ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Map Fusion and Heterogeneous Objects Tracking in Joint Sensing and Communication Networks

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

Published Version:
Favarelli, E., Matricardi, E., Pucci, L., Paolini, E., Xu, W., Giorgetti, A. (2023). Map Fusion and Heterogeneous Objects Tracking in Joint Sensing and Communication Networks [10.23919/EuRAD58043.2023.10289614].

Availability:
This version is available at: https://hdl.handle.net/11585/960096 since: 2024-02-23

Published:
DOI: http://doi.org/10.23919/EuRAD58043.2023.10289614

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)

07 October 2026

Map Fusion and Heterogeneous Objects Tracking in Joint Sensing and Communication Networks

Elia Favarelli[#], Elisabetta Matricardi[#], Lorenzo Pucci[#], Enrico Paolini[#], Wen Xu[*], Andrea Giorgetti[#]

[#]Wireless Communications Laboratory, CNIT, DEI, University of Bologna, Italy

[*]Munich Research Center, Huawei Technologies Duesseldorf GmbH, Germany

{elia.favarelli, elisabett.matricard3, lorenzo.pucci3, e.paolini, andrea.giorgetti}@unibo.it, wen.dr.xu@huawei.com

***Abstract*— We consider a joint sensing and communication (JSC) network where a set of base stations (BSs) cooperates through a fusion center (FC). The network is capable of tracking a set of heterogeneous objects and simultaneously communicating with a set of users. In particular, each BS acts as a monostatic radar that can scan the surrounding area, producing range-angle maps that are exploited to infer the targets positions. Such maps are fused at the FC, and then tracking algorithms, i.e., particle filter (PF) and multi-hypothesis tracker (MHT), are applied to estimate the targets states. Numerical results based on 5G new radio (NR) signals show that BSs are capable of serving user equipments (UEs) with a communication capacity of 850 Mbit/s while their cooperation for the radar task allows to localize point-like and extended targets with a root mean squared error (RMSE) less than 15 cm and 80 cm, respectively.**

I. Introduction

The upcoming generation of mobile radio networks will support various emerging functionalities, including radically new services. One such feature is the ability to perform effective sensing via radio frequency signals, made possible by the evolution toward higher frequency bands and larger antenna arrays [1], [2]. The interest in sensing is triggered by its ability to support a variety of applications, such as traffic monitoring, autonomous driving, and environment mapping. Additionally, with the advent of massive multiple-input multiple-output (MIMO) technology at millimeter-wave (mmWave), it will be possible to achieve not only high data rates but also very accurate position estimation of objects.

This work aims to investigate the possibility of using multi-sensor fusion techniques combined with tracking algorithms, to exploit range-angle radar maps obtained through a set of cooperating base stations (BSs) that transmit a waveform compliant with the 5G new radio (NR) standard. The main contributions can be summarized as follows: *i)* We propose a soft map fusion strategy based on range-angle maps obtained at the BSs. *ii)* We present a three-step clustering procedure to extract reliable detections from the fused soft maps. *iii)* We evaluate two different tracking strategies, particle filter (PF) and multi-hypothesis tracker (MHT), based on direct soft maps sampling, and on excision and clustering procedures, respectively. *iv)* The root mean squared error (RMSE) and capacity of the joint sensing and communication (JSC) system are evaluated to assess the radar and communication capabilities. Moreover, we investigate the impact of key resource parameters, i.e., the fraction of power, the fraction of subcarriers, and the fraction of time dedicated to the radar function.

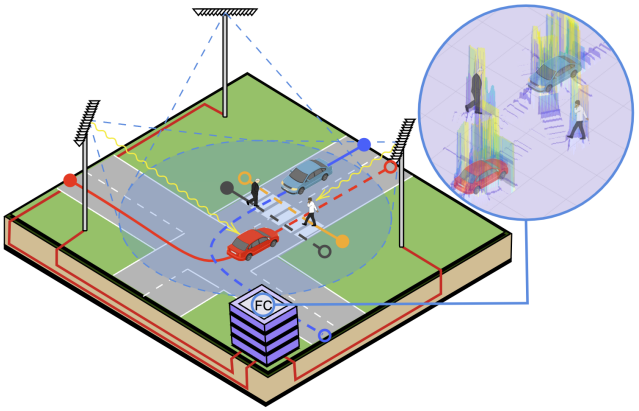

Fig. 1. An urban scenario with monostatic JSC BSs aiming at monitoring two pedestrians (point-like targets) and two vehicles (extended targets). BSs are communicating with UEs (represented with yellow wavy arrows) while simultaneously sensing the surrounding environment (blu area). Target trajectories are reported with red, blue, gray, and orange paths.

