

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

A track-before-detect algorithm for UWB radar sensor networks

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Yan, B., Giorgetti, A., Paolini, E. (2021). A track-before-detect algorithm for UWB radar sensor networks. SIGNAL PROCESSING, 189, 1-19 [10.1016/j.sigpro.2021.108257].

Availability:

This version is available at: <https://hdl.handle.net/11585/843075> since: 2025-01-23

Published:

DOI: <http://doi.org/10.1016/j.sigpro.2021.108257>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

A Track-Before-Detect Algorithm for UWB Radar Sensor Networks

Bo Yan, Andrea Giorgetti, and Enrico Paolini

Abstract

Precise localization and tracking of moving non-collaborative persons and objects using a network of ultra-wideband (UWB) radar nodes has been shown to represent a practical and effective approach. In UWB radar sensor networks (RSNs), existence of strong clutter, weak target echoes, and closely spaced targets are obstacles to achieving a satisfactory tracking performance. Using a track-before-detect (TBD) approach, the waveform obtained by each node during a time period are jointly processed. Both spatial information and temporal relationship between measurements are exploited in generating all possible candidate trajectories and only the best trajectories are selected as the outcome. The effectiveness of the developed TBD technique for UWB RSNs is confirmed by numerical simulations and by two experimental results, both carried out with actual UWB signals. In the first experiment, a human target is tracked by a monostatic radar network with an average localization error of 41.9 cm with no false alarm trajectory in a cluttered outdoor environment. In the second experiment, two targets are detected by multistatic radar network with localization errors of 25.4 cm and 19.7 cm, and detection rate of the two targets is 88.75%, and no false alarm trajectory.

Index Terms

Radar sensor network, track-before-detect, UWB radar, weak target.

This work was supported in part by MIUR under program “Departments of Excellence (2018-2022) – Precise-CPS,” in part by the POR FESR 2014-2020 program under CoACh project, and in part by the Fundamental Research Funds for the Central Universities and China Postdoctoral Science Foundation (No. 2019M663633). Part of this work has been presented at the 2020 IEEE Radar Conference (RadarConf20) [1].

B. Yan is with the School of Aerospace Science and Technology, XIDIAN University, China, and with the Department of Electrical, Electronic, and Information Engineering (DEI), University of Bologna. A. Giorgetti and E. Paolini are with CNIT, the Department of Electrical, Electronic, and Information Engineering (DEI), University of Bologna. Italy (e-mail: {bo.yan, andrea.giorgetti, e.paolini}@unibo.it).

I. INTRODUCTION

In recent years, the ultra-wideband (UWB) wireless technology has gained an increasing importance in a number of civilian and military radar applications [2]–[6]. In an impulse radio (IR)¹ UWB radar sensor network (RSN), sometimes also referred to as an IR UWB wireless sensor radar, radio nodes (or simply sensors), are deployed in the surveillance area to transmit and/or receive nanosecond-duration pulses; human targets or moving objects are then detected, located, and tracked by extracting information from all gathered waveforms and fusing it at a central node. Owing to ultra-wide signal bandwidth, the UWB technology allows achieving very high ranging accuracy and range resolution even in harsh environments affected by dense multipath propagation, such as indoor ones. However, it also poses several challenges.

As a main issue, the transmit power of UWB radio nodes is severely limited by worldwide regulations. Especially when working with targets characterized by a limited radar cross section (RCS), such as human beings, sensors in the network may experience poor signal-to-noise ratio (SNR) conditions even if relatively large pulse integration factors (e.g., in the order of 10^4) are used. In this regime, a target may not generate measurements in all scan periods, leading to a weak target problem. In the framework of a traditional approach, consisting of a detection step in which measurements acquired in one scan are processed to perform detection followed by a tracking step in which target points at different scans are associated to form the trajectories, the weak target issue leads to miss-detection events, inaccurate localization, and even loss of the tracks. In UWB RSNs, miss-detection events arise even when a weak target echo is received only by a subset of the sensors, since at least three sensors are necessary to locate the target and generate one point. It is worth noting that merely lowering the detection threshold to increase the detection rate is an ineffective solution, due to the considerable amount of false alarms arising as a consequence of the low SNR and the harsh propagation environment.

As an additional issue, the ultra-wide signal bandwidth makes the range resolution of UWB sensors in the order of 30 cm, smaller than the linear dimension of a target such as a human being. Therefore, a target may generate multiple measurements as it can in principle be detected in several resolution cells, leading to an extended target problem [5], [7]. Although a number of tracking methods have been developed to track extended targets, including methods based on the probability hypothesis density (PHD) filter [8]–[12] and methods based on the Bernoulli filter [13], it is known that, without a sufficient amount of a priori information about the target, the presence of multiple measurements deteriorates tracking performance. We also point out that targets in UWB RSNs are often human beings, whose motion may

¹Although the scope of UWB wireless technology has been extended beyond the IR technique, consisting of the transmission of sequences of very short duration pulses, in this paper the term UWB always refers to IR UWB.

be characterized by some degree of maneuverability, with course and speed changing within one or two scans.

Track-before-detect (TBD) is a well-known approach to cope with the weak target issue. Echos of weak targets are accumulated through multiple scans to reach the detection threshold. As such, rather than declaring the presence of targets relying on the measurements collected in a single scan, measurements received in multiple scans are jointly processed, keeping record of a number of candidate trajectories, and confirming only a subset of them (hence detecting the associated targets). Several TBD approaches are available in the literature (e.g., [12]–[41]), each tailored to specific tracking problems. For example, in the TBD framework, the issue of extended target has been addressed through point voting in Hough transform [20], [30] or particle filter based probability density estimation of extended target state [24]. The existing TBD implementations can be categorized into four classes, summarized hereafter.

Particle filter (PF) based TBD [22]–[24], [34], [35]: Particle filters can sequentially approximate the a posteriori probability density function with any dynamic and measurement models, making PF-TBD superior to other methods in the presence of nonlinear or non-Gaussian noise. Some PHD filters [11], [12] and Bernoulli filters [13] within TBD strategies are also implemented as particle filters.

Dynamic programming based TBD [25]–[27], [38]: DP-TBD is a grid-based method that estimates target trajectories by searching all physically admissible paths in a discrete state space. Some grid-based TBD techniques [39]–[41] perform target detection via sliding time window and multi-frame tests.

Hough transform based TBD [28]–[30]: The points on a straight line collapse into a single point in the transformed domain. So Hough transform-based TBD is effective in detecting the points of the trajectory on a straight-line.

Optimization based TBD [31]–[33]: Using optimization algorithms such as the Grey Wolf optimizer or greedy algorithms, it is possible to search the optimal association of points in the solution space efficiently.

In the above-mentioned methods (e.g., [8]–[35]) the input is typically represented by points featuring position information and a timestamp. Since, as previously remarked, in UWB RSNs target echoes can be very weak (sometimes lower than waveform detection threshold) and the propagation environment can be harsh (especially in indoor), the tracking system must cope with frequent miss-detection events and clutter points even when state-of-the-art point generation techniques are employed. For this reason, direct application of the above-mentioned TBD methods to the UWB RSNs framework, even in conjunction with UWB-tailored waveform detection and point generation techniques, such as the one proposed in [5], based on in-sensor detection, range estimation and spatial relationship between targets and sensors, tends to be ineffective since the TBD performance is degraded by imperfect point generation. Therefore,

in contrast with current TBD approaches, in this paper we design a new TBD method which, rather than focusing directly on the target points, organizes points into tracklets (short segments, corresponding to a number of subsequent scan periods) and seeks to detect target tracklets by processing waveforms incoming from multiple sensors through multiple scans. As such, it builds a deeper fusion framework which includes point generation plus track detection. Although the proposed approach does not strictly belong to any of the TBD categories mentioned above, it shares some commonalities with DP-TBD since in some steps it applies a grid cell strategy. However, rather than building a large number of track candidate represented by grid cells, grid cells of multiple scans are jointly processed by image processing methods to obtain the most promising tracks directly.

In the proposed method, soft waveform samples from the several sensors are jointly processed by a voting algorithm to obtain an image of surveillance area.² Multiple images, constructed through subsequent scan periods, are then stacked and jointly processed by applying efficient image processing techniques on the obtained three-dimensional (3D) structure. Although the proposed processing of stacked images, coherently with the TBD philosophy, is mainly motivated by the need to enhance detection and tracking of weak targets, it turns out to be useful in reducing clutter and coping with extended targets. Trajectory detection is performed by first generating tracklets and then by associating them, an approach helpful in the presence of maneuvering targets.

The main innovative contributions of the RSN implementation are summarized as follows:

- 1) Design of an efficient target imaging method using multi-sensor UWB waveforms; development of a target region partition method to refine the target image. The two strategies together are effective in making each target represented by a region where it exists, which avoids extended target issues.
- 2) Detection rate enhancement by jointly processing information from multiple sensors and multiple scans, rather than confirm a detection at single sensor or single scan level. Possible target tracks can be obtained by waveform directly, this strategy addresses issue of weak target detection.

Although the proposed technique is not tailored to a specific type of target, we present results obtained with human targets, showing the potential of the proposed approach in intrusion detection applications.

The paper is organized as follows. A system overview is provided in Section II. The proposed processing is described in Section III, while numerical and experimental results are presented in Section IV. Computer simulation results are first presented in Section IV-A. Then, two cases studies with actual UWB waveforms, both focused on detection and tracking of human targets, are presented in Section IV-B and

²A different approach consists of performing a pre-processing of the UWB waveform by a constant false alarm (CFAR) detector [5]. However, according to the TBD approach, we avoid any such “waveform detection” step to exploit even weak target echoes.

Section IV-C. In particular, Section IV-B addresses the case of a RSN composed of monostatic UWB sensors, while Section IV-C the case of a multistatic UWB RSN. Conclusions are drawn in Section V. A subset of the results presented in the paper appeared in its conference version [1]. With respect to [1], the proposed processing techniques are here addressed in a more thorough way, providing all of the details. Moreover, the multistatic UWB RSN case has been added and a number of additional results are presented.

II. SYSTEM OVERVIEW AND NOTATION

In this work, we consider a UWB RSN aimed at detecting, locating, and tracking moving objects in the surveillance area. The RSN may be either monostatic (i.e., composed of $N_R \geq 3$ UWB sensors each configured as a monostatic radar), or multi-static (i.e., composed of one UWB transmitter and $N_R \geq 3$ UWB sensors configured in a receive-only mode³). In both network configurations, all nodes are placed in known positions and are connected to a fusion center (FC).

Fig. 1 shows a block diagram summarizing the whole processing performed by the RSN. Each yellow box represents a single sensor, and the light red blocks in it are the processing steps performed by the sensor. The data preprocessed by sensors are transferred to a FC, represented by the light blue block, performing tracking, detection, and localization according to a TBD approach. For convenience, these blocks are labeled with the subsection in which the corresponding processing is described. In the following, we overview the main processing steps performed by the RSN.

In every scan period t , each sensor performs filtering, sampling, pulse integration, non-coherent demodulation, and clutter removal; the resulting signals \mathbf{m}_n^t , $n = 1, \dots, N_R$, are delivered to the FC.⁴ Such signals are expected to contain the (possibly weak) target echoes plus a clutter residue due to imperfect waveform alignment, non-static clutter, and ghost artifacts [5].

Within the FC, the first stage consists of a voting process, yielding a two-dimensional (2D) *score map* of the surveillance area, \mathbf{S}^t , at scan period t . In the process, residual clutter is suppressed. The second stage introduces the temporal dimension in the processing. A 3D data structure \mathbf{M}^t , built up by stacking the 2D score maps, is processed in a sliding-window fashion with time window size W , representing the number of scans jointly handled by the TBD algorithm. This processing is aimed at identifying sets of grid cells (dubbed *regions*), resembling 3D trajectories that potentially correspond to targets. The image processing techniques named region growing and opening operation, both performed in a 3D setting, are

³The transmitter may also be configured as a monostatic radar, in which case it shall be included in the set of sensors.

⁴Note that, with a slight abuse of notation, the time in the received signals $r_n(t)$ is also indicated with t in Fig. 1.

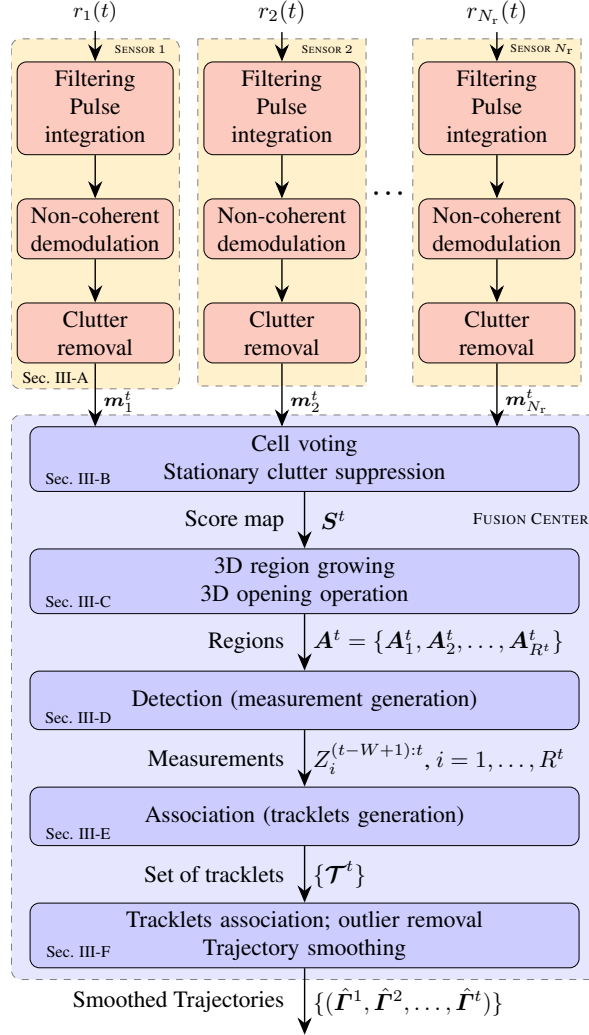


Fig. 1. Block diagram of the whole processing chain.

applied to this purpose. The number of obtained regions is R^t , and the generic region is denoted by A_i^t , $i = 1, \dots, R^t$.

The subsequent stage aims at extracting a set of *measurements* (also referred to as *points*) $Z_i^{(t-W+1):t}$ out of each region A_i^t , where each measurement carries information about position, scan index, and score. Next, tracklet detection is performed, in which points with scan indexes belonging to a time window of size W are associated to form *tracklets* \mathcal{T} , i.e., fragments of potential trajectories. The output of this stage is the set of detected tracklets. Tracklets are then associated with each other and combined to form trajectories. The generic trajectory point at scan period t (i.e., a point considered to be a detection) is $\mathbf{I}^t = (x^t, y^t, t)$. More accurate trajectories can be obtained by outliers removal and application of a

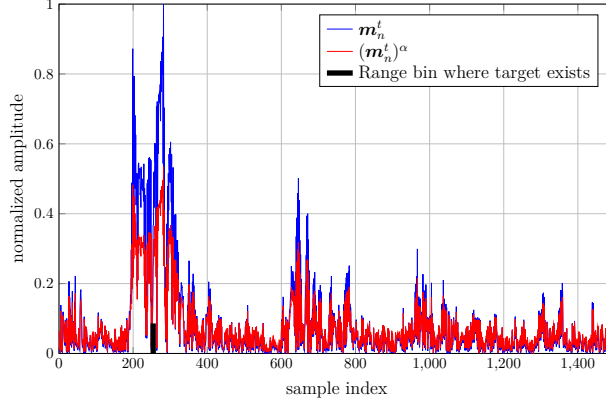


Fig. 2. Example of signal m_n^t along with the corresponding peak-suppressed samples $(m_n^t)^\alpha$ with $\alpha = 0.75$.

smoothing filter. The smoothed target trajectories represent the RSN output. The point corresponding to \mathbf{I}^t after outlier removal and trajectory smoothing is denoted by $\hat{\mathbf{I}}^t = (\hat{x}^t, \hat{y}^t, t)$.

