Kim, "Detection of Eye Blinking Using Doppler Sensor With Principal Component Analysis," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 123–126, 2015.

[147] S. Capobianco, L. Facheris, F. Cuccoli, and S. Marinai, "Vehicle Classification Based on Convolutional Networks Applied to FMCW Radar Signals," in *Traffic Mining Appl. to Police Activities - Proc. 1st Italian Conf. Traffic Police (TRAP-2017)*, vol. 728, Oct. 2017, pp. 115–128.

[148] T. Giese, J. Klappstein, J. Dickmann, and C. Wöhler, "Road course estimation using deep learning on radar data," in *Proc. 18th Int. Radar Symp. (IRS 2017)*, June 2017, pp. 1–7.

[149] H. Dbouk, H. Geng, C. M. Vineyard, and N. R. Shanbhag, "Low-Complexity Fixed-Point Convolutional Neural Networks For Automatic Target Recognition," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 1598–1602.

[150] K. Patel, K. Rambach, T. Visentin, D. Rusev, M. Pfeiffer, and B. Yang, "Deep Learning-based Object Classification on Automotive Radar Spectra," in *Proc. 2019 IEEE Radar Conf. (RadarConf)*, July 2019.

[151] J. M. García, D. Zoeke, and M. Vossiek, "MIMO-FMCW Radar-Based Parking Monitoring Application With a Modified Convolutional Neural Network With Spatial Priors," *IEEE Access*, vol. 6, pp. 41 391–41 398, 2018.

[152] B. Major, D. Fontijne, A. Ansari, R. T. Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. Subramanian, "Vehicle Detection With Automotive Radar Using Deep Learning on Range-Azimuth-Doppler Tensors," in *Proc. 2019 IEEE/CVF Int. Conf. Comput. Vision Workshop (ICCVW)*, Oct. 2019, pp. 924–932.

[153] J. Mun, H. Kim, and J. Lee, "A Deep Learning Approach for Automotive Radar Interference Mitigation," in *Proc. 2018 IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, 2018, pp. 1–5.

[154] J. Mun, S. Ha, and J. Lee, "Automotive Radar Signal Interference Mitigation Using RNN with Self Attention," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 3802–3806.

[155] H. Rohling, "Radar CFAR Thresholding in Clutter and Multiple Target Situations," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 608–621, 1983.

[156] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Neural Inf. Process. Syst.*, vol. 25, Jan. 2012.

[157] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. 2009 IEEE Conf. Comput. Vision Pattern Recogn.*, 2009, pp. 248–255.

[158] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[159] Z. Zheng, T. Ruan, Y. Wei, and Y. Yang, "VehicleNet: Learning Robust Feature Representation for Vehicle Re-identification," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recogn. (CVPR)*, June 2019.

[160] M. S. Seyfioğlu and S. Z. Gürbüz, "Deep Neural Network Initialization Methods for Micro-Doppler Classification With Low Training Sample Support," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2462–2466, Dec. 2017.

[161] J. Park, R. Javier, T. Moon, and Y. Kim, "Micro-Doppler Based Classification of Human Aquatic Activities via Transfer Learning of Convolutional Neural Networks," *Sensors*, vol. 16, no. 12, p. 1990, Nov. 2016.

[162] M. S. Seyfioglu, B. Erol, S. Z. Gurbuz, and M. G. Amin, "Diversified radar micro-Doppler simulations as training data for deep residual neural networks," in *Proc. 2018 IEEE Radar Conf. (RadarConf18)*, Apr. 2018, pp. 0612–0617.

[163] R. Girshick, "Fast R-CNN," in *Proc. 2015 IEEE Int. Conf. Comput. Vision (ICCV)*, 2015, pp. 1440–1448.

[164] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, June 2017.

[165] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. 2016 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR)*, 2016, pp. 779–788.

[166] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. 2017 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR)*, 2017, pp. 6517–6525.

[167] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. 2015 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR)*, June 2015, pp. 3431–3440.

[168] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[169] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Med. Image Comput. and Comput.-Assisted Intervention (MICCAI 2015)*, 2015, pp. 234–241.

[170] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. 2017 IEEE Int. Conf. Comput. Vision (ICCV)*, 2017, pp. 2980–2988.

[171] G. Zhang, H. Li, and F. Wenger, "Object Detection and 3d Estimation Via an FMCW Radar Using a Fully Convolutional Network," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 4487–4491.

[172] O. Bialer, D. Shapiro, and A. Jonas, "Object Surface Estimation from Radar Images," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 4132–4136.

[173] S. Gasperini, M. Paschali, C. Hopke, D. Wittmann, and N. Navab, "Signal Clustering With Class-Independent Segmentation," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2020, pp. 3982–3986.

[174] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler, "Semantic Segmentation on Radar Point Clouds," *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, pp. 2179–2186, 2018.

[175] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *Proc. 2017 IEEE Conf. Comput. Vision Pattern Recogn. (CVPR)*, 2017, pp. 77–85.