Capital and lesscase boldface letters denote matrices and vectors, respectively; $\mathbf{I}_n$ is the $n\times n$ identity matrix; $\|\cdot\|_p$ stands for the p-norm; $\delta(\cdot)$ is the Dirac delta function; $(\cdot)^c$ stands for conjugate; $\mathbf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{\Sigma})$ denotes a zero-mean circularly symmetric complex Gaussian random vector with covariance $\mathbf{\Sigma}$, and $\mathbf{x}\sim\mathcal{N}(\boldsymbol{\mu},\mathbf{\Sigma})$ denotes the real-valued Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$.

II. System Model

We analyze a JSC network composed of monostatic MIMO BSs operating at mmWave frequencies and utilizing orthogonal frequency division multiplexing (OFDM) waveform according to the 5G NR standard. The BSs cooperate through a fusion center (FC) to perform target detection and tracking, as well as resource allocation; multi-cell interference is managed by frequency division (FD) among the BSs. Furthermore, while scanning the surrounding environment, the BSs simultaneously communicate with UEs in their respective cells. The network and scenario are portrayed in Fig. 1.

Considering the downlink communication towards a UE, a BS transmits frames consisting of M OFDM symbols and $K=\rho_{\mathrm{f}}K_0$ subcarriers according to the 5G NR resource blocks allocated to the considered user; $\rho_{\mathrm{f}}\in[0,1]$ is, therefore, the

fraction of subcarriers among the K_0 available. To scan the environment, each BS steers the sensing beam and collects a subset $M_\mathrm{s} < M$ of OFDM symbols in each direction. Therefore, the transmitted signal for each sensing direction can be represented by a matrix $\mathbf{X}_\mathrm{s} \in \mathbb{C}^{K \times M_\mathrm{s}}$ whose elements $x_k^{(m)}$ are complex modulation symbols [3].

The BSs are quipped with half-wavelength spaced uniform linear arrays (ULAs) for transmission and reception with N_T and N_R antennas, respectively. The beamforming vector $\mathbf{w}_\mathrm{T} \in \mathbb{C}^{N_\mathrm{T} \times 1}$ is applied to the transmitted symbols at each antenna, $\tilde{\mathbf{x}}_k^{(m)} = \mathbf{w}_\mathrm{T} x_k^{(m)}$, to form the communication beam (directed towards the UE) and the sensing beam (steered to scan the area), and to split the available power between them, i.e.,

$$\mathbf{w}_\mathrm{T} = \frac{\sqrt{P_\mathrm{T} G_\mathrm{T}^\mathrm{a}}}{N_\mathrm{T}} \left(\sqrt{\rho_\mathrm{p}} \mathbf{a}_\mathrm{T}^c(\theta_{\mathrm{T,s}}) + \sqrt{1-\rho_\mathrm{p}} \mathbf{a}_\mathrm{T}^c(\theta_{\mathrm{T,c}}) \right) \quad (1)$$

where $\rho_\mathrm{p} \in [0,1]$ is fraction of power reserved for the sensing beam, P_T is the transmit power, G_T^a is the transmit array gain along the beam steering direction, and $\mathbf{a}_\mathrm{T}(\theta_{\mathrm{T,c}}) \in \mathbb{C}^{N_\mathrm{T} \times 1}$ and $\mathbf{a}_\mathrm{T}(\theta_{\mathrm{T,s}}) \in \mathbb{C}^{N_\mathrm{T} \times 1}$ are the steering vectors for communication and sensing directions, $\theta_{\mathrm{T,c}}$ and $\theta_{\mathrm{T,s}}$, respectively [3].