The proposed processing framework deviates significantly from the one that is usually performed in RSNs. Conventional RSN processing features a CFAR detection step at single sensor level followed by a point detection step at FC level. Target tracking methods (e.g., [8]–[23], [25]–[35]), fed with the detected points, can then be applied to generate the target tracks. In UWB RSNs, however, this approach suffers from the fact that weak target echo components may be filter out throughout CFAR detection in any sensor or in point generation at the FC. Even though powerful tracking method are used to process the detected points, methods compliant with the traditional framework lead to shortage in terms of weak target detection, when no target points are generated. A comparison between the proposed approach and some more conventional methods will be described in Section IV-A.

III. PROCESSING

A. In-Sensor Processing

At the n -th sensor, the received signal $r_n(t)$ is first processed by a bandpass filter to remove out-of-band noise spectral components and then sampled. Next, pulse integration is performed, where N_s pulses received within a scan are coherently averaged, yielding a gain in terms of SNR. The resulting vector in scan period t is indicated by \mathbf{r}_n^t and has length N_c , the number of samples per scan. Noncoherent demodulation, consisting of an envelope detector (implemented by a square-law device followed by a low-pass filter) is then applied to \mathbf{r}_n^t yielding the signal ϵ_n^t [5].

The waveform ϵ_n^t contains the echoes of both moving targets and non-static or static objects, together with noise. Echoes from static objects can be canceled out effectively by applying a clutter removal

algorithm. In UWB RSNs, the moving target indication (MTI) clutter removal technique turns out to be effective, especially when applied after noncoherent demodulation, owing to the possibility of performing very accurate time-of-arrival (ToA) estimation and therefore to accurately align the waveforms ϵ_n^t at consecutive scans [5]. MTI is implemented by calculating a reference waveform $\epsilon_n^{t,\text{av}}$ by averaging the waveforms received within a time window, of 2κ scans, preceding the current scan, i.e.,

$$\epsilon_n^{t,\text{av}} = \frac{1}{2\kappa} \sum_{p=1}^{2\kappa} \epsilon_n^{t-\kappa-p}. \quad (1)$$

Then, clutter removal is performed as $\mathbf{m}_n^t = \|\epsilon_n^t - \epsilon_n^{t,\text{av}}\|$, where $\|\cdot\|$ is the element-wise absolute value operator. The signal \mathbf{m}_n^t is expected to contain target echoes along with noise and non-static clutter residue. An example of real waveform \mathbf{m}_n^t collected in the measurement campaign described in Section IV-B, is presented in Fig. 2.

B. Voting for the Surveillance Area

The surveillance area is divided into $N_x \times N_y$ grid cells, of size $\Delta_x \times \Delta_y$. We use the notation $(i_x, i_y) \in \{1, \dots, N_x\} \times \{1, \dots, N_y\}$ to denote the pair of indexes identifying a cell in the score map. Moreover, we use (x, y) to denote the actual position of a point in the map. The position of the center of cell (i_x, i_y) is $(x_c(i_x, i_y), y_c(i_x, i_y)) = ((i_x - 1/2)\Delta_x, (i_y - 1/2)\Delta_y)$.

Processing at the FC starts from the vectors \mathbf{m}_n^t , $n = 1, \dots, N_R$, collected from the N_R sensors. The j -th element of \mathbf{m}_n^t is denoted by $m_n^t(j)$, $j = 1, \dots, N_c$. Each sample in \mathbf{m}_n^t votes for a subset of the grid cells, as follows. We denote by $\tau_n(j)$ the ToA corresponding to sample $m_n^t(j)$ (a multiple of the sampling time T_c), defined as the delay with respect to the time of transmission.

$$\tau_n(j) = \begin{cases} \tau_n^{\text{m}}(j), & \text{monostatic RSN} \\ \tau_n^{\text{b}}(j), & \text{multistatic RSN.} \end{cases}$$

The $\tau_n^{\text{m}}(j)$ and $\tau_n^{\text{b}}(j)$ means time delay in monostatic RSN and multistatic RSN respectively.

In a monostatic RSN, each sensor synchronizes on the time of transmission of the UWB pulse. In this case, the sample $m_n^t(j)$ votes for all cells that are crossed by the circle of equation

$$\sqrt{(x - x_n)^2 + (y - y_n)^2} = \frac{c \tau_n^{\text{m}}(j)}{2}$$

where (x_n, y_n) is the position of sensor n and c is the speed of light, i.e., the circle centered in the sensor and having radius $c \tau_n^{\text{m}}(j)/2$.

In a multistatic RSN, each sensor synchronizes on the ToA of the UWB pulse received along the direct path from the transmitter to the sensor; $\tau_n^b(j)$ means the excess delay with respect to the ToA of the direct pulse. In this case, denoting by (x_T, y_T) the position of the transmitter and by L_n the distance between the transmitter and sensor n , the sample $m_n^t(j)$ votes for all cells that are crossed by an ellipse with foci in (x_n, y_n) and (x_T, y_T) and major axis $L_n + c\tau_n^b(j)$, namely, the ellipse of equation

$$\sqrt{(x - x_T)^2 + (y - y_T)^2} + \sqrt{(x - x_n)^2 + (y - y_n)^2} = L_n + c\tau_n^b(j).$$

The voting algorithm starts by defining the amplitude of the circle or ellipse associated with $m_n^t(j)$ as

$$V_n^t(j) = \frac{(m_n^t(j))^\alpha}{\frac{1}{N_c} \sum_{l=1}^{N_c} (m_n^t(l))^\alpha} \quad (2)$$

where $0 < \alpha < 1$. A total of N_c circle or ellipse amplitudes are thus obtained from each vector \mathbf{m}_n^t .

The score (or *vote*) received by each of the $N_x \times N_y$ grid cells is defined as

$$S_n^t(i_x, i_y) = \max_{j \in \{1, \dots, N_c\}} \{V_n^t(j) I_n(j, i_x, i_y)\} \quad (3)$$

where the indicator function $I_n(j, i_x, i_y)$ equals 1 if the j -th circle or ellipse passes through the cell (i_x, i_y) and 0 otherwise. Depending on the sampling time T_c and on the cell dimension, one grid cell may be voted by several circles or ellipses; according to (3), the cell score is the largest received vote. At the end of this voting process, we obtain $\mathbf{S}_n^t = \{S_n^t(i_x, i_y)\}_{1 \leq i_x \leq N_x, 1 \leq i_y \leq N_y}$ as the score map associated with \mathbf{m}_n^t . The exponent α applied in (2) is used to mitigate strong clutter residues, avoiding excessively large cell scores from clutter.

The score maps constructed from all vectors \mathbf{m}_n^t , $n = 1, \dots, N_R$, are first multiplied to obtain an overall score map Σ^t , namely,

$$\Sigma^t = \left\{ \Sigma^t(i_x, i_y) = \prod_{n=1}^{N_R} S_n^t(i_x, i_y) \right\}_{1 \leq i_x \leq N_x, 1 \leq i_y \leq N_y}. \quad (4)$$

Then, score values that are below a threshold are forced to zero, yielding the final score map at scan t ,

$\mathbf{S}^t = \{S^t(i_x, i_y)\}_{1 \leq i_x \leq N_x, 1 \leq i_y \leq N_y}$ where

$$S^t(i_x, i_y) = \begin{cases} \Sigma^t(i_x, i_y), & \text{if } \Sigma^t(i_x, i_y) > \eta_{\text{score}}^t \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

The score threshold in (5) is adaptive and proportional to the average score map. Specifically, we have

$$\eta_{\text{score}}^t = \frac{\beta}{N_x N_y} \sum_{(i_x, i_y)} \Sigma^t(i_x, i_y) \quad (6)$$

for some positive β . As a main feature of the proposed approach, the threshold η_{score}^t may be kept relatively low (compared with track-after-detect approaches), which is very helpful in detecting weak targets. The high number of generated false alarms are then dealt with in the subsequent stages of the processing, in which the temporal dimension is introduced. In some scenarios (especially some indoor ones) a clutter removal step may also be applied to the score maps generated by (5). The clutter removal procedure is described in Appendix A.

C. 3D Region Growing and 3D Opening Operation

The scan maps \mathbf{S}^t generated through the voting process are stacked one over another, yielding a 3D data structure $\mathbf{M}^t = [M^t(i_x, i_y, k)]$, $1 \leq i_x \leq N_x$, $1 \leq i_y \leq N_y$, $1 \leq k \leq w$, where $M^t(i_x, i_y, k) = S^k(i_x, i_y)$. Hereafter, we denote by $\mathbf{M}^{k_1:k_2}$ the 3D structure composed of all layers of \mathbf{M}^t between scan period k_1 and scan period k_2 , with k_1 and k_2 included. We refer to each element of \mathbf{M}^t as a *cell* and to $M^t(i_x, i_y, k)$ as the score of cell (i_x, i_y, k) . Moreover, we say that two cells (i_x, i_y, k) and (j_x, j_y, h) are neighboring cells when all of the following conditions are true: $\|i_x - j_x\| \leq 1$, $\|i_y - j_y\| \leq 1$, $\|k - h\| \leq 1$. As such, every cell has at most 26 neighbors.

The data structure \mathbf{M}^t is processed in a sliding window fashion with time window size W . In particular, the top (i.e., most recent) W layers of \mathbf{M}^t , $\mathbf{M}^{(t-W+1):t}$, are processed by applying two operations borrowed from digital image processing, namely, region growing and opening operation [42]. These two operations are applied on $\mathbf{M}^{(t-W+1):t}$ every s scan periods, for some $1 \leq s \leq W$, i.e., at scan periods $t = W + q \cdot s$, $q \in \mathbb{N}$. For example, if $W = 4$ and $s = 2$ then region growing and open operation are applied on $\mathbf{M}^{1:4}$, $\mathbf{M}^{3:6}$, $\mathbf{M}^{5:8}$, and so on.⁵ Region growing and opening operation are described in the following. More details (e.g., the pseudocode of 3D region growing) are provided in Appendix B.

Region growing consists of grouping cells to form cell clusters called *regions*. Starting from a set of “seed” cells, regions are progressively grown by including in each of them new neighboring pixels that meet certain criteria. Region growing is performed in a 3D fashion, as described hereafter.

In the beginning, a number of cells in $\mathbf{M}^{(t-W+1):t}$ are declared as seed cells. Seed cells are evenly selected and their set is $\mathcal{C}_{\text{seed}}^{(t-W+1):t} = \{(ad_x, bd_y, t - W + cd_t)\}$, where d_x , d_y , and d_t are given positive

⁵If $s = W$ then every layer of \mathbf{M}^t is processed once. If instead $1 \leq s < W$ then the time windows are partially overlapped and every layer of \mathbf{M}^t is processed multiple times.

integers and $a = 1, \dots, \lfloor N_x/d_x \rfloor$, $b = 1, \dots, \lfloor N_y/d_y \rfloor$, $c = 1, \dots, \lfloor W/d_t \rfloor$. At the beginning, all seed cells are declared to be active. Seed cells are then processed in order.⁶ If a seed cell is active and its score is nonzero, a new region \mathcal{R} is initialized, having the seed cell as its sole elements. All neighbors of the seed cell are then considered and any of them having a nonzero score is included in \mathcal{R} . Next, all newly included cells are processed. The neighbors of each of them are considered and, if any such neighbor is not yet in \mathcal{R} and has a nonzero score, it is included in \mathcal{R} . This processing is iterated until no new cell can be included in the region. Importantly, if during some iteration the region grown from a seed cell includes another seed cell, this latter cell is declared to be inactive. This way, no new region will be initialized for it.

Not all of the built regions are further processed. In order for a region to become a confirmed one, the two conditions

$$\sum_{(i_x, i_y, k) \in \mathcal{R}} M^t(i_x, i_y, k) \geq \gamma_{\text{score}}^t \quad (7)$$

$$|\mathcal{R}| \geq \gamma_{\text{num}}^t \quad (8)$$

must be simultaneously fulfilled, where $|\cdot|$ denotes cardinality. This means that, to regard a region as potentially corresponding to a target, we require that it possesses a minimum score and a minimum cardinality. For all regions not fulfilling both (7) and (8), the score of all cells is forced to zero ($M^t(i_x, i_y, k) = 0$). Moreover, the score of any cell not included in any of the obtained region is also forced to zero, which is effective in removing isolated cells due to noise and clutter. Investigating the optimum values of the two threshold γ_{score}^t and γ_{num}^t analytically turns to be a hard task, as they depend on several factors that include the target size, radar cross-section, radar parameters, measurement noise, and grid cell area. In practice, suitable values of these thresholds can be found experimentally. The 3D region growing procedure is formalized in Appendix B.

At the end of 3D region growing, we obtain a set $\{\mathcal{R}_1^t, \mathcal{R}_2^t, \dots, \mathcal{R}_{R^t}^t\}$ of R^t regions. Note that R^t may be considered as a preliminary estimate of the number of targets at scan t , although no detection is performed at this stage. The W 2D layers of any region \mathcal{R}_i^t are usually irregular, especially due to the effect of clutter. Hence, a 3D opening operation [43]–[45] is performed to smooth the contour of \mathcal{R}_i^t , break narrow isthmuses, and eliminate thin protrusions. Opening operation is effective in improving the localization accuracy of targets corresponding to regions for which a detection is declared at the end of the whole processing, as well as in reducing false alarms.

⁶The order in which seed cells are processed is irrelevant, as the final set of generated regions is the same irrespective of the seed cell ordering.

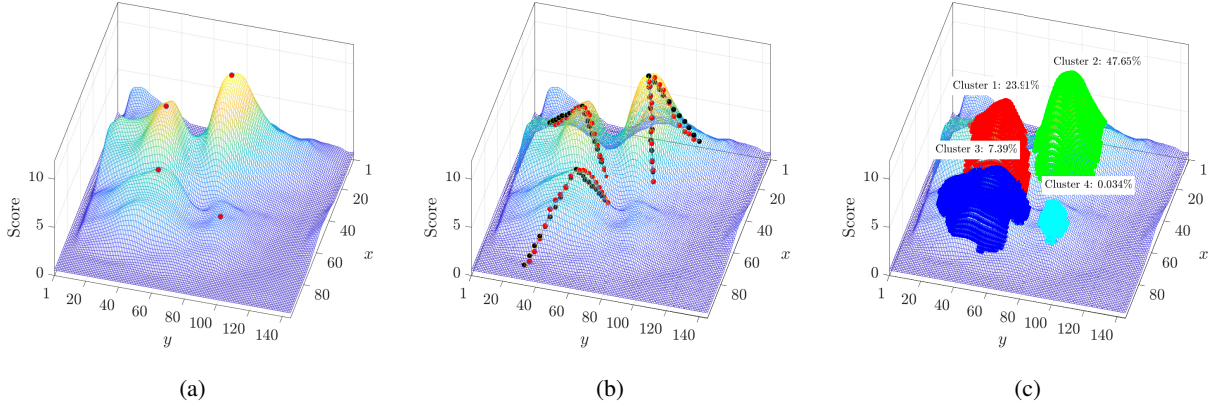


Fig. 3. (a) Example of surface S' and centers of local maximum cells. (b) Example of application of the criterion to defined clusters ($N = 10$). Black: points Q_h . Red: points \hat{Q}_h . (c) Portions of S' corresponding to the four clusters.

The opening of region \mathcal{R}_i^t by a structuring element (or kernel) B returns a new region A_i^t defined as

$$A_i^t = (\mathcal{R}_i^t \ominus B) \oplus B$$

where $\mathcal{R}_i^t \ominus B$ denotes erosion of \mathcal{R}_i^t by the structuring element B and $(\mathcal{R}_i^t \ominus B) \oplus B$ denotes dilation of $(\mathcal{R}_i^t \ominus B)$, again by B . The erosion and dilation operations (in a 3D setting) are defined as follows.

