

Covariance-based Imaging and Multi-View Fusion for Networked Sensing

Junyuan Gao, Weifeng Zhu, Yanmo Hu, Shuowen Zhang, Jiannong Cao, Yongpeng Wu, Giuseppe Caire, and Liang Liu

Abstract—This paper considers multi-view imaging in a sixth-generation (6G) integrated sensing and communication network, which consists of a transmit base-station (BS), multiple receive BSs connected to a central processing unit (CPU), and multiple extended targets. Our goal is to devise an effective multi-view imaging technique that can jointly leverage the targets’ echo signals at all the receive BSs to precisely construct the image of these targets. To achieve this goal, we propose a two-phase approach. In Phase I, each receive BS recovers an individual image based on the sample covariance matrix of its received signals. Specifically, we propose a novel covariance-based imaging framework to jointly estimate effective scattering intensity and grid positions, which reduces the number of estimated parameters leveraging channel statistical properties and allows grid adjustment to conform to target geometry. In Phase II, the CPU fuses the individual images of all the receivers to construct a high-quality image of all the targets. Specifically, we design edge-preserving natural neighbor interpolation (EP-NNI) to map individual heterogeneous images onto common and finer grids, and then propose a joint optimization framework to estimate fused scattering intensity and BS fields of view. Extensive numerical results show that the proposed scheme significantly enhances imaging performance, facilitating high-quality environment reconstruction for future 6G networks.

Index Terms—Imaging, integrated sensing and communication (ISAC), multi-view fusion, networked sensing.

I. INTRODUCTION

A. Motivation

Integrated sensing and communication (ISAC) emerges as one of the primary usage scenarios of future sixth-generation (6G) wireless systems. In 6G networks, sensing is performed by enabling devices to collect reflected signals from the targets and extract valuable information for decision-making, exploiting the same frequency band and hardware as those used for communication [1]. Due to this resource-sharing capability, 6G-oriented sensing is expected to play a crucial role in many emerging applications, such as smart transportation, augmented reality, and factory automation.

In most existing ISAC works, the sensing functionality is achieved based on the assumption of point targets. Under this

assumption, the reflected signal of a target contains information on classical “radar-type” parameters such as delay (range), angle of arrival (AoA), and Doppler (velocity), which in turn yields information for localization and tracking, irrespective of the target’s physical shape [2], [3]. However, it should be noted that the point target model remains valid only when the target size is much smaller than system resolution. In practical environments, there are a large number of targets, such as vehicles and industrial machinery, whose sizes substantially exceed resolution limits. In this case, the point target assumption does not hold, making it necessary to employ the extended target model. Compared to point targets, extended targets usually comprise numerous inseparable scattering points, rendering precise parameter estimation for individual scatterers impossible [4], [5]. Thus, for extended targets, the core sensing objective shifts to imaging - constructing a high-resolution reflectivity map of the environment, from which the shape, size, localization, and structure of targets can be estimated. Environment imaging is regarded as a key sensing functionality in 6G networks, which is therefore the topic of this work.

Compared with radar systems, in which radars usually work independently due to their sparse deployment and the absence of a globally unified standard, networked sensing is a unique advantage of 6G-oriented sensing [2]. In typical communication systems, such as cloud radio access network and cell-free multiple-input multiple-output (MIMO) network, base-stations (BSs) are densely deployed and connected via the fronthaul links for cooperative communication. This inherently enables networked sensing, where the signals reflected by targets can be received by multiple BSs and these BSs can share their local information to enhance sensing performance. It is worth noting that such cooperation is indispensable for environment imaging. Specifically, due to the occlusion effect and weak scattering at some sites, each receiver only has a limited field of view about the target, and this field is usually unknown in advance [6]–[8]. Even for imaging within this field of view, the performance is fundamentally limited by the resolution of each receiver. Therefore, constructing a high-resolution image requires global environment awareness beyond the capabilities of individual receivers. Moreover, typical impairments that traditional synthetic aperture radar (SAR) imaging must cope with, such as synchronization errors between sensors and uncertain positions (in the case of fly-over SAR imaging from planes and drones) are not problems in 6G networks, since BSs are in fixed and precisely known locations, and virtually perfect synchronization is achieved with standard fronthaul technologies, which makes the wireless network well-suited

J. Gao, W. Zhu, Y. Hu, S. Zhang, J. Cao, and L. Liu are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR (e-mail: {junyuan.gao, eee-wf.zhu, yanmo.hu, shuowen.zhang, jiannong.cao, liang-eie.liu}@polyu.edu.hk).