The received symbols after OFDM demodulation are represented by $\tilde{\mathbf{y}}_k^{(m)} = \mathbf{H}_k^{(m)} \tilde{\mathbf{x}}_k^{(m)} + \tilde{\mathbf{n}}_k$, where $\mathbf{H}_k^{(m)} \in \mathbb{C}^{N_\mathrm{R} \times N_\mathrm{T}}$ is the channel matrix for the mth symbol and the kth subcarrier, and $\tilde{\mathbf{n}}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_\mathrm{N}^2 \mathbf{I}_{N_\mathrm{R}})$.[1] Spatial combining is then performed in the considered sensing direction, $\theta_{\mathrm{R,s}} = \theta_{\mathrm{T,s}}$, by using the receiving beamforming vector $\mathbf{w}_\mathrm{R} = \mathbf{a}_\mathrm{R}^c(\theta_{\mathrm{R,s}})$. This yields the grid of the received symbols $\mathbf{Y}_\mathrm{s} \in \mathbb{C}^{K \times M_\mathrm{s}}$, where $y_k^{(m)} = \mathbf{w}_\mathrm{R}^T \tilde{\mathbf{y}}_k^{(m)}$. The symbols $\mathbf{Y}_\mathrm{s}$ collected in each sensing direction are then used to generate range-angle maps.

A. Target Models

According to the scenario depicted in Fig. 1, we consider point-like targets (pedestrians) and extended objects (vehicles). Vehicles are modeled with 12 reflection points, including 4 plane reflections due to the front, back, and sides (with a narrow visibility function and large radar cross-section (RCS)), 4 accounting for the wheelhouses, and 4 to mimic the corners [4]. The channel matrix in the presence of L point reflectors is

$$\mathbf{H}_k^{(m)} = \sum_{l=1}^{L} \alpha_l e^{j2\pi m T_\mathrm{s} f_{\mathrm{D},l}} e^{-j2\pi k \Delta f \tau_l} \mathbf{a}_\mathrm{R}(\theta_l) \mathbf{a}_\mathrm{T}^T(\theta_l) \quad (2)$$

where T_s is the OFDM symbol duration, Δf is the subcarrier spacing, $f_{\mathrm{D},l}$, τ_l, θ_l, and $\mathbf{a}_\mathrm{R}(\theta_l)$ are the Doppler shift, the round-trip delay, the direction of arrival (DoA) of the lth back-scattered signal, respectively. The complex term $\alpha_l = |\alpha_l| e^{j\phi_l}$ includes phase shift and attenuation along the lth propagation path. The signal-to-noise ratio (SNR) at each receiving antenna, related to the lth reflection point is

$$\mathrm{SNR}_l^{(\mathrm{s})} = \rho_\mathrm{p} \cdot \gamma_l \cdot \frac{P_\mathrm{T} G_\mathrm{T}^\mathrm{a} G_\mathrm{R}}{\sigma_\mathrm{N}^2} \frac{c^2 \sigma_{\mathrm{rcs},l}}{(4\pi)^3 f_\mathrm{c}^2 d_l^4} \quad (3)$$

[1]We consider negligible inter-carrier interference (ICI) and inter-symbol interference (ISI).

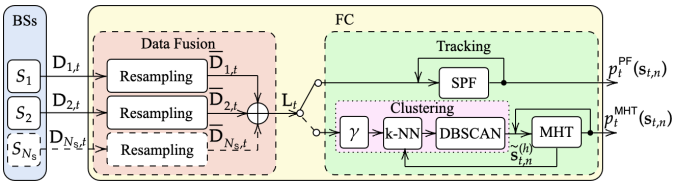

Fig. 2. Block diagram of the radar processing chain. The BSs sense the environment generating range-angle maps and share them with the FC. Range-angle maps are then resampled and fused in a single map (red block) and processed with tracking algorithms to obtain the object states estimation (green block). Note the presence of excision and clustering procedures (pink block) adopted to extract detection points for MHT.

where G_R is the single element antenna gain at the Rx, $\gamma_l = |\mathrm{AF}(\theta_{\mathrm{T,s}} - \theta_l)|^2 \in [0,1]$ is the normalized array gain at the Tx to account for the imperfect alignment between the sensing direction and the target DoA. We denote $\sigma_\mathrm{N}^2 = N_0 K \Delta f$ with N_0 the one-sided noise power spectral density (PSD), and d_l the distance between the lth reflection point and the BS and $\sigma_{\mathrm{rcs},l}$ its RCS; f_c and c are the carrier frequency and the speed of light, respectively.