Let F be a 3D region and B be a 3D structuring element, one cell of which is regarded as the origin of B . Then, $F \ominus B$ is the set of all cells (i_x, i_y, k) such that the structural element with origin in (i_x, i_y, k) is included in F . Formally:

$$F \ominus B = \{(i_x, i_y, k) | B_{i_x, i_y, k} \subseteq F\}$$

where $B_{i_x, i_y, k}$ is the structural element with origin in (i_x, i_y, k) . Moreover, $F \oplus B$ is the set of all cells (i_x, i_y, k) such that the structural element with origin in (i_x, i_y, k) has a nonzero intersection with F . Formally:

$$F \oplus B = \{(i_x, i_y, k) | B_{i_x, i_y, k} \cap F \neq \emptyset\}.$$

The employed 3D structuring element is composed of 7 cells, with a central cell playing the role of the origin and the 6 cells adjacent to its 6 facets (regarding a 3D cells as a parallelepiped). After application of opening operation to each region \mathcal{R}_i^t , we obtain a new region set $\{A_1^t, A_2^t, \dots, A_{R^t}^t\}$ that is further processed for generation of points.

D. Points Generation

We aim at generating a set of points, also referred to as *measurements*, out of the R^t regions constructed via region growing and opening operation. The starting point is a 3D data structure whose k -th layer, $k = t - W + 1, \dots, t$, is a 2D score map denoted by $\hat{m}^k(i_x, i_y)$. Each 2D score map $\hat{m}^k(i_x, i_y)$ is processed separately. Note that this score map is contributed by layer k of all 3D regions A_i^t identified in the previous step.

A 2D image smoothing operation is first performed on $\hat{m}^k(i_x, i_y)$, yielding a smoothed score map $m^k(i_x, i_y)$. Specifically, we perform

$$m^k(i_x, i_y) = \frac{1}{(2w + 1)^2} \sum_{p=-w}^w \sum_{q=-w}^w \hat{m}^k(i_x + p, i_y + q) \quad (9)$$

where w is the width of the smoothing window. Next, we search for cells being local maxima in $m^k(i_x, i_y)$. A cell is said to be a local maximum when its score is larger than the score of all of its 8 nearest neighbors in the 2D layer.

Once all local maxima in $m^k(i_x, i_y)$ have been identified, we grow a 2D region (hereafter referred to as a *cluster* associated with (i_x, i_y)) around each of them, according to the following strategy. Let $(i_{x,1}, i_{y,1})$ and $(i_{x,2}, i_{y,2})$ be a local maximum cell and a cell under test, respectively. Let $(x_{c,1}, y_{c,1})$ and $(x_{c,2}, y_{c,2})$ be the positions of the centers of cells $(i_{x,1}, i_{y,1})$ and $(i_{x,2}, i_{y,2})$, respectively. Moreover, let $m_1^k = m^k(i_{x,1}, i_{y,1})$ and $m_2^k = m^k(i_{x,2}, i_{y,2})$.

Consider the two points $P_1 = (x_{c,1}, y_{c,1}, m_1^k)$ and $P_2 = (x_{c,2}, y_{c,2}, m_2^k)$, draw the 3D segment having P_1 and P_2 as endpoints, and find $N - 1$ equally-spaced points on this segment. Letting the h -th such point be $Q_h = (\chi_h, \psi_h, c_h^k)$, $h = 1, \dots, N - 1$, we have

$$(\chi_h, \psi_h, c_h^k) = \frac{N - h}{N} (x_{c,1}, y_{c,1}, m_1^k) + \frac{h}{N} (x_{c,2}, y_{c,2}, m_2^k). \quad (10)$$

The smoothed score map may be represented as a 3D surface \mathcal{S} . Note that the z -axis value of points in \mathcal{S} is cell-wise constant. A second surface \mathcal{S}' can be obtained from \mathcal{S} through a 2D interpolation operation. Each point in \mathcal{S}' is in the form $(x, y, f^k(x, y))$, where $f^k(x, y)$ is constructed as follows. Let (a_1, b_1) , (a_1, b_2) , (a_2, b_1) , and (a_2, b_2) be the coordinates of the centers of the four cells closest to (x, y) , where $a_1 < x < a_2$ and $b_1 < y < b_2$, and let m_{a_1, b_1}^k , m_{a_1, b_2}^k , m_{a_2, b_1}^k , and m_{a_2, b_2}^k be the scores of the four

corresponding cells. Then

$$f^k(x, y) = \frac{a_2 - x}{a_2 - a_1} \frac{b_2 - y}{b_2 - b_1} m_{a_1, b_1}^k + \frac{a_2 - x}{a_2 - a_1} \frac{y - b_1}{b_2 - b_1} m_{a_1, b_2}^k \\ + \frac{x - a_1}{a_2 - a_1} \frac{b_2 - y}{b_2 - b_1} m_{a_2, b_1}^k + \frac{x - a_1}{a_2 - a_1} \frac{y - b_1}{b_2 - b_1} m_{a_2, b_2}^k.$$

Note that, when (x, y) coincides with any of the four centers, $f^k(x, y)$ is equal to the score of the corresponding cell.

For each of the above-defined points Q_h , we consider a second point $\tilde{Q}_h = (\chi_h, \psi_h, f_h^k)$ where $f_h^k = f^k(\chi_h, \psi_h)$. We then say that the cell under test $(i_{x,2}, i_{y,2})$ belongs to the subset associated with $(i_{x,1}, i_{y,1})$ if and only if

$$f_h^k \geq c_h^k \quad \forall h = 1, \dots, N - 1. \quad (11)$$

Operatively, for each local maximum cell, we start by testing its 8 neighbors, including the neighbor in the cluster if (11) is fulfilled. We then proceed by testing each neighbor of newly included cells (if not already tested), and we go ahead until no further cells can be included in the cluster.

Discussion: Condition (11) imposes that all of the $N - 1$ points Q_h , lying on the segment connecting the centers of a local maximum cell and a cell under test, falls “below” the surface S' . The net result is that the portion of S' corresponding to a cluster has an approximately convex shape. This is exemplified in Fig. 3, obtained with real measurements. The surface S' is first depicted⁷ along with markers corresponding to the centers of the four local maximum cells; examples of points $Q_h = (\chi_h, \psi_h, c_h^k)$ (black) and $\tilde{Q}_h = (\chi_h, \psi_h, f_h^k)$ (red) are then shown, for three local maximum cells and two test cells per maximum; finally, the portions of S' corresponding to the four clusters are highlighted (in red, green, blue, and cyan). The proposed approach to consider local maxima and growing clusters around them according to (11) is motivated by practical considerations. In our experiments, the image of a human target at this stage of the processing is typically a group of contiguous cells with a pronounced maximum in terms of score, a pronounced difference between the score of inner cells and the score of cells on the border (this is in part due to the smoothing operation (9)), and an approximately convex shape of the portion of the surface obtained via interpolation. Thus, the described processing aims at identifying groups of cells exhibiting these features.

The cluster grown around each local maximum cell is intrinsically associated with (layer k of) one of the regions A_i^t . In fact, since at the beginning of the points generation step the score of every cell not

⁷The surface S' is here shown only for illustration purposes. In the proposed method, only $N - 1$ z -axis values of S' shall be computed for each cell under test.

belonging to any region \mathbf{A}_i^t is equal to zero, every local maximum belongs to (layer k of) some region. To make this connection explicit, we denote the identified clusters at layer k by $U_{i,j}^k$, where i is the index of the region and j is the index of the cluster associated with the region. Exactly one point is extracted from each cluster $U_{i,j}^k$; moreover, only points characterized by a large enough *score* become useful for tracklet generation, as follows. The mean score and score variance in $U_{i,j}^k$ are defined as

$$\bar{m}_{i,j}^k = \frac{1}{|U_{i,j}^k|} \sum_{(i_x, i_y) \in U_{i,j}^k} m_{i,j}^k(i_x, i_y) \quad \text{and} \quad v_{i,j}^k = \frac{1}{|U_{i,j}^k|} \sum_{(i_x, i_y) \in U_{i,j}^k} (m(i_x, i_y) - \bar{m}_{i,j}^k)^2$$

respectively. Clusters with large mean score, score variance, and cardinality are more likely to be generated by a target. (A large variance arises from the above-mentioned difference between the score of inner cells and the score of cells on the border of a cluster.) Hence, the point corresponding to $U_{i,j}^k$ is constructed as $\mathbf{z}_{i,j}^k = (x_{i,j}^k, y_{i,j}^k, k, p_{i,j}^k)$, where

$$x_{i,j}^k = \frac{\sum_{(i_x, i_y) \in U_{i,j}^k} x_c(i_x, i_y) \cdot m(i_x, i_y)}{\sum_{(i_x, i_y) \in U_{i,j}^k} m(i_x, i_y)} \quad y_{i,j}^k = \frac{\sum_{(i_x, i_y) \in U_{i,j}^k} y_c(i_x, i_y) \cdot m(i_x, i_y)}{\sum_{(i_x, i_y) \in U_{i,j}^k} m(i_x, i_y)} \quad (12)$$

$$p_{i,j}^k = \bar{m}_{i,j}^k \cdot v_{i,j}^k \cdot |U_{i,j}^k|.$$

Note that $(x_{i,j}^k, y_{i,j}^k)$ is the cluster centroid, while $p_{i,j}^k$ assumes the meaning of score of $\mathbf{z}_{i,j}^k$. Measurements whose score is below a threshold are removed.

At the end of the process, a collection of points is generated for each region \mathbf{A}_i^t , $i = 1, \dots, R^t$. This is expressed as $\mathcal{Z}_i^{(t-W+1):t} = \{\mathcal{Z}_i^{t-W+1}, \mathcal{Z}_i^{t-W+2}, \dots, \mathcal{Z}_i^t\}$ where, for all $k \in \{t-W+1, \dots, t\}$, we have denoted by $\mathcal{Z}_i^k = \{\mathbf{z}_{i,j}^k | j = 1, \dots, N_i^k\}$ the subset of points corresponding to layer k of \mathbf{A}_i^t . The final set of points is

$$\mathcal{Z}^{(t-W+1):t} = \{\mathcal{Z}^{t-W+1}, \mathcal{Z}^{t-W+2}, \dots, \mathcal{Z}^t\} \quad (13)$$

where $\mathcal{Z}^k = \bigcup_{i=1}^{R^t} \mathcal{Z}_i^k$.

E. Tracklets Generation

Points generated by targets, collected in subsequent scans, tend to concentrate around tracklets, i.e., segments in the 3D space⁸ whose directions correspond to the kinematic states of the targets. For this reason, successive points within a time window can be associated to obtain tracklets.

We exploit partially overlapping time windows, each containing points of W scans; the points of the most recent window are given by (13). Tracklets are detected by the points in each time window

⁸The 3D space dimensions are represented by the x coordinate, y coordinate, and time.

independently. In this work, we employed the tracklet generation technique proposed in [30], adapted to a human target case. In particular, since the direction of a human target can change in one or two scans a narrower time window, i.e., a small enough W , should be used in our setting. Note that several tracklets may be generated in a time window due to false alarm points.

The l -th tracklet at scan t is denoted by \mathcal{T}_l^t and is a collection of W points belonging to \mathcal{Z}^{t-W+1} , \mathcal{Z}^{t-W+2} , \dots , and \mathcal{Z}^t , respectively. To keep the computational complexity under control, only tracklets with a large enough score are forwarded to the next step, while the others are discarded. To further limit complexity, depending on the application it is also possible to set a maximum number of detected tracklets per time window. The tracklet score is defined considering both the score of the corresponding points and an error measure for each point. More specifically, let the W points z^{t-W+1}, \dots, z^t be considered to form a possible tracklet. Then, the tracklet score is $(\prod_{k=t-W+1}^t p^k) / (\prod_{k=t-W+1}^t d^k)$ where p^k is the measurement score defined in the third equation of (12) and d^k is an error metric for $z^k = (x^k, y^k, k, p^k)$. This is calculated as the Euclidean distance between (x^k, y^k) and (\hat{x}^k, \hat{y}^k) , where $(\hat{x}^k, \hat{y}^k, k)$ is obtained by linear fitting of the $W - 1$ points (x^i, y^i, i) , $i \neq k$, with a least-squares criterion.

F. Tracklets Association, Outlier Removal and Trajectory Smoothing

The tracklets originated from the same target should be associated (i.e., joined) with each other to obtain the whole target trajectory [46]. In this work we have employed the multiple hypothesis tracking (MHT) tracklet association algorithm proposed in [30], with minor adjustments. Since adjacent windows are partially overlapping, tracklets including the same points are regarded as more likely to be originated from the same target. The algorithm also attempts to identify and remove false alarm tracklets to possibly let each target be represented by a unique point in every scan.

After tracklet association, at the generic scan period t we obtain a set of active trajectories $\{\Gamma\}^t$, whose time-variant cardinality $N^t = |\{\Gamma\}^t|$ is the number of detected targets at scan t . Each such detected target, at some scan $k \leq t$, is represented by a point belonging to one of these trajectories, along with its score and its error metric. Denoting by $\xi^k(\Gamma)$ the point of trajectory Γ at scan k , the structure of this point is $\xi^k(\Gamma) = (x^k, y^k, k; p^k, d^k)$.

Position errors along the trajectory may still be too large, due to outliers and measurement noise. Therefore, outliers removal and trajectory smoothing are performed to obtain the final smoothed trajectory. Outlier removal is based again on the metric d_k . Point $\xi^k(\Gamma)$ is regarded as an outlier whenever the error metrics of points along trajectory Γ fulfill

$$d^k \geq \frac{\nu}{2D} \left(\sum_{n=-D}^{-1} d^{k+n} + \sum_{n=1}^D d^{k+n} \right) \quad (14)$$

where $D = \lfloor W/2 \rfloor$ and $\nu > 0$ is a system design parameter. The above search for outliers is performed in a sliding window fashion, with a window size of $W + 1$ scans (apart from the initial and final scan periods). Then the point $\xi^k(\Gamma)$ is replaced by the smoothed point $\hat{\xi}^k(\Gamma) = (\hat{x}^k, \hat{y}^k, k; p^k, \hat{d}^k)$, obtained by linear fitting the other W points in the time window, with a least-squares criterion.

IV. EXPERIMENTAL RESULTS

An experimental campaign was performed to assess the effectiveness of the proposed processing chain. The performance analysis was carried out both by computer simulations (synthetic data) and by employing actual UWB waveforms collected in real environments.

Concerning computer simulations, two sets of simulations were performed. In the first set, the performance of a monostatic UWB RSN with $N_R = 4$ sensors was analyzed, both with two tracks and with three tracks; in all cases the tracks include maneuvering as well as walking-stop-walking targets. In the second set of simulations, a multistatic UWB RSN with $N_R = 4$ sensors and one transmitter was considered. Again, two scenarios with two and three targets were analyzed.

Regarding the analysis with actual UWB waveforms, two case studies were investigated. In the first case study we used a monostatic RSN to detect and track one human target in an outdoor environment (specifically, a balcony); in the second case study we used a multistatic RSN to detect two human targets in an indoor environment. We employed Humatics UWB nodes (P440 and P410) and a laptop acting as the FC. The nodes operate in the [3.1, 4.8] GHz band, and have two antenna ports (for transmission and reception), a power interface, and a control port. In all simulation and case studies, the sampling frequency at the receiver is $f_s = 16.48$ GHz (sampling time $T_s = 61$ ps), the scan period is 0.45 s and the cell dimensions are $\Delta_x = 10$ cm and $\Delta_y = 10$ cm.

The value of α in (2) was set to 0.75 in all experiments, and the maximum number of detected tracklets per time window was set to 5. The proposed approach involves a number of thresholds, whose values often need to be set by manual tuning. We used $\beta = 2$ in the adaptive threshold η_{score}^t in (6). Moreover, good results were obtained by setting the adaptive threshold γ_{score}^t in (7) equal to η_{score}^t , i.e., $\gamma_{\text{score}}^t = \eta_{\text{score}}^t$. Table I summarizes the values of the several parameters used in the experiments. A single value indicates that the corresponding parameter was did not change in all experiments. The third column of the table provides a brief description of each parameter along with its influence to the processing performance.