Y. Wu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang 200240, China (e-mail: yongpeng.wu@sjtu.edu.cn).

G. Caire is with the Communications and Information Theory Group, Technische Universität Berlin, Berlin 10587, Germany (e-mail: caire@tu-berlin.de).

for the task of networked sensing. Thus, in this paper, we present an investigation into environment imaging within the 6G networked sensing framework.

B. Related Works

Environment imaging is recognized as an indispensable sensing functionality in future 6G wireless networks, attracting significant interests from researchers. Back projection (BP) and fast Fourier transform (FFT) / inverse FFT (IFFT) algorithms have been widely adopted for imaging [9]–[11]. However, their reconstruction resolution is fundamentally constrained by physical limitations, such as antenna aperture and bandwidth. To enhance imaging performance, a number of super-resolution compressed sensing (CS) based methods have been proposed, where imaging is formulated as an inverse problem aiming at estimating instantaneous channel attenuation coefficients over a set of grid points. Specifically, the authors in [12] developed a generalized approximate message passing based imaging algorithm, where the ℓ_1 -norm regularization term was induced to exploit the sparsity of attenuation coefficients. In [13], the complex approximated message passing method was extended to an $\ell_{(1/2)}$ -norm based iterative thresholding scheme. Since discrete gradients of attenuation coefficients are sparse, i.e., discrete gradients only contain high values when the attenuation coefficients of adjacent grids are significantly different, [14] proposed to construct an image based on the total variation (TV) regularization. A sparse Bayesian learning based algorithm was proposed in [15] to solve the CS problem in environment reconstruction.

The aforementioned algorithms aim to estimate high-dimensional instantaneous channels from the echo signal in each snapshot, and thus are computationally expensive. In many scenarios, such as monitoring pedestrians or stationary infrastructure, the shape, position, and structure of extended targets exhibit little variation across multiple snapshots. As a result, a more effective approach is to reformulate the imaging task as a statistic scattering intensity estimation problem, instead of relying on instantaneous channel estimation. The multiple signal classification (MUSIC) algorithm is a classic super-resolution technique based on this principle, which, however, requires that the scatterers are uncorrelated with each other. Spatial smoothing was proposed to handle coherent scatterers [16]. However, this method remains effective only in scenarios involving a small number of correlated scatterers. For extended targets that exhibit dense clouds of scatters, the investigation of imaging based on statistic parameter estimation is still in its infancy.

Networked sensing substantially enhances sensing performance by enabling cooperation among multiple wireless nodes. Several works have explored networked sensing techniques for the localization of point targets. Specifically, the authors in [17] and [18] considered cell-free networks, where multiple BSs share their echo signals to jointly localize the targets. Effective schemes were proposed to deal with the data association issue. The authors in [19] proposed to integrate the massive communication and point target localization functionalities, where the uplink signals received at multiple

BSs were leveraged to achieve joint activity detection, channel estimation, and target localization. Networked sensing has also been leveraged to improve the imaging performance of extended targets in existing literatures. Specifically, in [20], [21], after obtaining single-view images at individual BSs via the BP technique, data fusion was performed by summing the estimates at all BSs together, under the assumption that BSs share the same field of view. However, in practice, receivers usually observe different scattering profiles of extended targets due to unknown occlusion effect and weak scattering at some sites, and thus their fields of view can be distinct and unknown in advance [7]. Under this condition, a clustering method was employed in [6] to identify the subset of receivers having high-confidence estimates at a given grid point, subsequently performing weighted averaging over this subset, which, however, relies on the Gaussian mixture assumption on local estimates. For general local estimates without this assumption, the authors in [8] proposed to fuse individual images applying the OR operation, where the existence of a scattering point was affirmed upon it was detected by one receiver. In [22], a Gaussian-based filter was applied before summing the multi-view estimates. However, these approaches overlook the geometric characteristics of extended targets, thereby limiting their fusion performance. Thus, it is necessary to develop advanced fusion methods that incorporate target geometric characteristics while accommodating distinct and unknown fields of view at receivers.