B. Measurement Model

As introduced before, the communication beam is directed towards a UE, while the sensing beam changes its direction over time with predefined angular steps. The scan duration, T_scan, depends on the number of sensing directions and the symbol duration T_s. After acquiring all symbols to form the matrix $\mathbf{Y}_\mathrm{s}$, an element-wise division between $\mathbf{Y}_\mathrm{s}$ and $\mathbf{X}_\mathrm{s}$ is performed to remove the transmitted symbols, resulting in the least square channel matrix estimate $\mathbf{G}_\mathrm{s}$. The next step is to perform a double-periodogram on the rows and columns of $\mathbf{G}_\mathrm{s}$ to obtain a range-Doppler map [3]. A range-angle map $\mathbf{D}_{n,t}$ is then obtained, at the nth BS and tth scan, from the range-Doppler map by taking the column of the periodogram with the maximum value and associating it with the considered sensing direction.[2]

It should be noted that a fixed scan duration T_scan is considered, and the time between two consecutive measurements (range-angle maps update) T_meas can be varied. Thus, $\mathrm{D} = T_\mathrm{meas}/T_\mathrm{scan}$ represents the decimation factor, and $\rho_\mathrm{t} = 1/\mathrm{D}$ represents the fraction of time allocated for sensing (i.e., with sensing beam turned on).

III. Data Processing

Following the block diagram in Fig. 2, each BS shares the range-angle maps $\mathbf{D}_{n,t}$ with the FC. At the FC, a linear uniform grid with resolution Δ_x and Δ_y (with N_x and N_y points) is used as a baseline, and maps are resampled in those points, ensuring consistent map fusion. Then, the resampled range-angle maps $\overline{\mathbf{D}}_{s,t}$ are fused using element-wise summation to obtain the soft map $\mathbf{L}_t = \sum_{s=1}^{N_\mathrm{s}} \overline{\mathbf{D}}_{s,t}$.[3]

[2]Since the inference of target parameters turns out to be a frequency estimation problem, the periodogram represents (asymptotically) the log-likelihood, hence, the column with the maximum value is chosen.

[3]$\mathbf{L}_t$ can be roughly interpreted as a target likelihood map.

A. Clustering

A three-step clustering procedure is implemented to extract detections from the soft maps and manage extended objects. First, an excision filter with threshold γ is applied to the matrix $\mathbf{L}_t$, to sift map points with higher values. Second, a k-nearest neighbors (k-NN) algorithm with $k = 1$ and gate γ_{n} (to ignore residual points distant from each target) is applied to cluster data that likely belong to a previously detected target [5]. Finally, the remaining points (i.e., map points higher than γ and outside the gate γ_{n}) are clustered through the density-based spatial clustering of applications with noise (DBSCAN) algorithm, with a maximum distance between points belonging to the same cluster γ_{d}, and a minimum number of points to form a cluster N_{d} [6]. Each cluster centroid is then stored in the matrix $\mathbf{Z}_t$ that contains target detections extracted from the soft maps.

B. Tracking

To exploit soft likelihood maps, PF has been selected as a good candidate to perform multiple target tracking [7]. Let us consider a set of N_{obj} objects with state vectors $\mathbf{s}_{t,n}$ containing targets position coordinates and velocity components, where t and n are the time and target indexes. Let us also assume that the initial target states $\mathbf{s}_{0,n}$, and the number of objects is known. At the time $t = 1$, a set of particles $\mathbf{s}_{t-1,n}^{(q)}$, where q is the particle index, is generated for each target to approximate its prior state distribution. The target state distribution can be written as

$$p_{t-1}^{\mathrm{PF}}(\mathbf{s}_{t-1,n}) = \sum_{q=1}^{N_{\mathrm{p}}} w_{t-1,n}^{(q)} \delta\left(\mathbf{s}_{t-1,n} - \mathbf{s}_{t-1,n}^{(q)}\right) \qquad (4)$$

where N_{p} is the number of particles and $w_{t-1}^{(q)}$ is the weight of the qth particle at time $t-1$. All particle weights $w_{t-1,n}^{(q)}$ are initialized to $1/N_{\mathrm{p}}$. In the prediction step, particles are propagated using a linear transition matrix and additive Gaussian noise, i.e., $\mathbf{s}_{t,n}^{(q)} = \mathbf{F}\,\mathbf{s}_{t-1,n}^{(q)} + \boldsymbol{\nu}_{t,n}$, where $\mathbf{F}$ is the transition matrix and $\boldsymbol{\nu}_{t,n} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{\mathrm{PF}})$ with $\mathbf{Q}^{\mathrm{PF}}$ the prediction covariance matrix. Particles coordinates are then discretized with the same maps grid resolution Δ_{x} and Δ_{y}.