A. Computer Simulations

As mentioned above, in order to numerically assess the performance of the proposed processing chain, we performed two sets of simulations, hereafter referred to as “Experiment 1” and “Experiment 2” and

TABLE I
NUMERICAL VALUES OF THE PARAMETERS INVOLVED IN THE PROCESSING

| Parameter | Value | Description |
|----------------------------------|---|--|
| (N_x, N_y, K) | Experiment 1: (150, 150, 40) Experiment 2: (150, 150, 40) Case study 1: (120, 80, 124) Case study 2: (120, 60, 40) | - (N_x, N_y) proportional to the surveillance area and inversely proportional to grid cell size. - K inversely proportional to the scan period. |
| Number of sensors, N_R | Experiment 1: $N_R = 4$ Experiment 2: $N_R = 4$ (+ 1 Tx) Case study 1: $N_R = 3$ Case study 2: $N_R = 4$ (+ 1 Tx) | - Monostatic RSN in simulation set 1 and case study 1. - Multistatic RSN in simulation set 2 and case study 2. |
| Pulse integration factor, N_s | Experiment 1: 4096 Simulation 2: 20800 Case study 1: 4096 Case study 2: 20800 | - A larger N_s improves the SNR, beneficial to weak target detection. - A smaller N_s enables a lower scan period, beneficial to maneuvering target detection. |
| Samples per scan, N_c | 1500 | N_c is proportional to coverage range of sensor. |
| κ in (1) | 10 | - A smaller κ reduces complexity in static clutter removal, but brings a performance deterioration for slowly moving targets. |
| α in (2) | 0.75 | - $\alpha \in (0, 1)$ helps suppressing strong clutter; a smaller α should be applied in strongly cluttered environments. |
| β in (6) | 2 | - A larger β helps decreasing the false alarms in a single scan, but tends to hinder detection of weak targets. |
| (W, s) | (4, 2) | - Larger W and s are beneficial to weak target detection but deteriorate tracking performance of maneuvering targets. |
| (d_x, d_y) | (5, 5) | - Related to the size of target region; larger (d_x, d_y) reduces complexity but tend to deteriorate detection performance. |
| γ_{score}^t in (7) | η_{score}^t in (6) | - γ_{score}^t and η_{score}^t are related to the SNR of target echoes; lower values favor weak target detection but yield more false alarms. |
| γ_{num}^t in (8) | 150 | - Proportional to the target size; a smaller γ_{num}^t is beneficial detection of smaller targets but brings more false alarms. |
| w in (9) | 3 | - Related to the target size and distribution of noise; a larger w is beneficial to suppress the ghost echoes but increases complexity and hinders partition of regions of closely-spaced targets. |
| N in (10) | 10 | - A larger N is beneficial to target region partition but increases complexity. |
| ν in (14) | 3 | - A smaller ν is beneficial to outlier removal but may deteriorate localization accuracy when target is highly maneuvering. |

corresponding to a monostatic and a multistatic UWB RSN, respectively. In both sets of simulations, the signals produced by the UWB sensors, representing the input of the FC, were generated according to distributions extracted from real data, as described hereafter.

Measurements were first collected with Humatics UWB radios both in a monostatic configuration in an outdoor environment and in a multistatic configuration in an indoor environment (hence, the two sets of measurements reflect the above-mentioned case studies), using human targets in all cases. Then, for each collected UWB waveform, the corresponding signal \mathbf{m}_n^t was obtained by applying the in-sensor processing described in Section III-A. This way, two sets of signals \mathbf{m}_n^t were built, one relevant to the monostatic configuration (in outdoor environment) and one to the multistatic configuration (in indoor environment). Next, separately for each of the two signal sets, the distribution of the generic element of \mathbf{m}_n^t was obtained, both for those range bins (elements of \mathbf{m}_n^t) in which a target echo exists and for those ones associated with clutter echoes and noise only. During each simulation, given the positions of the targets at scan t and given a sensor n , each sample of \mathbf{m}_n^t was generated randomly by applying the

appropriate distribution.⁹

A method for points generation in presence of UWB signals has been developed in [5]. To benchmark the performance of the proposed method measured via numerical simulations, the point generation method of [5] has been employed and the points generated by this method have been further processed for track detection by two different techniques, namely, the PHD filter from [8] and the TBD technique from [20], both approaches targeting tracking of extended targets.

The detected tracks of a monostatic and of a multistatic RSN (Experiment 1 and Experiment 2, respectively) using the proposed method, both in presence of two targets (“Test 1”) and of three targets (“Test 2”), are presented in Fig. 4. Each subfigure shows the target ground truths together with the final detection points produced by the proposed method (red points), after trajectory smoothing. In particular, the superposition of all detection points generated through 100 simulations is illustrated. Both in presence of two targets and of three targets, the target detection rate is higher than 90% with an average positional error lower than 12 cm. The points generated by the detection approach in [5] are shown in Fig. 5. We can see that a relatively large number false of alarm points arise, obscuring the target points. Compared with the case of two targets, in the three target case the target detection rate is lower and more false alarm points are generated. A comparison between Fig. 4 and Fig. 5 allows appreciating the remarkably better performance offered by the proposed method with respect to a “points generation only” approach in presence of weak UWB echoes, especially in cluttered environments with multiple (e.g., three) targets.

Fig. 6 shows the detected tracks returned by the TBD algorithm recently proposed in [20], using as input the points generated by the detection technique in [5]. A lower detection rate is achieved and and more false tracks are generated. In [20], a TBD approach based on a 3D Projection was developed and shown to be able to outperform other methods, including the 3D Hough transform TBD algorithm proposed in [30] and the PHD filter based algorithm proposed in [8]. The fact that the method here developed performs better than the one of [20] indicates that it can outperform other tracking approaches in presence of UWB waveforms.

The results of Experiment 1 and Experiment 2 (both Test 1 and Test 2 for each of them) are summarized in Table II, Table III, Table IV, and Table V, including the detection rate and the positional error of each target, the number of average false alarm tracks per scan, the optimal subpattern assignment (OSPA) distance [47], and the average running time over 100 Monte Carlo experiments. As it can be observed from the reported data, the detection rate of the proposed method is remarkably higher than that of the

⁹The statistical distributions according to which the signals \mathbf{m}_n^t are generated throughout the simulations are therefore tied to the specific in-sensor processing described in Section III-A. This is not an issue because this in-sensor processing has been specifically designed for UWB signals and because the same distributions are employed to generate the signals \mathbf{m}_n^t at the input of all FC algorithms whose detection and tracking performance we seek to compare.

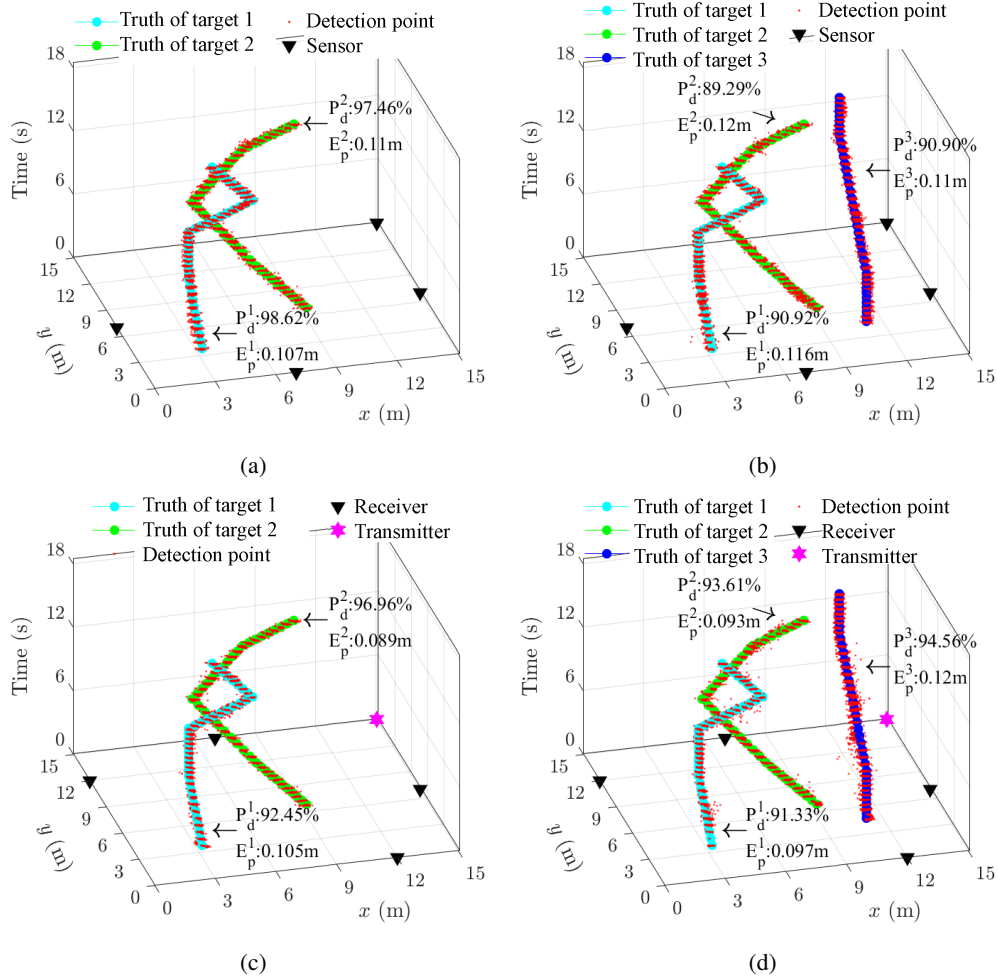


Fig. 4. Results of the proposed method through 100 numerical simulations. (a) Experiment 1 (monostatic RSN) – Test 1 (two targets); (b) Experiment 1 – Test 2 (three targets); (c) Experiment 2 (multistatic RSN) – Test 1; (d) Experiment 2 – Test 2.

benchmark methods, with less false alarm tracks. As a consequence, the proposed method achieves a lower OSPA distance. In contrast, the average running time of proposed method is higher than that of the other two methods. This fact mainly comes as a consequence of the voting (Sec. III-B) and the 3D image processing (Sec. III-C) stages. Importantly, we can trade-off the computational complexity and the performance of the proposed method by adjusting the size of grid cell, defined by Δ_x and Δ_y . Besides $\Delta_x = \Delta_y = \Delta = 0.1$ m, tests we carried out for $\Delta = 0.05$ m and $\Delta = 0.2$ m, corresponding to a smaller and a larger grid cell, respectively. As from the tabular data, using a larger Δ , such as $\Delta = 0.2$ m, yields considerable savings in complexity at the price of a slight deterioration in performance, which however remains generally better than that of the other techniques. A smaller Δ , such as $\Delta = 0.05$ m achieves very marginal gains over $\Delta = 0.1$ m while imposing unacceptable complexity.

Fig. 7 illustrates the OSPA distance through different scans for the proposed method and the two

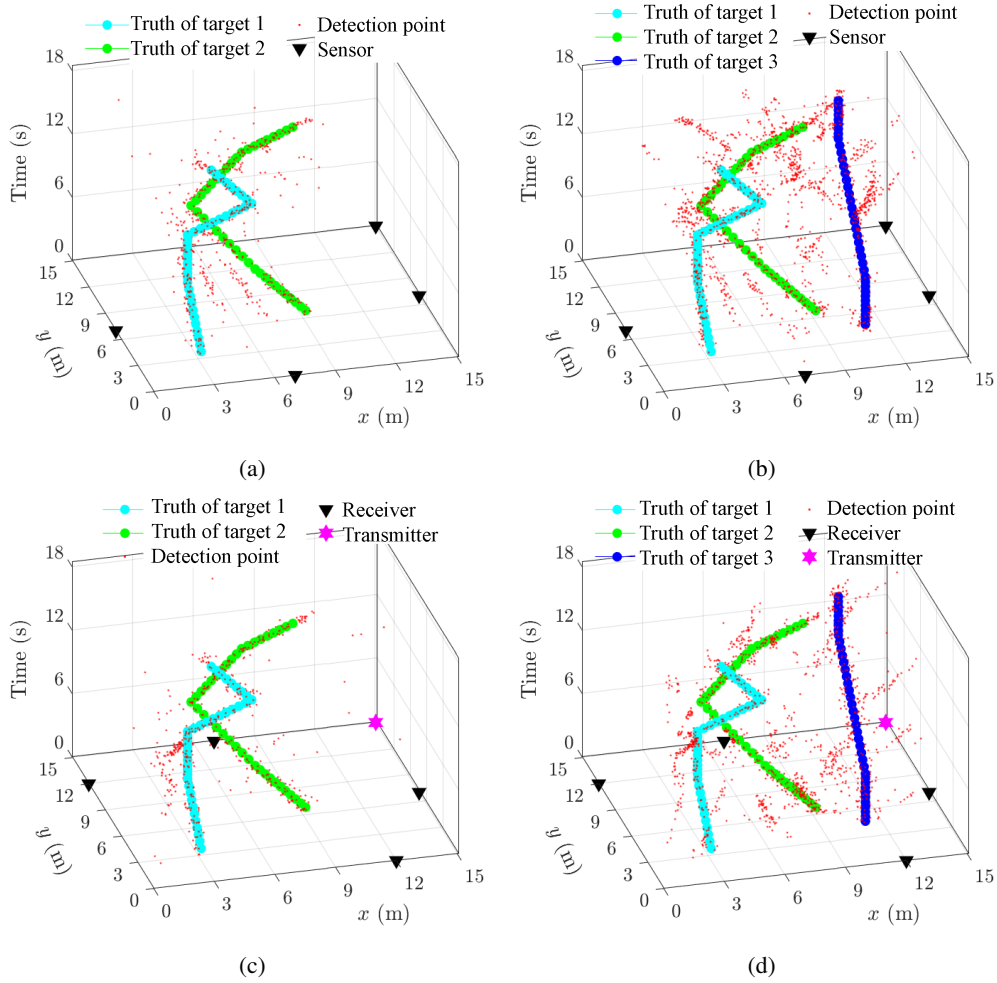


Fig. 5. Detection points generated by the method [5] through 10 numerical simulations. (a) Experiment 1 – Test 1; (b) Experiment 1 – Test 2; (c) Experiment 2 – Test 1; (d) Experiment 2 – Test 2.

methods used as benchmarks; the applied OSPA metric employs the Euclidean distance with a cut-off parameter 0.7 meters and an order parameter equal to 1. In all cases (Experiments 1 and 2, Tests 1 and 2), the proposed method achieves an OSPA distance lower than the one achieved by competing approaches. As we can observe in Fig. 7, the OSPA distance values achieved by the proposed method at the 11th, 21th, and 31th scans are higher than those in the other scans. This fact can be attributed to the least-square criterion adopted in trajectory smoothing: A larger positional error is unavoidable when the target is maneuvering because of the mismatch between the smoothing model and actual target motion state [48]. The smoothing model trades off positional error between target in straight line and in maneuvering. A smaller scan period is helpful to decrease the positional error in those scans in which the target is maneuvering; this can be achieved by reducing pulse integration factor N_s , but at the expense of decreasing the target detection rate.