[176] M. Meyer and G. Kuschk, "Automotive Radar Dataset for Deep Learning Based 3D Object Detection," in *Proc. 16th Eur. Radar Conf. (EuRAD)*, Oct. 2019, pp. 129–132.

[177] M. Meyer and G. Kuschk, "Deep Learning Based 3D Object Detection for Automotive Radar and Camera," in *Proc. 16th Eur. Radar Conf. (EuRAD)*, 2019, pp. 133–136.

[178] A. B. Arrieta, N. Díaz-Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82 – 115, 2020.

[179] M. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Proc. Eur. Conf. Comput. Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., vol. 8689, Nov. 2013.

[180] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," in *Proc. Workshop at Int. Conf. Learn. Representations*, 2014.

[181] "Inras," 2020. [Online]. Available: http://www.inras.at/

[182] J. Zhu, S. Rosset, H. Zou, and T. Hastie, "Multi-class AdaBoost," *Statist. and its interface*, vol. 2, Feb. 2006.

[183] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[184] "Parallax inc." [Online]. Available: https://www.parallax.com/

[185] L. Al Shalabi and Z. Shaaban, "Normalization as a Preprocessing Engine for Data Mining and the Approach of Preference Matrix," *Proc. 2006 Int. Conf. Dependability Comput. Sys.*, pp. 207–214, 2006.

[186] J. Ludikhuize, S. Smorenburg, and Rooij, "Identification of Deteriorating Patients on General Wards; Measurement of Vital Parameters and Potential Effectiveness of the Modified Early Warning Score," *J. Crit. Care*, vol. 27, pp. 424 7–424 13, 2012.

[187] D. Dias and J. Paulo Silva Cunha, "Wearable Health Devices-Vital Sign Monitoring, Systems and Technologies," *Sensors*, vol. 18, no. 8, p. 2414, 2018.

[188] S. Suzuki, T. Matsui, S. Gotoh, Y. Mori, B. Takase, and M. Ishihara, "Development of Non-Contact Monitoring System of Heart Rate Variability (HRV) - an Approach of Remote Sensing for Ubiquitous Technology," in *International Conference on Ergonomics and Health Aspects of Work with Computers*, 2009, pp. 195–203.

[189] C. Gu, C. Li, J. Lin, J. Long, J. Huangfu, and L. Ran, "Instrument-Based Noncontact Doppler Radar Vital Sign Detection System Using Heterodyne Digital Quadrature Demodulation Architecture," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 6, pp. 1580–1588, Jun. 2010.

[190] S. M. M. Islam, F. Fioranelli, and V. M. Lubecke, "Can Radar Remote Life Sensing Technology Help Combat COVID-19?" *Front. Commun. Netw.*, vol. 2, p. 3, 2021.

[191] G. Wang, J.-M. Munoz-Ferreras, C. Gu, C. Li, and R. Gomez-Garcia, "Application of Linear-Frequency-Modulated Continuous-Wave (LFMCW) Radars for Tracking of Vital Signs," *IEEE Trans. Microwave Theory Techn.*, vol. 62, no. 6, pp. 1387–1399, Jun. 2014.

[192] G. Wang, C. Gu, T. Inoue, and C. Li, "A Hybrid FMCW-Interferometry Radar for Indoor Precise Positioning and Versatile Life Activity Monitoring," *IEEE Trans. Microwave Theory Techn.*, vol. 62, no. 11, pp. 2812–2822, Nov. 2014.

[193] J. C. Lin, "Noninvasive Microwave Measurement of Respiration," *Proc. IEEE*, vol. 63, no. 10, pp. 1530–1530, Oct. 1975.

[194] ——, "Microwave sensing of physiological movement and volume change: A review," *Bioelectromagnetics*, vol. 13, no. 6, pp. 557–565, 1992.

[195] T. E. McEwan, "Ultra-Wideband Radar Motion Sensor," Patent 5 361 070, 1994.

[196] M. Varanini, P. Berardi, F. Conforti, M. Micalizzi, D. Neglia, and A. Macerata, "Cardiac and Respiratory Monitoring through Non-Invasive and Contactless Radar Technique," in *2008 Computers in Cardiology*, 2008, pp. 149–152.

[197] E. Greneker, "Radar Sensing of Heartbeat and Respiration at a Distance with Applications of the Technology," in *Radar 97 (Conf. Publ. No. 449)*, 1997, pp. 150–154.

[198] E. Staderini, "UWB Radar in Medicine," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 17, pp. 13–18, Feb. 2002.

[199] C. Li, J. Lin, and Y. Xiao, "Robust Overnight Monitoring of Human Vital Signs by a Non-Contact Respiration and Heartbeat Detector," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2006, pp. 2235–2238.

[200] J.-Y. Kim, J.-H. Park, S.-Y. Jang, and J.-R. Yang, "Peak Detection Algorithm for Vital Sign Detection Using Doppler Radar Sensors," *Sensors*, vol. 19, no. 7, p. 1575, Apr. 2019.

[201] D. Zito, D. Pepe, M. Mincica, F. Zito, A. Tognetti, A. Lanata, and D. De Rossi, "SoC CMOS UWB Pulse Radar Sensor for Contactless Respiratory Rate Monitoring," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 6, pp. 503–510, Dec. 2011.