Particles weights are then updated with the values of $\mathbf{L}_t$ corresponding to the position coordinates of the relative particles, and normalized according to $w_{t,n}^{(q)} = \widetilde{w}_{t,n}^{(q)} \big(\sum_{q=1}^{N_{\mathrm{p}}} \widetilde{w}_{t,n}^{(q)}\big)^{-1}$, where $\widetilde{w}_{t,n}^{(q)}$ represents unnormalized weights. To prevent weights drift, a resampling procedure is performed at each time step [8]. Finally, posterior target state estimation $\hat{\mathbf{s}}_{t,n}^{\mathrm{PF}}$ is extracted using (4).

An alternative to PF that directly samples the matrix $\mathbf{L}_t$, is the MHT [9]. This algorithm can be used in the framework of Bayesian filtering approaches, where the prior target distribution is assumed to be a Gaussian mixture, namely,

$$p_{t-1}^{\mathrm{MHT}}(\mathbf{s}_{t-1,n}) = \sum_{h=1}^{N_{\mathrm{h}}} w_{t-1,n}^{(h)} \mathcal{N}(\mathbf{s}_{t-1,n}^{(h)}, \mathbf{P}_{t-1,n}^{(h)}) \qquad (5)$$

where N_{h} represents the number of components or hypotheses, $w_{t-1,n}^{(h)}$ its weight, and $\mathbf{s}_{t-1,n}^{(h)}$ and $\mathbf{P}_{t-1,n}^{(h)}$ the mean value and covariance of the hth hypothesis, respectively. Starting from this configuration, a Kalman linear prediction is performed to infer the target position in the consecutive time step, i.e., $\widetilde{\mathbf{s}}_{t,n}^{(h)} = \mathbf{F}\,\widetilde{\mathbf{s}}_{t-1,n}^{(h)}$, and $\widetilde{\mathbf{P}}_{t,n}^{(h)} = \mathbf{F}\,\mathbf{P}_{t-1,n}^{(h)}\,\mathbf{F}^{\mathrm{T}} + \mathbf{Q}^{\mathrm{MHT}}$, where $\mathbf{Q}^{\mathrm{MHT}}$ represent the MHT prediction covariance matrix. MHT prediction is then corrected through the Kalman update procedure, exploiting the hard measurements $\mathbf{Z}_t$, their covariance matrix $\mathbf{R}$, and considering the most likely data association hypotheses [9]. In MHT, the clutter intensity λ and the detection probability P_{D} are key parameters to need to be set. Maximum a posteriori (MAP) criterion is used to estimate targets position $\hat{\mathbf{s}}_{t,n}^{\mathrm{MHT}}$ from the distribution $p_t^{\mathrm{MHT}}(\mathbf{s}_{t,n})$.

IV. Numerical Results

We considered a scenario with three BSs, two extended targets (vehicles), and two point-like targets (pedestrians). Targets motions are illustrated in Fig. 1; pedestrians move with uniform linear motion, whereas vehicles move with a combination of constant velocity turn, linear accelerated/decelerated, static, and uniform linear motions. The area monitored is of size $x \in [-20, 20]\,\mathrm{m}$ and $y \in [-20, 20]\,\mathrm{m}$. The transmission parameters are: QPSK modulation, $f_{\mathrm{c}} = 28\,\mathrm{GHz}$, $\Delta f = 120\,\mathrm{kHz}$, $K_0 = 3168$ (i.e., about $400\,\mathrm{MHz}$ bandwidth), $M = 1120$, and $M_{\mathrm{s}} = 112$. The effective isotropic radiated power (EIRP) is set to $30\,\mathrm{dBm}$, and the noise PSD is $N_0 = 4 \cdot 10^{-20}\,\mathrm{W/Hz}$. All the three BSs are equipped with $N_{\mathrm{T}} = N_{\mathrm{R}} = 50$ antennas.