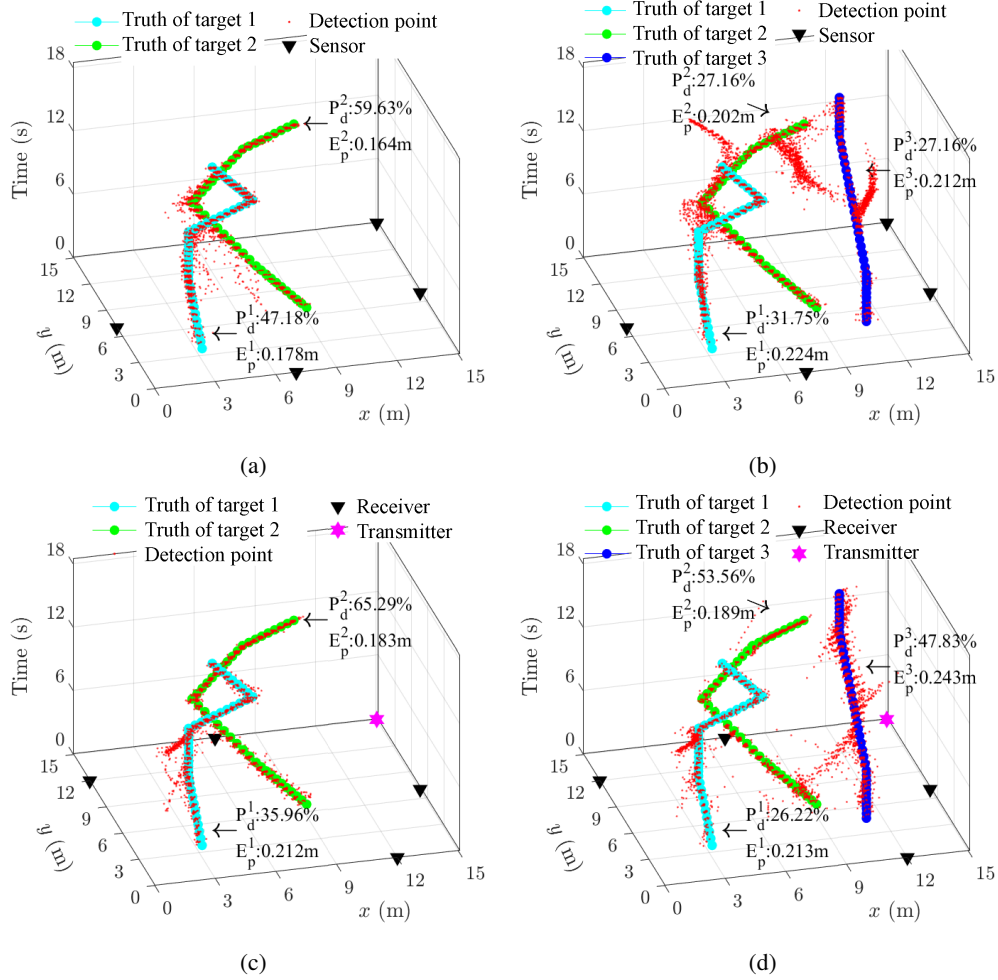


Fig. 6. Results of the TBD method in [20] fed with the points generated by the detection technique in [5] through 100 numerical simulations. (a) Experiment 1 – Test 1; (b) Experiment 1 – Test 2; (c) Experiment 2 – Test 1; (d) Experiment 2 – Test 2.

TABLE II
RESULTS OF EXPERIMENT 1 – TEST 1 (MONOSTATIC RSN, 2 TARGETS).

| | P_d^1 ^(a) | E_p^1 ^(b) | P_d^2 | E_p^2 | N_{FA} ^(c) | OSPA | Time |
|---------------------------------------|------------------------|------------------------|---------|---------|-------------------------|-------|---------|
| Proposed method ($\Delta = 0.1$ m) | 98.62% | 0.107 | 97.46% | 0.114 | 0.0087 | 0.123 | 21.87 s |
| Point generation [5] + TBD [20] | 47.18% | 0.178 | 59.63% | 0.164 | 0.0318 | 0.426 | 12.45 s |
| Point generation [5] + PHD filter [8] | 49.40% | 0.329 | 63.01% | 0.334 | 0.9810 | 0.531 | 07.44 s |
| Proposed method ($\Delta = 0.2$ m) | 98.28% | 0.167 | 94.87% | 0.155 | 0.0052 | 0.175 | 15.87 s |
| Proposed method ($\Delta = 0.05$ m) | 98.72% | 0.105 | 95.98% | 0.110 | 0.0089 | 0.164 | 61.51 s |

^(a) P_d^i : detection rate of target i .

^(b) E_p^i : average positional error of target i .

^(c) N_{FA} : average number of false alarms per scan.

B. Case Study 1

In the first experiment, the goal was to detect a human target, of weight 80 kg and height 170 cm, walking over a balcony. The experimental scenario is portrayed in Fig. 8(a). The RSN was composed of

TABLE III
RESULTS OF EXPERIMENT 1 – TEST 2 (MONOSTATIC RSN, 3 TARGETS).

| | P_d^1 | E_p^1 | P_d^2 | E_p^2 | P_d^3 | E_p^3 | N_{FA} | OSPA | Time |
|---------------------------------------|---------|---------|---------|---------|---------|---------|----------|-------|---------|
| Proposed method ($\Delta = 0.1$ m) | 90.92% | 0.116 | 89.29% | 0.122 | 90.89% | 0.118 | 0.0280 | 0.206 | 22.38 s |
| Point generation [5] + TBD [20] | 31.75% | 0.224 | 46.86% | 0.202 | 27.16% | 0.212 | 0.4010 | 0.537 | 13.86 s |
| Point generation [5] + PHD filter [8] | 53.45% | 0.314 | 61.47% | 0.366 | 51.10% | 0.392 | 3.5780 | 0.587 | 07.84 s |
| Proposed method ($\Delta = 0.2$ m) | 86.67% | 0.228 | 88.60% | 0.197 | 87.56% | 0.197 | 0.0220 | 0.244 | 16.12 s |
| Proposed method ($\Delta = 0.05$ m) | 92.92% | 0.114 | 86.29% | 0.121 | 92.90% | 0.109 | 0.0028 | 0.208 | 71.85 s |

TABLE IV
RESULTS OF EXPERIMENT 2 – TEST 1 (MULTISTATIC RSN, 2 TARGETS).

| | P_d^1 | E_p^1 | P_d^2 | E_p^2 | N_{FA} | OSPA | Time |
|---------------------------------------|---------|---------|---------|---------|----------|-------|---------|
| Proposed method ($\Delta = 0.1$ m) | 96.64% | 0.105 | 92.45% | 0.089 | 0.0045 | 0.132 | 19.02s |
| Point generation [5] + TBD [20] | 35.96% | 0.212 | 65.29% | 0.183 | 0.0650 | 0.453 | 15.89s |
| Point generation [5] + PHD filter [8] | 51.13% | 0.331 | 65.75% | 0.341 | 0.7540 | 0.513 | 12.21 s |
| Proposed method ($\Delta = 0.2$ m) | 95.89% | 0.157 | 92.82% | 0.150 | 0.0091 | 0.210 | 16.49 s |
| Proposed method ($\Delta = 0.05$ m) | 93.84% | 0.110 | 91.02% | 0.094 | 0.0094 | 0.180 | 64.05 s |

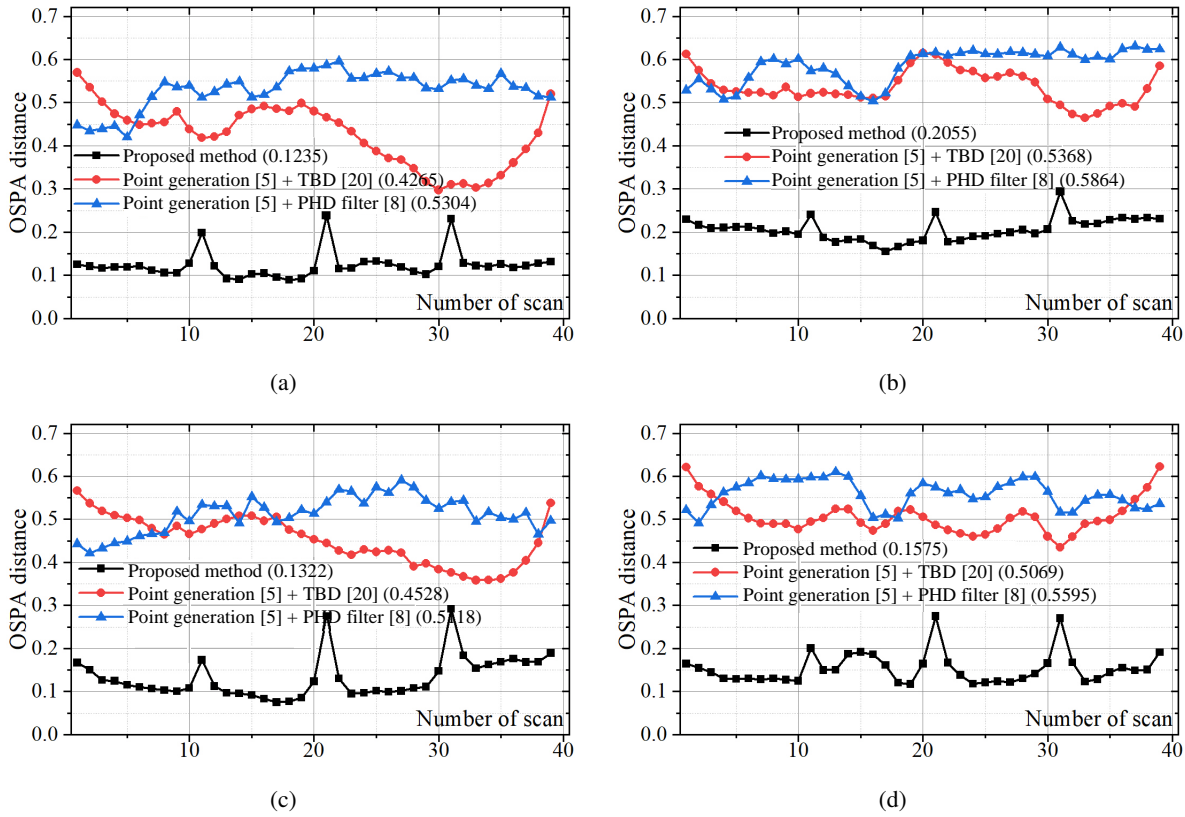
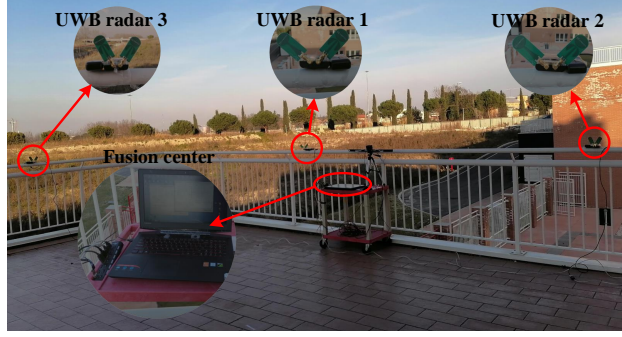


Fig. 7. OSPA distance of methods. (a) Experiment 1 test 1. (b) Experiment 1 test 2 . (c) Experiment 2 test 1. (d) Experiment 2 test 2.

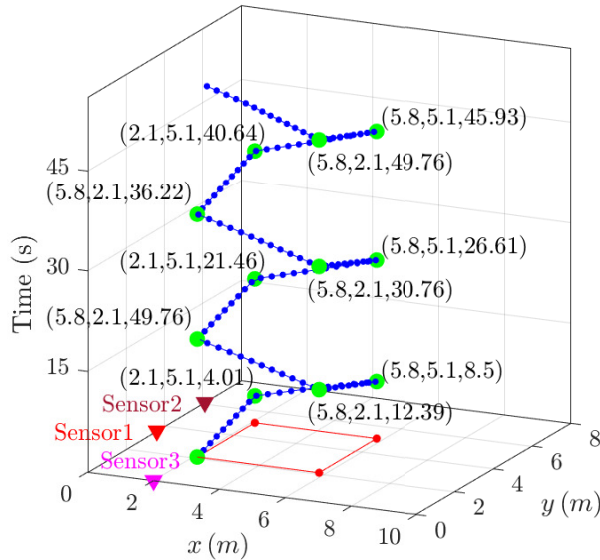
$N_R = 3$ UWB sensors, each configured as a monostatic radar. Fig. 8(b) illustrates, on the $x-y$ plane, both the sensor positions and the target trajectory, a rectangular path of size $3.7\text{ m} \times 3\text{ m}$ covered three times

TABLE V
RESULTS OF EXPERIMENT 2 – TEST 2 (MULTISTATIC RSN, 3 TARGETS).

| | P_d^1 | E_p^1 | P_d^2 | E_p^2 | P_d^3 | E_p^3 | N_{FA} | OSPA | Time |
|---------------------------------------|---------|---------|---------|---------|---------|---------|----------|-------|---------|
| Proposed method ($\Delta = 0.1$ m) | 91.33% | 0.097 | 93.61% | 0.093 | 94.56% | 0.121 | 0.0092 | 0.157 | 20.25 s |
| Point generation [5] + TBD [20] | 26.22% | 0.213 | 53.56% | 0.189 | 47.83% | 0.243 | 0.1620 | 0.507 | 16.76 s |
| Point generation [5] + PHD filter [8] | 53.42% | 0.334 | 56.75% | 0.362 | 62.35% | 0.340 | 2.6200 | 0.559 | 11.37 s |
| Proposed method ($\Delta = 0.2$ m) | 93.64% | 0.199 | 92.61% | 0.182 | 93.64% | 0.193 | 0.0190 | 0.214 | 16.39 s |
| Proposed method ($\Delta = 0.05$ m) | 92.89% | 0.129 | 96.71% | 0.096 | 94.23% | 0.110 | 0.0097 | 0.178 | 72.95 s |



(a)



(b)

Fig. 8. (a) Picture of the experimental setup for case study 1. (b) Positions of the sensors and ground truth.

in about 56 s with an approximately constant speed of 0.717 m/s. The evolution of the target trajectory in the 3D $x-y-t$ space is also shown. The total number of scans is $K = 124$.

The signals m_n^t (i.e., the result of the in-sensor processing) are represented in Fig 9(a) for the sensor $n = 3$ located in position (2, 0). Each row corresponds to one of the $K = 124$ scan periods. In each row

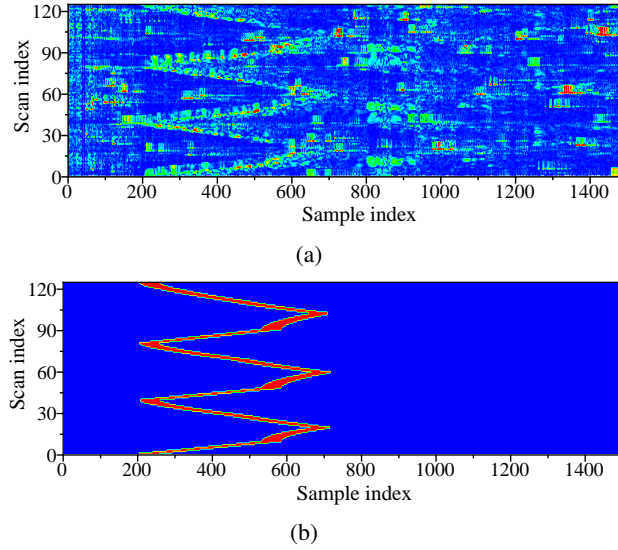


Fig. 9. (a) Illustration of the signals m_n^t for $n = 3$ and $1 \leq t \leq 124$. (b) Actual target ToA.

t , $1 \leq t \leq 124$, all samples $m_n^t(j)$ are depicted by means of colors, where blue represents zero and red the largest sample value; every two consecutive samples correspond to a difference of 61 ps in terms of ToA and of 0.91 cm in terms of range. The actual target ToA is depicted in Fig. 9(b). It is possible to observe how signals incoming to the FC are affected by residual clutter and some strong echoes, not due to the target but even stronger (by a factor up to 3) than the target's ones. We found that these echoes originated by cars going along a large street close to the balcony where the experiment was performed. Very similar figures are obtained for sensors $n = 1$ and $n = 2$.