[202] A. Droitcour, V. Lubecke, Jenshan Lin, and O. Boric-Lubecke, "A Microwave Radio for Doppler Radar Sensing of Vital Signs," in *2001 IEEE MTT-S International Microwave Sympsoium Digest (Cat. No.01CH37157)*, vol. 1, May 2001, pp. 175–178.

[203] C. Gu, R. Li, S. B. Jiang, and C. Li, "A Multi-Radar Wireless System for Respiratory Gating and Accurate Tumor Tracking in Lung Cancer Radiotherapy," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 417–420.

[204] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, "A Review on Recent Advances in Doppler Radar Sensors for Noncontact Healthcare Monitoring," *IEEE Trans Microw. Theory Techn*, vol. 61, no. 5, pp. 2046–2060, May 2013.

[205] C. Gu, "Short-Range Noncontact Sensors for Healthcare and Other Emerging Applications: A Review," *Sensors*, vol. 16, no. 8, p. 1169, Jul. 2016.

[206] M. Kebe, R. Gadhafi, B. Mohammad, M. Sanduleanu, H. Saleh, and M. Al-Qutayri, "Human Vital Signs Detection Methods and Potential Using Radars: A Review," *Sensors*, vol. 20, no. 5, p. 1454, 2020.

[207] E. Cardillo and A. Caddemi, "A review on biomedical mimo radars for vital sign detection and human localization," *Electronics*, vol. 9, no. 9, 2020.

[208] C. Li, Z. Peng, T.-Y. Huang, T. Fan, F.-K. Wang, T.-S. Horng, J.-M. Munoz-Ferreras, R. Gomez-Garcia, L. Ran, and J. Lin, "A Review on Recent Progress of Portable Short-Range Noncontact Microwave Radar Systems," *IEEE Trans. Microwave Theory Techn.*, vol. 65, no. 5, pp. 1692–1706, May 2017.

[209] S. Pisa, E. Pittella, and E. Piuzzi, "A survey of radar systems for medical applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 31, no. 11, pp. 64–81, Nov. 2016.

[210] A. Droitcour, O. Boric-Lubecke, V. Lubecke, J. Lin, and G. Kovacs, "Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for noncontact cardiopulmonary monitoring," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 838–848, Mar. 2004.

[211] S. Schellenberger, K. Shi, F. Michler, F. Lurz, R. Weigel, and A. Koelpin, "Continuous In-Bed Monitoring of Vital Signs Using a Multi Radar Setup for Freely Moving Patients," *Sensors*, vol. 20, no. 20, p. 5827, Oct. 2020.

[212] X. Yang, Y. Yu, H. Qian, X. Zhang, and L. Zhang, "Body orientation and vital sign measurement with ir-uwb radar network," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020, pp. 485–488.

[213] G. Sacco, E. Piuzzi, E. Pittella, and S. Pisa, "An FMCW Radar for Localization and Vital Signs Measurement for Different Chest Orientations," *Sensors*, vol. 20, no. 12, p. 3489, Jun. 2020.

[214] R. O. Bonow, D. L. Mann, D. P. Zipes, and P. Libby, *Braunwald's heart disease e-book: A textbook of cardiovascular medicine.* Elsevier Health Sciences, 2007.

[215] J. Hall, *Guyton and Hall Textbook of Medical Physiology*, 13th ed. Philadelphia, PA: Elsevier, 2016.

[216] M. G. Levitzky, "Pulmonary Physiology," in *Pulmonary Physiology, 7e.* New York: McGraw-Hill Medical, 2007.

[217] S. McGee, "Evidence-Based Physical Diagnosis," in *Evidence-Based Physical Diagnosis*, 4th ed., S. McGee, Ed. Philadelphia: Elsevier LTD, Oxford, 2021, p. iv.

[218] M. Nosrati and N. Tavassolian, "Accurate Doppler Radar-Based Cardiopulmonary Sensing Using Chest-Wall Acceleration," *IEEE J. Electromagn. RF Microw. Med. Biol.*, vol. 3, no. 1, pp. 41–47, 2019.

[219] A. De Groote, M. Wantier, G. Cheron, M. Estenne, and M. Paiva, "Chest Wall Motion during Tidal Breathing," *J. Appl. Physiol.*, vol. 83, no. 5, pp. 1531–1537, 1997.

[220] G. Shafiq and K. C. Veluvolu, "Surface Chest Motion Decomposition for Cardiovascular Monitoring," *Sci. Rep.*, vol. 4, no. 1, p. 5093, May 2014.

[221] A. Albanese, L. Cheng, M. Ursino, and N. W. Chbat, "An Integrated Mathematical Model of the Human Cardiopulmonary System: Model Development," *Am. J. Physiol.-Heart Circ. Physiol.*, vol. 310, no. 7, pp. H899–H921, 2016.

[222] D. Wang, S. Yoo, and S. H. Cho, "Experimental Comparison of IR-UWB Radar and FMCW Radar for Vital Signs," *Sensors*, vol. 20, no. 22, p. 6695, Nov. 2020.