The scan at each BS has a duration of $T_{\mathrm{scan}} = 50\,\mathrm{ms}$, and the overall scene is monitored for $10\,\mathrm{s}$, resulting in $N_{\mathrm{m}} = 200$ measurements (maps collected). Grid resolution is set to $\Delta_{\mathrm{x}} = 0.1\,\mathrm{m}$ and $\Delta_{\mathrm{y}} = 0.1\,\mathrm{m}$. The amount of power dedicated to sensing is set to $\rho_{\mathrm{p}} = 0.5$. The performance are derived with D ranging from 1 to 5, and varying the fraction of subcarriers ρ_{f} from 0.1 to 1. The tracking performance of the nth target is evaluated using the $\mathrm{RMSE} = \big(\frac{1}{N_{\mathrm{m}}} \sum_{t=1}^{N_{\mathrm{m}}} \|\mathbf{s}_{t,n} - \hat{\mathbf{s}}_{t,n}\|_2^2\big)^{1/2}$, where $\mathbf{s}_{t,n}$ and $\hat{\mathbf{s}}_{t,n}$ are the true and estimated target states.

The aggregate capacity of the BS in downlink, considering that some resources are devoted to sensing is

$$\begin{aligned} C(\rho_{\mathrm{f}}, \rho_{\mathrm{t}}, \rho_{\mathrm{p}}) = {} & \rho_{\mathrm{f}}\,\rho_{\mathrm{t}}\,\Delta f\,K_0\,\log_2(1 + (1-\rho_{\mathrm{p}})\,\mathrm{SNR}^{(\mathrm{c})}) \\ & + \rho_{\mathrm{f}}\,(1-\rho_{\mathrm{t}})\,\Delta f\,K_0\,\log_2(1 + \mathrm{SNR}^{(\mathrm{c})}) \\ & + (1-\rho_{\mathrm{f}})\,\Delta f\,K_0\,\log_2(1 + \mathrm{SNR}^{(\mathrm{c})}) \end{aligned} \qquad (6)$$

where $\mathrm{SNR}^{(\mathrm{c})}$ is the communication SNR experienced by the users; as a typical example we set $\mathrm{SNR}^{(\mathrm{c})} = 8\,\mathrm{dB}$ for all UEs.

The PF uses $N_{\mathrm{p}} = 1000$ particles, and resampling is applied at each iteration. The prediction covariance matrix is set to $\mathbf{Q}^{\mathrm{PF}} = \mathrm{diag}(0.25, 0.25)$. For the clustering procedure $\gamma = 10^{-7}$, and the k-NN gate is $\gamma_{\mathrm{n}} = 4\,\mathrm{m}$. The DBSCAN algorithm uses a minimum number of points per cluster $N_{\mathrm{d}} = 20$ and maximum distance $\gamma_{\mathrm{d}} = 2\,\mathrm{m}$. For the MHT algorithm, the prior covariance matrix is initialized as $\mathbf{P}_{t,n}^{(h)} = \mathrm{diag}(0.5, 0.5, 0.5, 0.5)$, the measurements covariance matrix is set to $\mathbf{R} = \mathrm{diag}(0.25, 0.25)$, the clutter

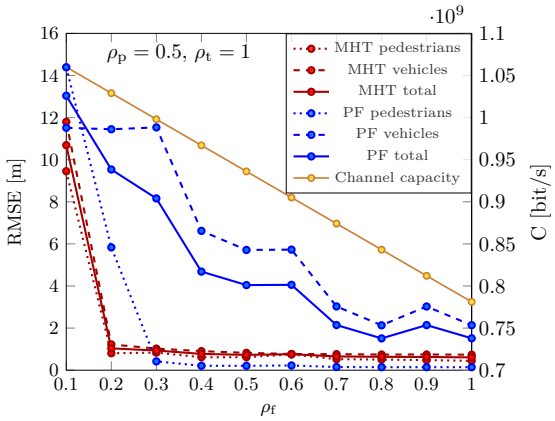

Fig. 3. Localization and communication performance by varying the fraction of frequencies ρ_f reserved for sensing.

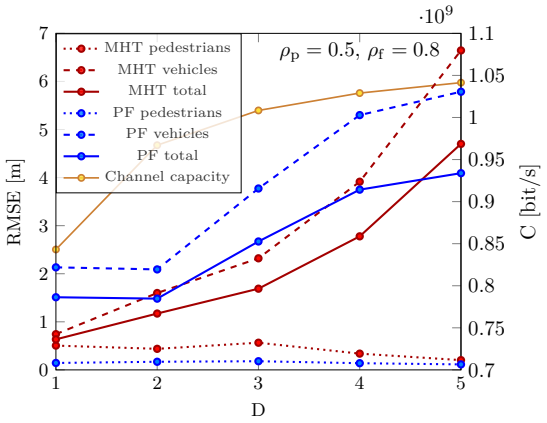

Fig. 4. Localization and communication performance by varying the decimation factor D.

intensity is $\lambda = 10^{-4}$, the detection probability is $P_D = 0.98$, and the prediction covariance matrix is $\mathbf{Q}^{\text{MHT}} = \text{diag}(2T_{\text{scan}}, 2T_{\text{scan}}, 2T_{\text{scan}}, 2T_{\text{scan}})$. The transition matrix is defined as $\mathbf{F} = (1, 0, T_{\text{scan}}, 0; 0, 1, 0, T_{\text{scan}}; 0, 0, 1, 0; 0, 0, 0, 1)$.