Fig. 2 shows the original samples $m_n^t(j)$, $j = 1, \dots, N_c$, generated by one of the sensors along with the corresponding peak-suppressed samples, $(m_n^t(j))^\alpha$, with $\alpha = 0.75$. The range bin where the target exists is marked by a thick black line and it reveals how the amplitude of target echoes may not be the largest, although it is larger than that of most non-target range bins.

An example of score maps S_1^t , S_2^t , and S_3^t , one for each receiver, is illustrated through color maps in Fig. 10(a), Fig. 10(b), and Fig. 10(c), respectively. The sensor positions are marked with triangles, the actual position of the target is indicated by a cyan star and, in each map, all scores are normalized to the peak one for visualization purposes. These score maps are the ones obtained during the first scan period ($t = 1$). The corresponding overall (normalized) score map, obtained via (4) and (5), is shown in Fig. 10(d). Note that the overall score map predicts a potential target in the correct position. However, the presence of a relatively-strong residual clutter prevents from performing a correct detection in a single scan, leading to a considerable number of false alarms for a low detection threshold or to excessive

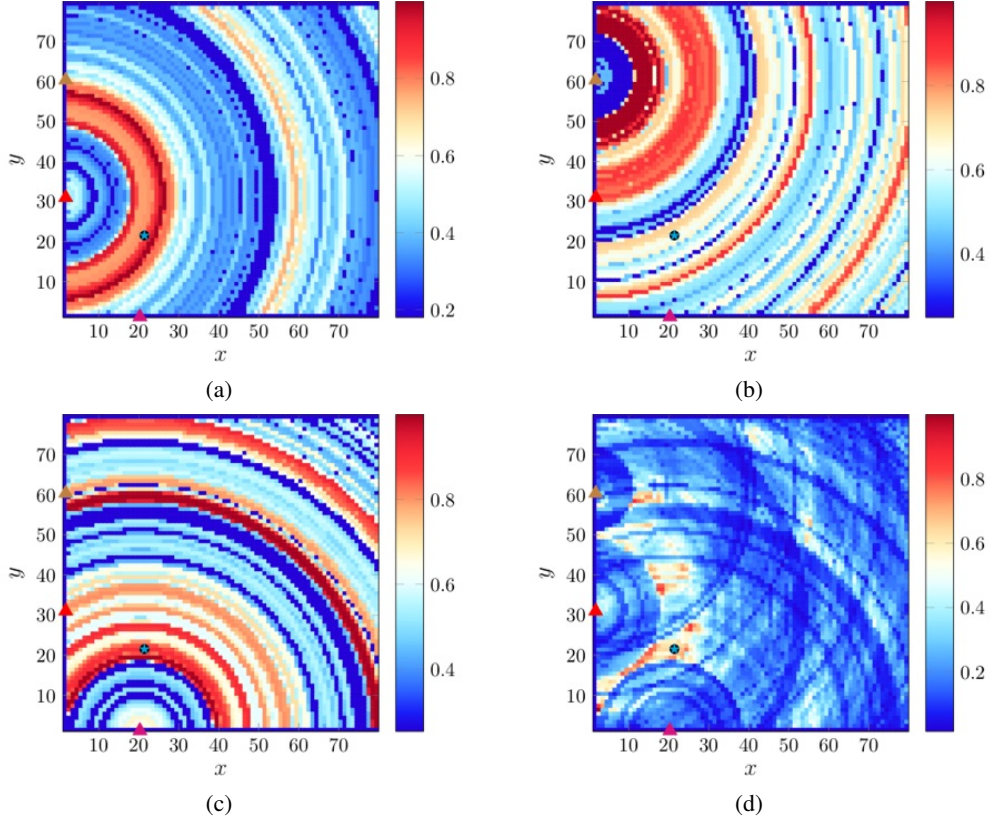


Fig. 10. Score maps in the first scan period ($t = 1$). (a) Score map \mathcal{S}_1^t for the sensor in position $(0, 3.1 \text{ m})$. (b) Score map \mathcal{S}_2^t for the sensor in position $(0, 6.1 \text{ m})$. (c) Score map \mathcal{S}_3^t for the sensor in position $(2 \text{ m}, 0)$. (d) Overall score map \mathcal{S}^t .

mis-detections of weak targets if the detection threshold is increased. In a classical track-after-detect configuration, this makes the tracking filter task very problematic, which justifies our TBD approach.

Fig. 11(a), 11(b), and 11(c) show the result of the 3D processing proposed in Section III-C. In particular, Fig. 11(a) illustrates how the 3D data structure M^t would appear in the last scan period ($t = 124$) by simply stacking score maps $S(i_x, i_y)$ obtained by (5) and (6) (cells with a null score are left blank). Several isolated cells with a nonzero score, originated by noise and clutter, can be observed. Fig. 11(b) shows the result of application of region growing only with $W = 4$ and $s = 2$. Only one region was found in all time windows ($R^t = 1$ for all $t = W + q \cdot s$, $q = 0, \dots, 59$). As it is possible to appreciate comparing Fig. 11(b) with Fig. 11(a), most isolated cells are now removed and the generated 3D region resembles the ground truth depicted in Fig. 8(b). An even closer match with the ground truth is obtained after application of 3D opening operation, as depicted in Fig. 11(c). The points obtained in each scan through the proposed point generation method (Section III-D) are depicted in Fig 11(d). These points are processed in each time window by the employed tracklet generation algorithm (Section III-E). In each time window, we employed the tracklet generation by enumeration. The non-discarded detected tracklets

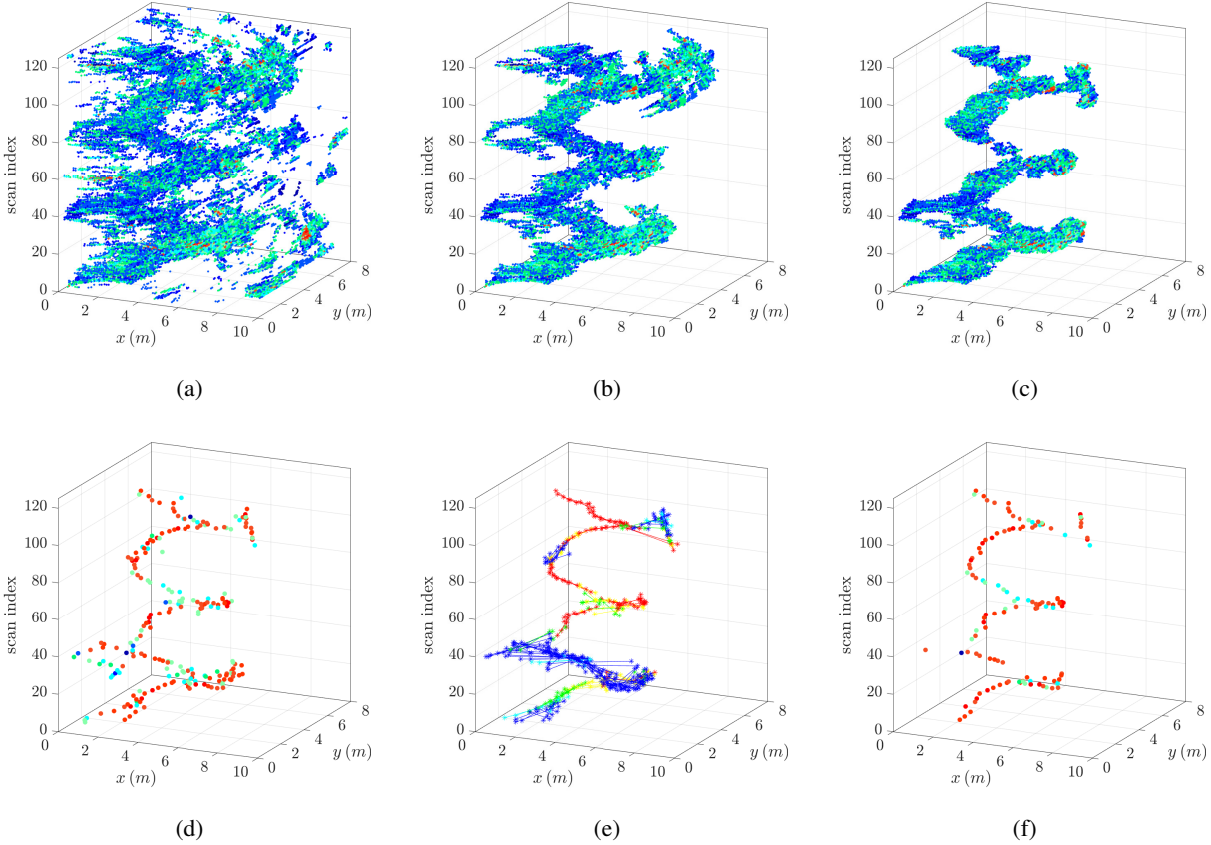


Fig. 11. (a) Example of 3D data structure applying voting, i.e., (5) only. Red and blue colors denote the largest and the smallest score, respectively. (b) The same data structure obtained by applying 3D region growing. (c) The same data structure obtained by applying also 3D opening operation. (d) The points obtained with the proposed generation method, and (e) the corresponding tracklets. (f) Trajectory points after tracklet association (before outlier removal and trajectory smoothing).

are presented in Fig 11(e), where the red color indicates that only one tracklet was generated in the time window and, on the opposite side, the blue color indicates that 5 tracklets (the largest number) were generated. Fig 11(f) shows the unique point obtained in each scan after tracklet association. Compared with the Fig 11(d), we can see how several false alarm points have been removed.

The smoothed 3D trajectory after outlier removal and smoothing is presented in Fig. 12 (red circles) together with the ground truth (green and blue triangles). Comparing it with the points in Fig. 11(f), we can see how the positional error is decreased after smoothing. The projection of the smoothed trajectory on the $x-y$ plane has an “oval” shape as an effect of smoothing, while the actual one is rectangular. Therefore, positioning errors close to the corners tend to be larger.

To account for both cardinality and localization errors, the tracking performance was assessed by OSPA distance [47]. Since the trajectory was detected in all 124 scans without false alarms, the OSPA distance corresponds to the positioning error (distance between the actual target position and the estimated one);

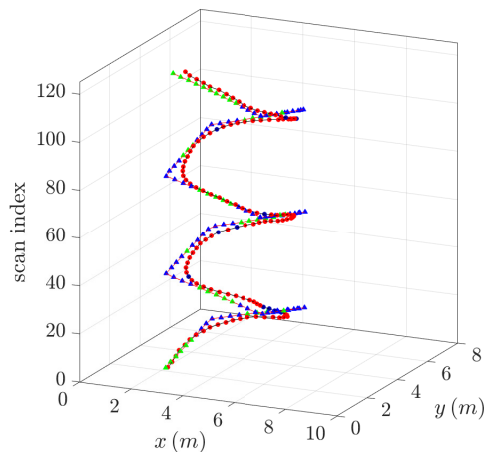


Fig. 12. The ground truth trajectory (triangles) and the smoothed one (circles). Green triangles indicate straight portions of the ground truth; blue triangles indicate a turning target.

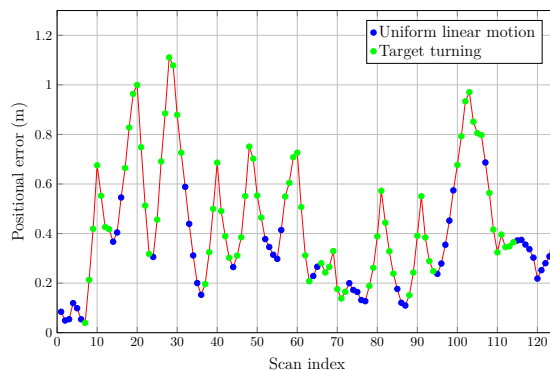


Fig. 13. The positional error of the smoothed trajectory among the 124 scans.

this is shown in Fig. 13 as a function of the scan index, where green and blue markers correspond to scan periods where the target is turning and where it moves with uniform linear motion, respectively. The average positioning error over all scans is 0.419 m. As expected, the performance is better during uniform linear motion (blue points), where the average error is 0.283 m, than during target turning (green points) where it reaches 0.502 m. Similar with the OSPA distance in Fig. 7, it is irreparable because of the mismatching between smooth model and actual target motion state when the target is maneuvering. However, the overall performance is remarkably good, considering that the average error is comparable with the target transverse size.

The effectiveness of the proposed points generation method (Section III-D) followed by tracklet detection and association can be observed by repeating the processing with the very same data and turning

TABLE VI
AVERAGE POSITIONING ERROR.

| | After point generation | After trajectory smoothing | Target turning | Uniform linear motion |
|-----------------|------------------------|----------------------------|----------------|-----------------------|
| Simple method | 0.535 | 0.499 | 0.526 | 0.455 |
| Proposed method | 0.514 | 0.419 | 0.502 | 0.283 |

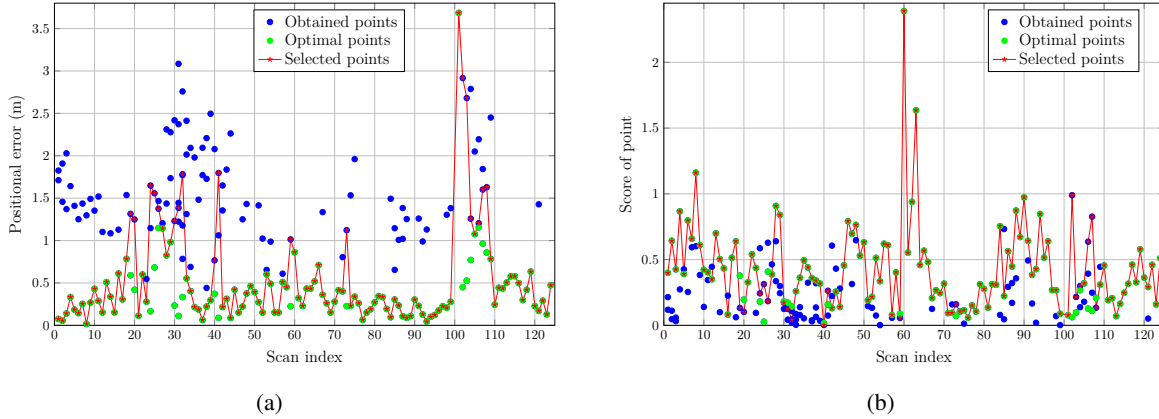


Fig. 14. (a) The positional error of the points before trajectory smooth. (b) The score of the points obtained in target detection.

all of these steps off. Specifically, instead of considering layer k of the region \mathbf{A}_i^t , finding clusters $U_{i,j}^k$ out of it, extracting one point per cluster, and applying tracklet detection and association to them, we can imagine to extract one point from layer k of \mathbf{A}_i^t , in particular its centroid, and to feed the smoothing filter with these points directly. This latter technique is here referred to as the “simple method”. Doing this, we would obtain an average positional error of 0.535 m, slightly worse than the one of the proposed technique after points generation, equal to 0.514 m. The gap increases substantially after smoothing and outlier removal, which are almost ineffective on the points generated by the simple method. The overall average positioning error of smoothed trajectory is 0.499 m; in particular, it equals 0.526 m during target turning and 0.455 m during uniform linear motion. The superiority of the proposed approach is mainly due to clutter cells being entirely removed from the calculation of the target position. The comparison is summarized in Table VI.