C. Main Contributions

In this paper, we consider a networked sensing system consisting of a transmit BS, multiple receive BSs that are connected to a CPU, and multiple extended targets that behave as dense clouds of scatterers. The transmitted signals are reflected by extended targets before arriving at receivers, and thus the characteristics of targets, such as shape and size, are embedded in received signals. This work aims to achieve precise imaging of extended targets by leveraging all received signals. To this end, we devise an advanced two-phase approach. In Phase I, each receive BS recovers an individual image using limited resources. In Phase II, the CPU fuses these individual results to construct a high-quality image. The main contributions of this paper are summarized as follows.

- In Phase I, we propose a novel covariance-based framework for single-view imaging. We reveal that the sample covariance matrix of the received signals is sufficient for imaging. Specifically, based on the maximum likelihood (ML) criterion and target geometric characteristics, we formulate the imaging problem as an estimation problem of effective scattering intensities over discretized grid points, where the target is recovered by grids exhibiting high intensity values. Compared to FFT and CS based methods that need to estimate instantaneous channels, our approach significantly reduces the number of parameters to be estimated by leveraging channel statistical properties, thereby enabling high-quality imaging even in the scenario with narrowband and small aperture.
- To enhance the imaging performance, we incorporate grid position optimization into the proposed covariance-based

single-view imaging framework. This enables dynamic adjustment of grid points to better align with target geometries, thereby improving the quality of environment reconstruction. Both theoretical analysis and simulation results demonstrate the significant advantage of employing dynamic grids over employing fixed and excessively dense grid points for imaging.

- We propose a novel multi-view fusion scheme for Phase II, where cross-disciplinary concepts and tools from computer vision that are highly new in the field of imaging are introduced. The fusion task is challenging because grid topologies are heterogeneous and receiver fields of view are distinct and unknown. To this end, we develop an edge-preserving natural neighbor interpolation (EP-NNI) method to map each single-view image onto a set of common and finer grids while preserving edge features. Then, we establish an optimization framework to fuse these interpolated results, which selects the subset of receivers that are “informative” on the scattering intensity of each grid, and updates the fused intensity based on this subset and target geometric features in an alternative manner.
- We provide extensive simulation results to verify the superiority of the proposed scheme. Specifically, the proposed covariance-based imaging scheme consistently outperforms FFT and CS based benchmark schemes, particularly in scenarios where resources, such as the number of antennas, transmit power, and pilot length, are limited. We interpret this performance gain as the exploitation of channel statistical properties and target geometric properties in our scheme. Moreover, our simulation results demonstrate that the proposed multi-view fusion scheme is effective in enhancing imaging performance and suppressing sidelobe artifacts existed in individual reconstructions.

Organizations: The rest of this paper is organized as follows. Section II describes the imaging system model. Section III formulates the imaging problem after grid discretization. In Section IV, we devise the covariance-based single-view imaging scheme and theoretically analyze the advantages of adopting dynamic grids. Section V presents the multi-view fusion scheme for high-quality reconstruction based on local images. Numerical results are presented in Section VI. Finally, Section VII concludes this paper.

Notations: Throughout this paper, uppercase and lowercase boldface letters denote matrices and vectors, respectively. The notation $[\mathbf{X}]_{m,:}$ denotes the m -th row of $\mathbf{X}$ and $[\mathbf{X}]_{:,n}$ denotes the n -th column of $\mathbf{X}$. The notation $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. We use $(\cdot)^T$, $(\cdot)^H$, $\|\mathbf{x}\|_p$, and $\|\mathbf{X}\|_F$ to denote transpose, conjugate transpose, ℓ_p -norm, and Frobenius norm, respectively. We use $\otimes$ and $\odot$ to denote Kronecker product and element-wise product, respectively. We use $\text{diag}\{\mathbf{x}\}$ to denote a diagonal matrix with $\mathbf{x}$ comprising its diagonal elements. The operation of the vectorization of a matrix is denoted as $\text{vec}(\cdot)$. Given any complex variable, vector, or matrix, we use $\Re(\cdot)$ and $\Im(\cdot)$ to return its real and imaginary parts, respectively. We use $\cdot \setminus \cdot$ to denote set subtraction and $|\mathcal{A}|$ to denote the cardinality of a set $\mathcal{A}$. For an integer $k > 0$, let