A. Sensing vs Communication Trade-off

In Fig. 3, the localization performance is reported by varying the fraction of subcarriers ρ_f. Dotted, dashed, and solid curves refer to point-like objects, extended objects, and the entire set of objects, respectively. As it can be seen, the MHT can track both extended and point-like targets with RMSE less than 1 m, thanks to its capability to cluster measurements originating from the same object. Differently, PF performs better with point-like objects due to its capability to directly sample soft maps, resulting in a RMSE less than 0.2 m. Notably, the MHT allows reducing ρ_f with respect to the PF.

Fig. 4 highlights the effect of sensing duty-cycle on the position estimation error. As it can be seen, the RMSE for point-like objects remains almost constant for both algorithms. In contrast, extended targets experience an increasing error for low ρ_t; indeed, extended objects represent vehicles whose motion is less predictable than pedestrians.

In both figures, the aggregate capacity of the BS is shown in yellow. As expected, by reducing the number of subcarriers devoted to sensing, the channel capacity increases linearly (see also Eq. (6)). Similarly, the capacity increases when reducing ρ_t as the resources available for communication increases.

Focusing on the overall system, with $\rho_f = 0.8$, $\rho_p = 0.5$, and reducing ρ_t from 1 to 0.5, pedestrians can still be localized with an error less than 0.2 m and vehicles with RMSE around 1.5 m, while the aggregate communication capacity of the BS increases by 13%.

V. Conclusion

This study investigated the application of multi-sensor data fusion and tracking algorithms in a JSC system using cooperative monostatic BSs. We showed that the trade-off between sensing and communication could be set by varying the available resources, i.e., the fraction of subcarriers, power, and time. Numerical results, based on the 5G NR waveform, showed that, e.g., fixing $\rho_p = 0.5$, $\rho_f = 0.8$, and $\rho_t = 1$, both tracking algorithms are effective with a RMSE less than 0.2 m for pedestrians tracked with PF, and less than 0.8 m for vehicles tracked by MHT while the aggregate capacity of the BS is reduced by 22%.

Acknowledgment

This work was supported in part by the CNIT National Laboratory WiLab and the WiLab-Huawei Joint Innovation Center and in part by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART").

References

[1] R. Thomä, T. Dallmann, S. Jovanoska, P. Knott, and A. Schmeink, "Joint communication and radar sensing: An overview," in *Europ. Conf. on Ant. and Prop. (EuCAP)*, Dusseldorf, Germany, Mar. 2021, pp. 1–5.

[2] S. Schieler, C. Schneider, C. Andrich, M. Döbereiner, J. Luo, A. Schwind, R. S. Thomä, and G. Del Galdo, "OFDM waveform for distributed radar sensing in automotive scenarios," *Int. J. of Microwave and Wireless Tech.*, vol. 12, no. 8, p. 716–722, 2020.

[3] L. Pucci, E. Paolini, and A. Giorgetti, "System-level analysis of joint sensing and communication based on 5G new radio," in *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, July 2022, pp. 2043–2055.

[4] M. Bühren and B. Yang, "Simulation of automotive radar target lists using a novel approach of object representation," in *IEEE Intelligent Veh. Symp.*, 2006, pp. 314–319.

[5] J. Watt, R. Borhani, and A. K. Katsaggelos, *Machine Learning Refined*. Cambridge University Press, 2016.

[6] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Int. Conf. on Know. Disc. in Data Mining*, Portland, Oregon, 1996, pp. 226–231.

[7] P. Djuric, J. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, 2003.

[8] J. D. Hol, T. B. Schon, and F. Gustafsson, "On resampling algorithms for particle filters," in *IEEE Non. Stat. Sig. Proc. Work.*, 2006, pp. 79–82.

[9] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, 2003.