[223] M. Mercuri, Y.-H. Liu, I. Lorato, T. Torfs, A. Bourdoux, and C. Van Hoof, "Frequency-tracking cw doppler radar solving small-angle approximation and null point issues in non-contact vital signs monitoring," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 3, pp. 671–680, 2017.

[224] M. Skolnik, "Role of Radar in Microwaves," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 3, pp. 625–632, 2002.

[225] F.-K. Wang, P.-H. Juan, D.-M. Chian, and C.-K. Wen, "Multiple Range and Vital Sign Detection Based on Single-Conversion Self-Injection-Locked Hybrid Mode Radar With a Novel Frequency Estimation Algorithm," *IEEE Trans. Microwave Theory Techn.*, vol. 68, no. 5, pp. 1908–1920, May 2020.

[226] T. Wu, T. S. Rappaport, and C. M. Collins, "The human body and millimeter-wave wireless communication systems: Interactions and implications," in *2015 IEEE International Conference on Communications (ICC)*, 2015, pp. 2423–2429.

[227] F.-K. Wang, C.-T. M. Wu, T.-S. Horng, C.-H. Tseng, S.-H. Yu, C.-C. Chang, P.-H. Juan, and Y. Yuan, "Review of Self-Injection-Locked Radar Systems for Noncontact Detection of Vital Signs," *IEEE J. Electromagn. RF Microw. Med. Biol.*, vol. 4, no. 4, pp. 294–307, Dec. 2020.

[228] C.-H. Tseng, T.-J. Tseng, and C.-Z. Wu, "Cuffless Blood Pressure Measurement Using a Microwave Near-Field Self-Injection-Locked Wrist Pulse Sensor," *IEEE Trans. Microw. Theory Tech.*, vol. 68, no. 11, pp. 4865–4874, Nov. 2020.

[229] A. Mishra, W. McDonnell, J. Wang, D. Rodriguez, and C. Li, "Intermodulation-Based Nonlinear Smart Health Sensing of Human Vital Signs and Location," *IEEE Access*, vol. 7, pp. 158 284–158 295, 2019.

[230] Y. Yuan, A. Y.-K. Chen, and C.-T. M. Wu, "Super-Regenerative Oscillator-Based High-Sensitivity Radar Architecture for Motion Sensing and Vital Sign Detection," *IEEE Trans. Microwave Theory Techn.*, vol. 69, no. 3, pp. 1974–1984, Mar. 2021.

[231] Y. Yuan and C.-T. M. Wu, "Non-contact Fingertip Microwave Plethysmography Based on Near-Field Sensing with Super-regenerative Oscillator," in *2021 IEEE MTT- Int. Microw. Symp. IMS*, Jun. 2021, pp. 362–365.

[232] B.-K. Park, O. Boric-Lubecke, and V. M. Lubecke, "Arctangent Demodulation with DC Offset Compensation in Quadrature Doppler Radar Receiver Systems," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp. 1073–1079, 2007.

[233] A. Davoli, G. Guerzoni, and G. M. Vitetta, "Machine Learning and Deep Learning Techniques for Colocated MIMO Radars: A Tutorial Overview," *IEEE Access*, vol. 9, pp. 33 704–33 755, 2021.

[234] P. Bello, "Time-Frequency Duality," *IEEE Trans. Inf. Theory*, vol. 10, no. 1, pp. 18–33, Jan. 1964.

[235] P. Di Viesti, A. Davoli, G. Guerzoni, and G. M. Vitetta, "Novel Deterministic Detection and Estimation Algorithms for Colocated Multiple-Input Multiple-Output Radars," *IEEE Access*, vol. 10, pp. 2216–2255, 2022.

[236] K. Naishadham, J. E. Piou, L. Ren, and A. E. Fathy, "Estimation of Cardiopulmonary Parameters from Ultra Wideband Radar Measurements Using the State Space Method," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1037–1046, Dec. 2016.

[237] A. Ahmad, J. C. Roh, D. Wang, and A. Dubey, "Vital Signs Monitoring of Multiple People Using a FMCW Millimeter-Wave Sensor," in *2018 IEEE Radar Conference (RadarConf18)*, Apr. 2018, pp. 1450–1455.

[238] G. Vinci, S. Lindner, F. Barbon, S. Mann, M. Hofmann, A. Duda, R. Weigel, and A. Koelpin, "Six-Port Radar Sensor for Remote Respiration Rate and Heartbeat Vital-Sign Monitoring," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 5, pp. 2093–2100, May 2013.

[239] F.-K. Wang, C.-J. Li, C.-H. Hsiao, T.-S. Horng, J. Lin, K.-C. Peng, J.-K. Jau, J.-Y. Li, and C.-C. Chen, "An injection-locked detector for concurrent spectrum and vital sign sensing," in *2010 IEEE MTT- Int. Microw. Symp.* Anaheim, CA, USA: IEEE, May 2010, pp. 768–771.