To provide insights about the tracklet generation and association steps (Section III-E and Section III-F), in Fig. 14(a) we show the positioning error of all points generated through all scan periods (blue and green markers – recall from Section III-D that multiple points per scan may be generated). For each scan index, green markers are associated with the optimal (least error) point and blue markers with the other points. The average positioning error of all generated points (green and blue) and optimal points (green only) are 0.909 m and 0.387 m, respectively. The selected points, corresponding to the points

TABLE VII
RESULTS OF CASE STUDY 1

| | P_d^1 & E_p^1 | N_{FA} | OSPA |
|---------------------------------------|-------------------|----------|-------|
| Proposed method ($\Delta=0.1m$) | 124/124 & 0.419 | 0 | 0.419 |
| Point generation [5] + TBD [20] | 47/124 & 0.591 | 3/124 | 0.659 |
| Point generation [5] + PHD filter [8] | 59/124 & 0.583 | 19/124 | 0.652 |

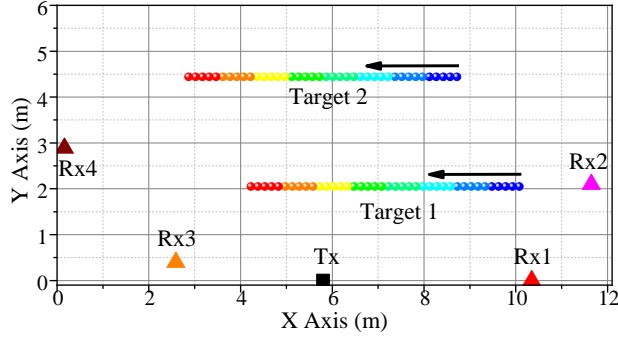


Fig. 15. Case study 2 scenario. Tx: UWB transmitter. Rx: UWB sensor.

in Fig. 11(f), are marked with a red star in the figure. Out of all 124 scan periods, the optimal point is selected as a trajectory points 105 times, with an average positional error of the selected points as low as 0.535m which reveals the benefits provided by tracklets generation and association in terms of localization accuracy. The score of the points generated through all scans is presented in Fig. 14(b). As we can see, in most scans the optimal points (green marks) exhibit the largest score and in the majority of scans the points with the largest score are the ones selected as the target points. This shows how the adopted score is beneficial in terms of trajectory detection.

The overall results of Case Study 1 are presented in Table VII. The detection rate of the proposed method is remarkably higher than that of the other two methods. The detection rate of points generation [5] followed by the TBD method of [20] is slightly lower than the one achieved by the same points generation technique followed by PHD filtering, because the TBD technique in [20] is tailored to straight-line tracklets. The number of false alarms with PHD filtering is higher because false tracks can be suppressed by multi-scan detection. According to these results, the proposed method can outperform the two competing methods within a monostatic RSN configuration.

C. Case Study 2

In the second experiment, we employ a multistatic RSN, composed of one UWB transmitter and $N_R = 4$ UWB sensors placed in a rectangular area of size 12 m \times 6 m, to detect two human targets in an indoor scenario. The ground truth of the two targets is provided in Fig. 15, where the transmitter and

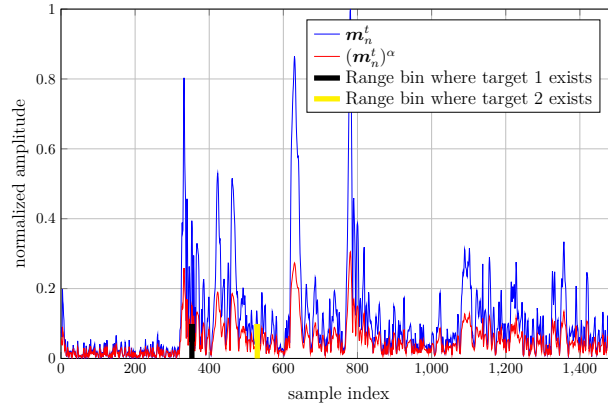


Fig. 16. Example of signal m_n^t collected during the second experiment.

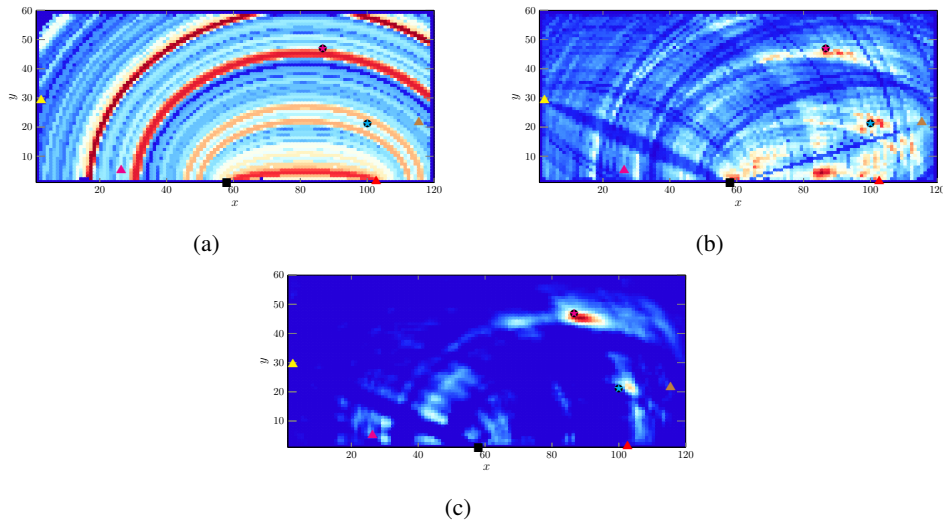


Fig. 17. (a) Example of score map S_n^t . (b) The overall score map S^t before clutter suppression. (c) The overall score map \hat{S}^t after clutter suppression.

four sensors are indicated by the black square and the four triangles, respectively. The total number of scan periods in this second experiment is $K = 40$.

In each bistatic pair, a single UWB pulse transmission causes the reception of at least two pulse replicas: the direct one via the transmitter-to-receive path and the target-reflected one via the transmitter-to-target-to-receiver path. Fig. 17(a) shows an example of the actual waveform m_n^t received by the first sensor ($n = 1$) in the first scan ($t = 1$). The corresponding score map S_n^t is depicted in Fig. 17(b), in which the transmitter, sensor, and two target positions are marked with a black square, a red triangle, and two stars, respectively. An example of overall score map S^t (again for $t = 1$) is shown in Fig 17(c). In this experiment, we applied the clutter removal procedure described in Appendix A to remove static

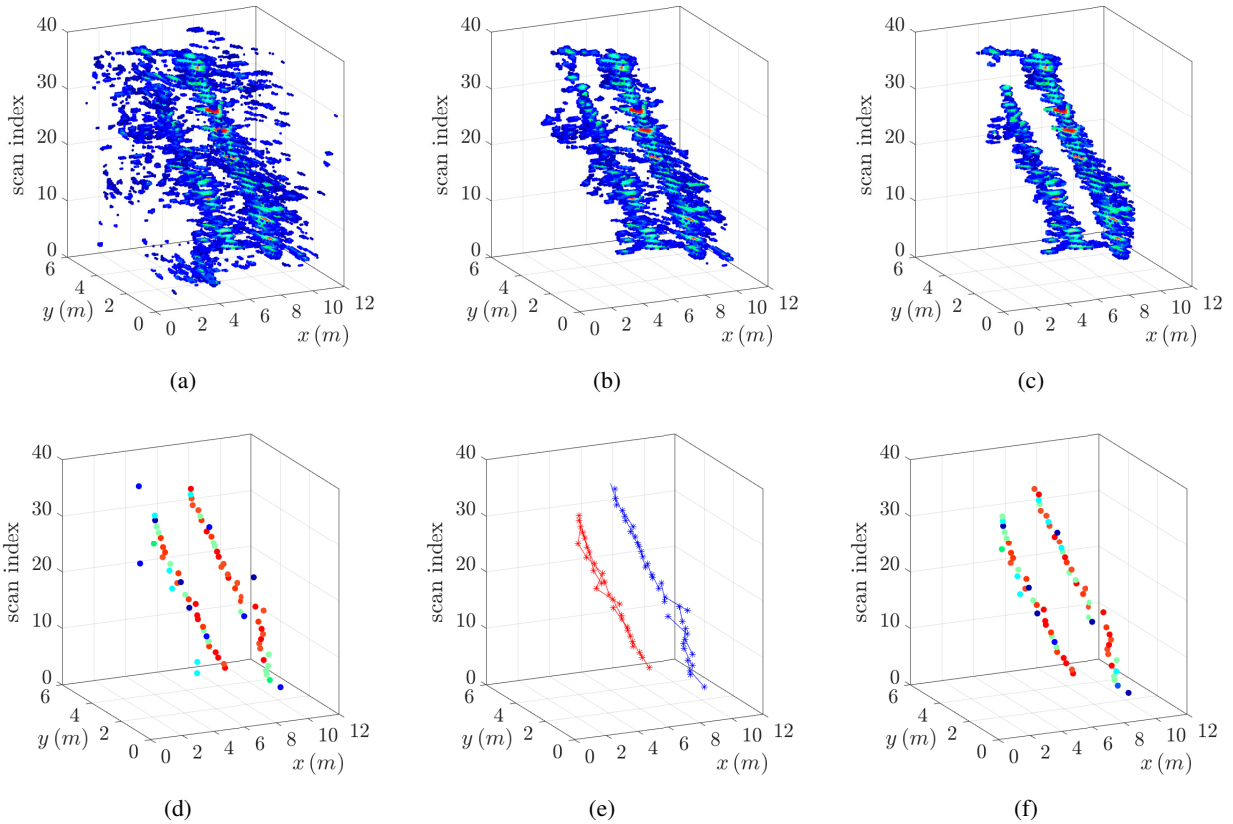


Fig. 18. (a) Example of 3D data structure applying (5) only. Red and blue colors denote the largest and the smallest score, respectively. (b) The same data structure obtained by applying 3D region growing. (c) The same data structure obtained by applying also 3D opening operation. (d) The points obtained with the proposed generation method, and (e) the corresponding tracklets. (f) Trajectory points after tracklet association (before outlier removal and trajectory smoothing).

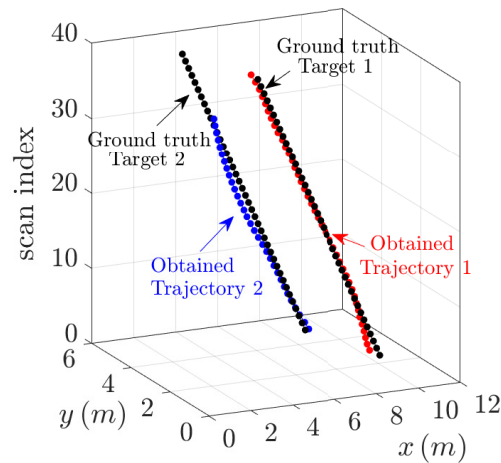


Fig. 19. The ground truth trajectory (black ball) and the smoothed one. Red balls indicate smoothed trajectory of target 1; blue balls indicate smoothed trajectory of target 2.

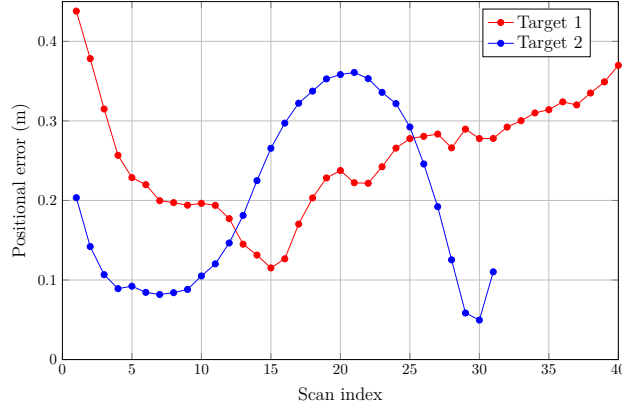


Fig. 20. Final positioning error (smoothed trajectories) through all scans where targets are detected.

clutter; the corresponding clutter-suppressed score map \hat{S}^t is shown in Fig. 17(d). As it is possible to observe, the score map after clutter suppression is much neater, and the target cells are better emphasized with respect to the surrounding ones.

As done for the first experiment, we also show the result of the several processing steps through all scan periods for the second one. This is depicted in Fig. 18. Specifically, we illustrate how the 3D data structure M^t would appear in the last scan period without application of 3D region growing and opening operation (Fig. 18(a)); how it would appear with application of 3D region growing only (Fig. 18(b)); how it appears by applying both techniques (Fig. 18(c)). Scores are indicated with colors. Comparing Fig. 18(a) and Fig. 18(c), we see that a number of isolated cells and boundary cells are removed. The points generated through all scan periods by application of the technique proposed in Section III-D are depicted in Fig. 18(d), where colors are used to indicate the point scores $p_{i,j}^k$ in (12); all tracklets detected by processing the generated points are shown in Fig. 18(e); all points obtained after tracklet association (before outlier removal and trajectory smoothing) are depicted in Fig. 18(f), where we can observe the presence of just a few residual false alarm points to be rejected by outliers removal.

Two trajectories are obtained after tracklet association. The two smoothed trajectories are presented in Fig. 19, in red for target 1 and in blue for target 2. The ground truth of each target is also depicted in black. As we can observe, target 1 is detected in all 40 scans; target 2 is also detected up to scan 31 after which, however, it is lost. Actually, as from Fig. 18(d), no point was extracted from this target in the last scan periods due to extremely low echoes. This phenomenon is explained by observing, from Fig. 15, that target 2 is shadowed by target 1 in the last scan periods, i.e., target 1 is between the transmitter and target 2. The average detection rate of the two targets is 88.75%. Target 1 is achieved through all scan periods; the detection rate equals 100% up to scan 31. Remarkably, no false alarm trajectories are built

TABLE VIII
RESULTS OF CASE STUDY 2

| | P_d^1 & E_p^1 | P_d^2 & E_p^2 | N_{FA} | OSPA |
|---------------------------------------|-------------------|--------------------|----------|-------|
| Proposed method ($\Delta=0.1m$) | 40/40 & 0.254 | 31/40 & 0.197 | 0 | 0.282 |
| Point generation [5] + TBD [20] | 8/40 & 0.463 | 0/40 & \emptyset | 0 | 0.676 |
| Point generation [5] + PHD filter [8] | 5/40 & 0.483 | 1/40 & 0.497 | 7/40 | 0.657 |

despite the heavy clutter affecting the initial processing stages.

The positioning error for each of the two targets is provided in Fig. 20 as a function of the scan index. The average positioning error of target 1 (through all scans) and target 2 (up to scan 31) are 25.4 cm and 19.7 cm, respectively. The positioning error values are generally very satisfactory, always below 40 cm except for target 1 in the first scan period. To provide a benchmark for the achieved performance, we can compare the proposed approach with the method described in [5]; this method works very well, even in indoor scenarios, when target echoes are relatively strong but is not specifically tailored to address weak targets. In this second experiment, target 2 generates very weak echoes, and with the same UWB waveforms, the method of [5] was not able to detect it in any scan. The proposed technique compares favorably with the benchmark, revealing its potential in the framework of UWB detection and tracking. Its better performance with weak targets outcomes from the application of a TBD approach, featuring the judicious combination of low thresholds in score map generation (in order not to cancel out echoes from weak targets) and of a sliding window based temporal processing, including generation of points, tracklet detection, and tracklet association, that allows handling a considerable number of false alarms.

The overall results of Case Study 2 are presented in Table VIII. The detection rate of two benchmark methods is lower than that achieved by the proposed technique, since two targets exist in this case study: The presence of multiple target yields a deterioration in point generation [5]. As from Table VIII, the proposed method can outperform the other two approaches also under a multistatic configuration.

V. CONCLUSIONS

In this paper, we proposed a complete TBD processing chain for (monostatic or multistatic) UWB RSNs, able to cope with weak targets. As opposed to track-after-detect approaches, in which target detection is first performed exploiting measurements collected by the several sensors in the same scan period and then detection points are processed by a tracking filter, the temporal dimension is here introduced in the detection process. Waveforms, pre-processed by the spatially distributed UWB sensors, are jointly processed for score map generation under a lower threshold for weak targets. Score maps are then jointly processed over the time, in a sliding window fashion, to generate points, tracklets, and finally

trajectories. The effectiveness of the proposed method has been verified with actual measurements (in outdoor and indoor environments) and human targets. Trajectory confirmation using several measurements over a time window turned out to be very beneficial to weak target detection and clutter suppression.