[240] Y. Xiao, J. Lin, O. Boric-Lubecke, and M. Lubecke, "Frequency-tuning technique for remote detection of heartbeat and respiration using low-power double-sideband transmission in the ka-band," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 5, pp. 2023–2032, May 2006.

[241] J.-M. Munoz-Ferreras, Z. Peng, R. Gomez-Garcia, G. Wang, C. Gu, and C. Li, "Isolate the Clutter: Pure and Hybrid Linear-Frequency-Modulated Continuous-Wave (LFMCW) Radars for Indoor Applications," *IEEE Microwave*, vol. 16, no. 4, pp. 40–54, May 2015.

[242] I.-S. Lee, J.-H. Park, and J.-R. Yang, "Detrending Technique for Denoising in CW Radar," *Sensors*, vol. 21, no. 19, p. 6376, 2021.

[243] H. Noguchi, H. Kubo, T. Mori, T. Sato, and H. Sanada, "Signal Phase Estimation for Measurement of Respiration Waveform Using a Microwave Doppler Sensor," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 6740–6743.

[244] Changzhan Gu, Ruijiang Li, Hualiang Zhang, A. Y. C. Fung, C. Torres, S. B. Jiang, and Changzhi Li, "Accurate Respiration Measurement Using DC-Coupled

Continuous-Wave Radar Sensor for Motion-Adaptive Cancer Radiotherapy," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3117–3123, Nov. 2012.

[245] J. Wang, X. Wang, L. Chen, J. Huangfu, C. Li, and L. Ran, "Noncontact Distance and Amplitude-Independent Vibration Measurement Based on an Extended DACM Algorithm," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 1, pp. 145–153, Jan. 2014.

[246] W. Xu, Y. Li, C. Gu, and J.-F. Mao, "Large Displacement Motion Interferometry With Modified Differentiate and Cross-Multiply Technique," *IEEE Trans. Microw. Theory Tech.*, vol. 69, no. 11, pp. 4879–4890, Nov. 2021.

[247] Changzhi Li and Jenshan Lin, "Complex Signal Demodulation and Random Body Movement Cancellation Techniques for Non-Contact Vital Sign Detection," in *2008 IEEE MTT-S International Microwave Symposium Digest.* IEEE, 2008, pp. 567–570.

[248] M. Alizadeh, G. Shaker, J. C. M. D. Almeida, P. P. Morita, and S. Safavi-Naeini, "Remote Monitoring of Human Vital Signs Using Mm-Wave FMCW Radar," *IEEE Access*, vol. 7, pp. 54 958–54 968, 2019.

[249] X. Hu and T. Jin, "Short-Range Vital Signs Sensing Based on EEMD and CWT Using IR-UWB Radar," *Sensors*, vol. 16, no. 12, p. 2025, Nov. 2016.

[250] Y. Xu, S. Dai, S. Wu, J. Chen, and G. Fang, "Vital Sign Detection Method Based on Multiple Higher Order Cumulant for Ultrawideband Radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1254–1265, 2012.

[251] T. K. Vodai, K. Oleksak, T. Kvelashvili, F. Foroughian, C. Bauder, P. Theilmann, A. Fathy, and O. Kilic, "Enhancement of Remote Vital Sign Monitoring Detection Accuracy Using Multiple-Input Multiple-Output 77 GHz FMCW Radar," *IEEE J. Electromagn. RF Microw. Med. Biol.*, pp. 1–1, 2021.

[252] Y. Wang, Y. Shui, X. Yang, Z. Li, and W. Wang, "Multi-Target Vital Signs Detection Using Frequency-Modulated Continuous Wave Radar," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, p. 103, Oct. 2021.

[253] I. Walterscheid, O. Biallawons, and P. Berens, "Contactless Respiration and Heartbeat Monitoring of Multiple People Using a 2-D Imaging Radar," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* Berlin, Germany: IEEE, Jul. 2019, pp. 3720–3725.

[254] S. Wang, S. Kueppers, H. Cetinkaya, and R. Herschel, "3D Localization and Vital Sign Detection of Human Subjects with a 120 GHz MIMO Radar," in *2019 20th International Radar Symposium (IRS)*, 2019, pp. 1–6.

[255] F. Liang, F. Qi, Q. An, H. Lv, F. Chen, Z. Li, and J. Wang, "Detection of Multiple Stationary Humans Using UWB MIMO Radar," *Sensors*, vol. 16, no. 11, p. 1922, Nov. 2016.

[256] X. Shang, J. Liu, and J. Li, "Multiple Object Localization and Vital Sign Monitoring Using IR-UWB MIMO Radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 6, pp. 4437–4450, 2020.

[257] M. Nosrati, S. Shahsavari, S. Lee, H. Wang, and N. Tavassolian, "A Concurrent Dual-Beam Phased-Array Doppler Radar Using MIMO Beamforming Techniques for Short-Range Vital-Signs Monitoring," *IEEE Trans. Antennas Propag.*, vol. 67, no. 4, pp. 2390–2404, 2019.

[258] N. T. P. Van, L. Tang, A. Singh, N. D. Minh, S. C. Mukhopadhyay, and S. F. Hasan, "Self-Identification Respiratory Disorder Based on Continuous Wave Radar Sensor System," *IEEE Access*, vol. 7, pp. 40 019–40 026, 2019.

[259] N. Malešević, V. Petrović, M. Belić, C. Antfolk, V. Mihajlović, and M. Janković, "Contactless Real-Time Heartbeat Detection via 24 GHz Continuous-Wave Doppler Radar Using Artificial Neural Networks," *Sensors*, vol. 20, no. 8, p. 2351, 2020.

[260] K. Shi, S. Schellenberger, F. Michler, T. Steigleder, A. Malessa, F. Lurz, C. Ostgathe, R. Weigel, and A. Koelpin, "Automatic Signal Quality Index Determination of Radar-Recorded Heart Sound Signals Using Ensemble Classification," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 773–785, Mar. 2020.

[261] C.-Y. Huang, G.-W. Fang, H.-R. Chuang, and C.-L. Yang, "Clutter-Resistant Vital Sign Detection Using Amplitude-Based Demodulation by EEMD-PCA-Correlation Algorithm for FMCW Radar Systems," in *2019 49th European Microwave Conference (EuMC)*. Paris, France: IEEE, Oct. 2019, pp. 928–931.

[262] S. Wu, T. Sakamoto, K. Oishi, T. Sato, K. Inoue, T. Fukuda, K. Mizutani, and H. Sakai, "Person-Specific Heart Rate Estimation with Ultra-Wideband Radar Using Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 168 484–168 494, 2019.

[263] A. Bin Obadi, P. J. Soh, O. Aldayel, M. H. Al-Doori, M. Mercuri, and D. Schreurs, "A Survey on Vital Signs Detection Using Radar Techniques and Processing with FPGA Implementation," *IEEE Circuits Syst. Mag.*, vol. 21, no. 1, pp. 41–74, 2021.

[264] O. Biallawons and J. Klare, "Person Localization by Detection of Breathing with the MIMO Radar MIRA-CLE Ka," in *EUSAR 2014; 10th European Conference on Synthetic Aperture Radar*, 2014, pp. 1–4.

[265] S. Gabriel, R. W. Lau, and C. Gabriel, "The Dielectric Properties of Biological Tissues: III. Parametric Models for the Dielectric Spectrum of Tissues," *Phys. Med. Biol.*, vol. 41, no. 11, pp. 2271–2293, Nov. 1996.

[266] M. G. Amin, "Radar for Indoor Monitoring: Detection, Classification, and Assessment," M. G. Amin, Ed. CRC Press, 2017.

[267] M. Zhadobov, N. Chahat, R. Sauleau, C. Le Quement, and Y. Le Drean, "Millimeter-Wave Interactions with the Human Body: State of Knowledge and Recent Advances," *Int. J. Microw. Wirel. Technol.*, vol. 3, no. 2, pp. 237–247, 2011.

[268] L. Ren, S. Nahar, A. E. Fathy, T. Phan, N. Tran, and O. Kilic, "Investigation of vital signs monitoring errors due to subject's orientation, clothing and distance from a sfcw radar," in *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*, 2016, pp. 1171–1172.

[269] "Infineon GmbH." [Online]. Available: https://www.infineon.com/

[270] Q. Liang, L. Xu, N. Bao, L. Qi, J. Shi, Y. Yang, and Y. Yao, "Research on Non-Contact Monitoring System for Human Physiological Signal and Body Movement," *Biosensors*, vol. 9, no. 2, 2019.

[271] K. Shi, R. Weigel, A. Koelpin, S. Schellenberger, L. Weber, J. P. Wiedemann, F. Michler, T. Steigleder, A. Malessa, F. Lurz, and C. Ostgathe, "Segmentation of Radar-Recorded Heart Sound Signals Using Bidirectional LSTM Networks," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany: IEEE, Jul. 2019, pp. 6677–6680.

[272] K. Shi, S. Schellenberger, C. Will, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, and A. Koelpin, "A Dataset of Radar-Recorded Heart Sounds and Vital Signs Including Synchronised Reference Sensor Signals," *Sci Data*, vol. 7, no. 1, p. 50, Dec. 2020.

[273] Y. Ding, X. Yu, C. Lei, Y. Sun, X. Xu, and J. Zhang, "A Novel Real-Time Human Heart Rate Estimation Method for Noncontact Vital Sign Radar Detection," *IEEE Access*, vol. 8, pp. 88 689–88 699, 2020.

[274] "Hexoskin Device." [Online]. Available: https://www.hexoskin.com/

[275] M. Fajkus, "Alternative Fiber Optic Sensor Based on Bragg Grating for Heart Rate Monitoring," *Int. J. Biosens. Bioelectron.*, vol. 4, Oct. 2018.

[276] "Shimmer Sensing." [Online]. Available: http://shimmersensing.com/

[277] A. K. Gupta, "Respiration Rate Measurement Based on Impedance Pneumography," 2011.