APPENDIX A

CLUTTER SUPPRESSION IN SCORE MAPS

The proposed clutter removal procedure relies on the score maps collected through the past $U(t) = t - \max\{1, t - V\}$ scans for some $V > 0$, namely, $\mathbf{S}^{t-U(t):t-1} = \{\mathbf{S}^{t-U(t)}, \mathbf{S}^{t-U(t)+1}, \dots, \mathbf{S}^{t-1}\}$. A *clutter density map* is obtained by computing the average grid cell scores through the past $U(t)$ scans. Denoting this map by $\mathbf{C} = \{C(i_x, i_y)_{1 \leq i_x \leq N_x, 1 \leq i_y \leq N_y}\}$, we have

$$C(i_x, i_y) = \frac{1}{U(t)} \sum_{i=1}^{U(t)} S^{t-1}(i_x, i_y).$$

The score map at scan t after clutter suppression, denoted by $\hat{\mathbf{S}}^t = \{\hat{S}^t(i_x, i_y)_{1 \leq i_x \leq N_x, 1 \leq i_y \leq N_y}\}$, is now computed as $\hat{S}^t(i_x, i_y) = S^t(i_x, i_y)/C(i_x, i_y)$. The score map $\hat{\mathbf{S}}^t$ is then employed instead of the original one \mathbf{S}^t in the subsequent processing steps.

APPENDIX B

3D REGION GROWING

The 3D region growing algorithm described in Section III-C is formalized in Algorithm 1. In the pseudocode, the generic cell (element of the input 3D data structure $\mathbf{M}^{(t-W+1):t}$) is denoted by Z and the generic seed cell by Z_s . The set of seed cells is denoted by \mathcal{S} . Moreover, for any cell Z , $\mathcal{N}_+(Z)$ represents the set of all cells that are neighbors of Z and whose score is nonzero. Note that there are at most 26 such neighbors. The status of a seed cell Z_s is denoted by $\Xi(Z_s)$ and can take two values, namely, $\Xi(Z_s) = 1$ (active seed cell) and $\Xi(Z_s) = 0$ (inactive seed cell). In the beginning, all seed cells are set to active (lines 2-4 in the algorithm).

Seed cells can be processed in any order. For each seed cell Z_s , a new region (containing in the beginning just the seed cell) is initialized if Z_s is active and its score is nonzero (lines 6-8). Moreover, the set \mathcal{P} is initialized with the neighbors of Z_s having a nonzero score (line 9). The cycle at lines 10-18 performs the following processing. In every iteration, all cells in \mathcal{P} are processed in any order. For each such cell Z , all neighbors of Z with a nonzero score, except the ones that are already in \mathcal{P} or in \mathcal{R}_n^t , are accumulated in the set \mathcal{T} (line 13), and then the cell Z is moved from \mathcal{P} to \mathcal{R}_n^t (lines 14-15). This way, at the end of each iteration of the while cycle, the set \mathcal{T} contains all new neighbors, not discovered before,

Algorithm 1: 3D Region Growing**Data:** 3D $(N_x \times N_y \times W)$ data structure $M^{(t-W+1):t}$, set of seed cells \mathcal{S} **Result:** $R^t, \{\mathcal{R}_1^t, \mathcal{R}_2^t, \dots, \mathcal{R}_{R^t}^t\}$

```

1  $n \leftarrow 0; \mathcal{Q} \leftarrow \{\};$ 
2 forall ( $Z_s \in \mathcal{S}$ ) do
3    $\Xi(Z_s) \leftarrow 1;$ 
4 end
5 forall ( $Z_s \in \mathcal{S}$ ) do
6   if ( $(M(Z_s) > 0) \wedge (\Xi(Z_s) \neq 0)$ ) then
7      $n \leftarrow n + 1;$ 
8      $\mathcal{R}_n^t \leftarrow \{Z_s\};$ 
9      $\mathcal{P} \leftarrow \mathcal{N}_+(Z_s);$ 
10    while ( $\mathcal{P} \neq \{\}$ ) do
11       $\mathcal{T} \leftarrow \{\};$ 
12      forall ( $Z \in \mathcal{P}$ ) do
13         $\mathcal{T} \leftarrow \mathcal{T} \cup (\mathcal{N}_+(Z) \setminus (\mathcal{P} \cup \mathcal{R}_n^t));$ 
14         $\mathcal{R}_n^t \leftarrow \mathcal{R}_n^t \cup \{Z\};$ 
15         $\mathcal{P} \leftarrow \mathcal{P} \setminus \{Z\};$ 
16      end
17       $\mathcal{P} \leftarrow \mathcal{P} \cup \mathcal{T};$ 
18    end
19    if ( $\mathcal{R}_n^t$  satisfies (7) and (8)) then
20       $\Xi(Z_s) = 0 \forall Z_s \text{ s.t. } Z_s \in \mathcal{R}_n^t;$ 
21       $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{\mathcal{R}_n^t\};$ 
22    end
23    else
24      forall ( $Z \in \mathcal{R}_n^t$ ) do
25         $M(Z) \leftarrow 0;$ 
26      end
27       $n \leftarrow n - 1;$ 
28    end
29  end
30 end
31 forall ( $Z \in M^{(t-W+1):t}$  s.t.  $Z \notin \cup_{i=1}^n \mathcal{R}_i^t$ ) do
32    $M(Z) \leftarrow 0$ 
33 end
34 return  $n, \mathcal{Q}$ 

```

with a nonzero score. All of these new cells are moved to \mathcal{P} (line 17). The cycle terminates whenever \mathcal{P} equals the empty set, meaning that no new neighbors with a nonzero score have been discovered.

The region \mathcal{R}_n^t grown from Z_s is now tested (lines 19-22): If it fulfills the conditions (7) and (8), then the region is included in the set of output regions \mathcal{Q} and the status of all seed cells in it is set to inactive, so that these seed cells will not be processed. Otherwise (lines 23-28), the region is rejected and the score of all of its cells is forced to zero. Before returning the number of obtained regions and

their set, the score of all cells not belonging to any of the the regions id forced to zero.

REFERENCES

- [1] B. Yan, A. Giorgetti, and E. Paolini, "A track-before-detect approach for UWB radar sensor networks," in *Proc. 2020 IEEE Radar Conf.*, Florence, Italy, Sep. 2020.
- [2] M. Z. Win, A. Conti, S. Mazuelas, Y. Shen, W. M. Gifford, D. Dardari, and M. Chiani, "Network localization and navigation via cooperation," *IEEE Commun. Mag.*, vol. 49, no. 5, pp. 56–62, May 2011.
- [3] P. Withington, H. Fluhler, and S. Nag, "Enhancing homeland security with advanced UWB sensors," *IEEE Microwave*, vol. 4, no. 3, pp. 51–58, Sep. 2003.
- [4] S. Bartoletti, A. Conti, A. Giorgetti, and M. Z. Win, "Sensor radar networks for indoor tracking," *IEEE Wireless Commun. Lett.*, vol. 3, no. 2, pp. 157–160, Apr. 2014.
- [5] M. Chiani, A. Giorgetti, and E. Paolini, "Sensor radar for object tracking," *Proc. IEEE*, vol. 106, no. 6, pp. 1022–1041, Jun. 2018.
- [6] B. Sobhani, E. Paolini, A. Giorgetti, M. Mazzotti, and M. Chiani, "Target tracking for UWB multistatic radar sensor networks," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 1, pp. 125–136, Feb. 2014.
- [7] Y. He, P. Aubry, F. L. Chevalier, and A. Yarovoy, "Keystone transform based range-Doppler processing for human target in UWB radar," in *Proc. IEEE Radar Conf.*, Cincinnati, OH, USA, May 2014, pp. 1–7.
- [8] K. Granstrom, C. Lundquist, and O. Orguner, "Extended target tracking using a Gaussian-mixture PHD filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.
- [9] K. Granström, U. Orguner, R. Mahler, and C. Lundquist, "Corrections on: "Extended target tracking using a Gaussian-mixture PHD filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 1055–1058, Apr. 2017.
- [10] K. Granström, A. Natale, P. Braca, G. Ludeno, and F. Serafino, "Gamma Gaussian inverse Wishart probability hypothesis density for extended target tracking using X-band marine radar data," *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 12, pp. 6617–6631, Dec. 2015.
- [11] C. Lundquist, K. Granström, and U. Orguner, "An extended target CPHD filter and a gamma Gaussian inverse Wishart implementation," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 472–483, Jun. 2013.
- [12] M. R. Leonard and A. M. Zoubir, "Multi-target tracking in distributed sensor networks using particle PHD filters," *Signal Processing*, vol. 159, pp. 130–146, 2019.
- [13] L. Chai, L. Kong, S. Li, and W. Yi, "The multiple model multi-Bernoulli filter based track-before-detect using a likelihood based adaptive birth distribution," *Signal Processing*, vol. 171, p. 107501, 2020.
- [14] N. C. Mohanty, "Computer tracking of moving point targets in space," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 3, no. 5, pp. 606–611, Sep. 1981.
- [15] J. Arnold, S. W. Shaw, and H. Pasternack, "Efficient target tracking using dynamic programming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 44–56, Jan. 1993.
- [16] S. M. Tonissen and Y. Bar-Shalom, "Maximum likelihood track-before-detect with fluctuating target amplitude," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 796–809, Jul. 1998.
- [17] I. Hyounjun and K. Taejeong, "Optimization of multiframe target detection schemes," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 176–187, Jan. 1999.
- [18] S. Buzzi, M. Lops, and L. Venturino, "Track-before-detect procedures for early detection of moving target from airborne radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 937–954, Jul. 2005.
- [19] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA: Artech House, 1999.

- [20] B. Yan, E. Paolini, N. Xu, Z. Sun, and L. Xu, "Multiple maneuvering extended targets detection by 3D projection and tracklet association," *Signal Processing*, vol. 179, p. 107821, 2021.
- [21] B. Ristic, L. Rosenberg, D. Y. Kim, and R. Guan, "Bernoulli filter for tracking maritime targets using point measurements with amplitude," *Signal Processing*, vol. 181, p. 107919, 2021.
- [22] L. Ubeda-Medina, Á. F. García-Fernández, and J. Grajal, "Adaptive auxiliary particle filter for track-before-detect with multiple targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 5, pp. 2317–2330, Oct. 2017.
- [23] W. Yi, L. Fu, Á. F. García-Fernández, L. Xu, and L. Kong, "Particle filtering based track-before-detect method for passive array sonar systems," *Signal Processing*, vol. 165, pp. 303–314, 2019.
- [24] B. Errasti-Alcala and P. Braca, "Track before detect algorithm for tracking extended targets applied to real-world data of X-band marine radar," in *Proc. 17th Int. Conf. Information Fusion*, Salamanca, Spain, Jul. 2014.
- [25] B. Yan, L. Xu, M. Li, and J. Z. Yan, "Track-before-detect algorithm based on dynamic programming for multi-extended-targets detection," *IET Signal Processing*, vol. 11, no. 6, pp. 674–686, Aug. 2017.
- [26] L. Wei, X. Zhang, and L. Fan, "A TBD algorithm based on improved randomized Hough transform for dim target detection," in *Proc. Int. Conf. on Signal Proc. Systems*, vol. 2, Dalian, China, 2010, pp. 241–245.
- [27] W. Yi, M. R. Morelande, L. Kong, and J. Yang, "An efficient multi-frame track-before-detect algorithm for multi-target tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 421–434, Jun. 2013.
- [28] B. D. Carlson, E. D. Evans, and S. Wilson, "Search radar detection and track with the Hough transform – I. System concept," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 1, pp. 102–108, Jan. 1994.
- [29] L. R. Moyer, J. Spak, and P. Lamanna, "A multi-dimensional Hough transform-based track-before-detect technique for detecting weak targets in strong clutter backgrounds," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 4, pp. 3062–3068, Oct. 2011.
- [30] B. Yan, N. Xu, G. Wang, S. Yang, and L. Xu, "Detection of multiple maneuvering extended targets by three-dimensional Hough transform and multiple hypothesis tracking," *IEEE Access*, vol. 7, pp. 80717–80732, Jun. 2019.
- [31] B. Yan, X. Y. Zhao, N. Xu, Y. Chen, and W. B. Zhao, "A Grey Wolf optimization-based track-before-detect method for maneuvering extended target detection and tracking," *Sensors*, vol. 19, no. 7, pp. 1–21, 2019.
- [32] E. Grossi, M. Lops, and L. Venturino, "A heuristic algorithm for track-before-detect with thresholded observations in radar systems," *IEEE Signal Process. Lett.*, vol. 20, no. 8, pp. 811–814, Aug. 2013.
- [33] H. Wang, J. Yi, and X. Wan, "Greedy algorithm-based track-before-detect in radar systems," *IEEE Sensors J.*, vol. 18, no. 17, pp. 7158–7165, Sep. 2018.
- [34] J. Kwon, N. Kwak, E. Yang, and K. Kim, "Particle filter based track-before-detect method in the range-doppler domain," in *Proc. IEEE Radar Conf.*, Boston, MA, USA, Apr. 2019, pp. 1–5.
- [35] D. J. Salmond and H. Birch, "A particle filter for track-before-detect," in *Proc. American Control Conf.*, Arlington, VA, USA, Jun. 2001, pp. 3755–3760.
- [36] W. Yi, Z. Fang, W. Li, R. Hoseinnezhad, and L. Kong, "Multi-frame track-before-detect algorithm for maneuvering target tracking," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4104–4118, 2020.
- [37] W. Li, W. Yi, M. Wen, and D. Orlando, "Multi-PRF and multi-frame track-before-detect algorithm in multiple PRF radar system," *Signal Processing*, vol. 174, p. 107648, 2020.
- [38] A. Aprile, E. Grossi, M. Lops, and L. Venturino, "Track-before-detect for sea clutter rejection: Tests with real data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 3, pp. 1035–1045, Jun. 2016.
- [39] L. Wang, G. Zhou, and P. Li, "A complex pseudo-spectrum based velocity filtering method for track-before-detect," *Signal Processing*, vol. 174, p. 107651, 2020.

- [40] H. Jiang, W. Yi, G. Cui, L. Kong, and X. Yang, "Track-before-detect strategies for range distributed target detection in compound-Gaussian clutter," *Signal Processing*, vol. 120, pp. 462–467, 2016.
- [41] C. Cao, Y. Zhao, X. Pang, B. Xu, and Z. Suo, "Sequential Monte Carlo cardinalized probability hypothesized density filter based on track-before-detect for fluctuating targets in heavy-tailed clutter," *Signal Processing*, vol. 169, p. 107367, 2020.
- [42] S.-Y. Wan and W. E. Higgins, "Symmetric region growing," *IEEE Trans. Image Processing*, vol. 12, no. 9, pp. 1007–1015, Sep. 2003.
- [43] J. A. Maintz, P. A. van den Elsen, and M. A. Viergever, "3D multimodality medical image registration using morphological tools," *Image and Vision Computing*, vol. 19, no. 1-2, pp. 53–62, Jan. 2001.
- [44] E. Grinzato and V. Vavilov, "Corrosion evaluation by thermal image processing and 3D modelling," *Revue Générale de Thermique*, vol. 37, no. 8, pp. 669–679, Sep. 1998.
- [45] R. Gonzalez, R. Woods, and B. Masters, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2008.
- [46] C.-Y. Chong, S. Mori, W. H. Barker, and K.-C. Chang, "Architectures and algorithms for track association and fusion," *IEEE Aerosp. Electron. Syst. Mag*, vol. 15, no. 1, pp. 5–13, Jan. 2000.
- [47] B. Ristic, B.-N. Vo, and D. Clark, "Performance evaluation of multi-target tracking using the OSPA metric," in *Proc. 13th Int. Conf. Information Fusion*, Edinburgh, UK, Jul. 2010.
- [48] Y. Chen, L. Xu, B. Yan, and C. Li, "A novel smooth variable structure smoother for robust estimation," *Sensors*, vol. 20, no. 6, p. 1781, 2020.