[278] E. Turppa, J. M. Kortelainen, O. Antropov, and T. Kiuru, "Vital Sign Monitoring Using FMCW Radar in Various Sleeping Scenarios," *Sensors*, vol. 20, no. 22, p. 6505, Nov. 2020.

[279] R. Ichapurapu, S. Jain, M. Kakade, D. Lie, and R. Banister, "A 2.4GHz Non-Contact Biosensor System for Continuous Vital-Signs Monitoring on a Single PCB," in *2009 IEEE 8th International Conference on ASIC*. IEEE, Nov. 2009, pp. 925–928.

[280] M. Li and J. Lin, "Wavelet-Transform-Based Data-Length-Variation Technique for Fast Heart Rate Detection Using 5.8-GHz CW Doppler Radar," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 568–576, 2018.

[281] F. Michler, K. Shi, S. Schellenberger, T. Steigleder, A. Malessa, L. Hameyer, N. Neumann, F. Lurz, C. Ostgathe, R. Weigel, and A. Koelpin, "A Clinically Evaluated Interferometric Continuous-Wave Radar System for the Contactless Measurement of Human Vital Parameters," *Sensors*, vol. 19, no. 11, p. 2492, May 2019.

[282] T.-J. Tseng and C.-H. Tseng, "Noncontact Wrist Pulse Waveform Detection Using 24-GHz Continuous-Wave Radar Sensor for Blood Pressure Estimation," in *2020 IEEEMTT- Int. Microw. Symp. IMS*. Los Angeles, CA, USA: IEEE, Aug. 2020, pp. 647–650.

[283] L. Wen, Y. Gao, C. Gu, and J. Mao, "PhysioChair: A Dual-Frequency Radar System for Noninvasive and Continuous Detection of Physiological Signatures," *IEEE Sens. J.*, vol. 22, no. 8, pp. 8224–8233, Apr. 2022.

[284] S. Suzuki, T. Matsui, H. Kawara, and al, "A Non-Contact Vital Sign Monitoring System for Ambulances Using Dual-Frequency Microwave Radars," *Med. Biol. Eng. Comput.*, vol. 47, no. 1, pp. 101–105, 2009.

[285] J. E. Johnson, O. Shay, C. Kim, and C. Liao, "Wearable Millimeter-Wave Device for Contactless Measurement of Arterial Pulses," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1525–1534, Dec. 2019.

[286] L. Anitori, A. de Jong, and F. Nennie, "FMCW Radar for Life-Sign Detection," in *2009 IEEE Radar Conference*, 2009, pp. 1–6.

[287] E. Antolinos, F. García-Rial, C. Hernández, D. Montesano, J. I. Godino-Llorente, and J. Grajal, "Cardiopulmonary Activity Monitoring Using Millimeter Wave Radars," *Remote Sens.*, vol. 12, no. 14, 2020.

[288] L. Liu and S. Liu, "Remote Detection of Human Vital Sign with Stepped-Frequency Continuous Wave Radar," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 7, no. 3, pp. 775–782, 2014.

[289] W.-C. Su, M.-C. Tang, R. E. Arif, T.-S. Horng, and F.-K. Wang, "Stepped-Frequency Continuous-Wave Radar with Self-Injection-Locking Technology for Monitoring Multiple Human Vital Signs," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 12, pp. 5396–5405, 2019.

[290] Z. Zhang, Y. Nian, J. Chen, and M. He, "An Experimental Study to Optimize the Stepped-Frequency Continuous-Wave Radar Parameters for Noncontact Multi-Target Vital Sign Monitoring," in *2019 IEEE International Conference on Computational Electromagnetics (ICCEM)*, 2019, pp. 1–4.

[291] F. Khan and S. Cho, "A Detailed Algorithm for Vital Sign Monitoring of a Stationary/Non-Stationary Human through IR-UWB Radar," *Sensors*, vol. 17, no. 2, p. 290, Feb. 2017.

[292] H. Shen, C. Xu, Y. Yang, L. Sun, Z. Cai, L. Bai, E. Clancy, and X. Huang, "Respiration and Heartbeat Rates Measurement Based on Autocorrelation Using IR-UWB Radar," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 65, no. 10, pp. 1470–1474, 2018.

[293] J.-Y. Park, Y. Lee, Y.-W. Choi, R. Heo, H.-K. Park, S.-H. Cho, S. H. Cho, and Y.-H. Lim, "Preclinical Evaluation of a Noncontact Simultaneous Monitoring Method for Respiration and Carotid Pulsation Using Impulse-Radio Ultra-Wideband Radar," *Sci Rep*, vol. 9, no. 1, p. 11892, Dec. 2019.

[294] J. D. Kim, W. H. Lee, Y. Lee, H. J. Lee, T. Cha, S. H. Kim, K.-M. Song, Y.-H. Lim, S. H. Cho, S. H. Cho, and H.-K. Park, "Non-Contact Respiration Monitoring Using Impulse Radio Ultrawideband Radar in Neonates," *R. Soc. open sci.*, vol. 6, no. 6, p. 190149, Jun. 2019.

[295] S. Singh, Q. Liang, D. Chen, and L. Sheng, "Sense through wall human detection using UWB radar," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 20, Jun. 2011. [Online]. Available: https://doi.org/10.1186/1687-1499-2011-20

[296] Z. Li, T. Jin, Y. Dai, and Y. Song, "Through-wall multi-subject localization and vital signs monitoring using uwb mimo imaging radar," *Remote Sensing*, vol. 13, no. 15, 2021.

[297] C. Will, K. Shi, S. Schellenberger, T. Steigleder, F. Michler, J. Fuchs, R. Weigel, C. Ostgathe, and A. Koelpin, "Radar-Based Heart Sound Detection," *Sci Rep*, vol. 8, no. 1, p. 11551, Dec. 2018.

[298] Z.-K. Yang, H. Shi, S. Zhao, and X.-D. Huang, "Vital Sign Detection during Large-Scale and Fast Body Movements Based on an Adaptive Noise Cancellation Algorithm Using a Single Doppler Radar Sensor," *Sensors*, vol. 20, no. 15, p. 4183, Jul. 2020.

[299] C. Gu, G. Wang, Y. Li, T. Inoue, and C. Li, "A Hybrid Radar-Camera Sensing System with Phase Compensation for Random Body Movement Cancellation in Doppler Vital Sign Detection," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 12, pp. 4678–4688, 2013.

[300] P. Zhao, C. X. Lu, B. Wang, C. Chen, L. Xie, M. Wang, N. Trigoni, and A. Markham, "Heart rate sensing with a robot mounted mmwave radar," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 2812–2818.

[301] Task Force of the European Society of Cardiology the North A Electrophysiology, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[302] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Health*, vol. 5, p. 258, 2017.

[303] K. C. BILCHICK and R. D. BERGER, "Heart Rate Variability," *Journal of Cardiovascular Electrophysiology*, vol. 17, no. 6, pp. 691–694, 2006.

[304] M. K. Lahiri, P. J. Kannankeril, and J. J. Goldberger, "Assessment of Autonomic Function in Cardiovascular Disease: Physiological Basis and Prognostic Implications," *Journal of the American College of Cardiology*, vol. 51, no. 18, pp. 1725–1733, 2008.

[305] I. Cygankiewicz and W. Zareba, "Chapter 31 - heart rate variability," in *Autonomic Nervous System*, ser. Handbook of Clinical Neurology, R. M. Buijs and D. F. Swaab, Eds. Elsevier, 2013, vol. 117, pp. 379–393.

[306] O. Boric-Lubecke, W. Massagram, V. M. Lubecke, A. Host-Madsen, and B. Jokanovic, "Heart Rate Variability Assessment Using Doppler Radar with Linear Demodulation," in *2008 38th European Microwave Conference*, 2008, pp. 420–423.

[307] K. Shi, T. Steigleder, S. Schellenberger, F. Michler, A. Malessa, F. Lurz, N. Rohleder, C. Ostgathe, R. Weigel, and A. Koelpin, "Contactless Analysis of Heart Rate Variability during Cold Pressor Test Using Radar Interferometry and Bidirectional LSTM Networks," *Sci. Rep.*, vol. 11, no. 1, p. 3025, Feb. 2021.

[308] J. Liu, Y. Chen, Y. Dong, Y. Wang, T. Zhao, and Y.-D. Yao, "Continuous User Verification via Respiratory Biometrics," in *IEEE INFOCOM 2020*, 2020, pp. 1–10.

[309] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "BreathPrint: Breathing Acoustics-Based User Authentication," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 278–291.

[310] H. J. Proctor DF, "Studies of Respiratory Air Flow; Significance of the Normal Pneumotachogram," *Bull Johns Hopkins Hosp.*, vol. 85, no. 4, pp. 253–280, 1949.

[311] G. Benchetrit, S. Shea, T. Dinh, S. Bodocco, P. Baconnier, and A. Guz, "Individuality of Breathing Patterns in Adults Assessed over Time," *Respir. Physiol.*, vol. 75, no. 2, pp. 199–209, 1989.

[312] A. Rahman, V. M. Lubecke, O. Boric-Lubecke, J. H. Prins, and T. Sakamoto, "Doppler Radar Techniques for Accurate Respiration Characterization and Subject Identification," *IEEE J Emerg Sel Top. Circuits Syst*, vol. 8, no. 2, pp. 350–359, Jun. 2018.

[313] S. M. M. Islam, A. Rahman, N. Prasad, O. Boric-Lubecke, and V. M. Lubecke, "Identity Authentication System Using a Support Vector Machine (SVM) on Radar Respiration Measurements," in *2019 93rd ARFTG Microwave Measurement Conference (ARFTG)*, 2019, pp. 1–5.

[314] S. M. M. Islam, A. Sylvester, G. Orpilla, and V. M. Lubecke, "Respiratory Feature Extraction for Radar-Based Continuous Identity Authentication," in *2020 IEEE Radio and Wireless Symposium (RWS)*, 2020, pp. 119–122.

[315] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., ser. Johns Hopkins Studies in the Mathematical Sciences. Baltimore: Johns Hopkins University Press, 1996.