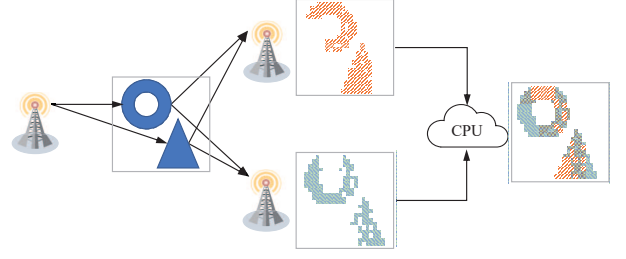


Fig. 1. System model.

$[k] = \{1, \dots, k\}$. We use $1[\cdot]$ to denote the indicator function. The distribution of a Bernoulli random variable x with mean λ is denoted as $\text{Ber}(\lambda)$. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted as $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

II. SYSTEM MODEL

We consider a multi-view imaging scenario in 6G cellular network as illustrated in Fig. 1, in which there are a transmit BS, K receive BSs that are connected to the CPU via fronthaul links, and B extended targets. The transmit BS and each receive BS are equipped with uniform linear arrays (ULAs) consisting of N_{tx} and N_{rx} antennas spaced by half of the carrier wavelength, respectively. Under a two-dimensional (2D) Cartesian coordinate system, the locations of the transmit BS and the k -th receive BS are denoted as $\mathbf{p}_{\text{tx}} = [x_{\text{tx}}, y_{\text{tx}}]^T \in \mathbb{R}^{2 \times 1}$ and $\mathbf{p}_{\text{rx},k} = [x_{\text{rx},k}, y_{\text{rx},k}]^T \in \mathbb{R}^{2 \times 1}$ in meter (m), respectively, which are known in advance¹.

We consider the transmission over M frames, each comprising L symbols. In each frame, denote the transmitted pilot signal as $\mathbf{X} \in \mathbb{C}^{N_{\text{tx}} \times L}$ satisfying $\|\mathbf{X}\|_F^2 = LP, \forall n \in [N_{\text{tx}}]$. This signal reaches the receiver through propagation channels that include the scattering due to extended targets. The signal received at the k -th receiver in the m -th frame is given by²

$$\mathbf{Y}_{k,m} = \mathbf{H}_{k,m} \mathbf{X} + \mathbf{Z}_{k,m} \in \mathbb{C}^{N_{\text{rx}} \times L}, \quad \forall k, m, \quad (1)$$

where $\mathbf{H}_{k,m} \in \mathbb{C}^{N_{\text{rx}} \times N_{\text{tx}}}$ denotes the channel matrix between the transmitter and the k -th receiver in the m -th frame and $\mathbf{Z}_{k,m}$ denotes the noise with each element following i.i.d. $\mathcal{CN}(0, \sigma_z^2)$ distribution.

According to the Born approximation [25] for the scattered field, the channel matrix $\mathbf{H}_{k,m}$ in (1) can be expressed as

$$\mathbf{H}_{k,m} = \iint_{\mathcal{S}} c_{x,y,k,m} \mathbf{b}(\theta_{x,y,k}) \mathbf{a}^T(\phi_{x,y}) dx dy, \quad \forall k, m, \quad (2)$$

where $\mathcal{S}$ denotes the spatial region occupied by extended targets, $[x, y]^T$ denotes the position of a general point on the

¹As in [8], [13], [23], [24], this paper focuses on 2D imaging for the purpose of exposition. This can be extended to three-dimensional (3D) imaging by deploying uniform planar arrays at BSs, which enables estimation of both azimuth and elevation angles and thus allows for detecting whether there exist scatterers in 3D space.

²Notice that (1) is a narrowband model, corresponding for example to one subcarrier of an orthogonal frequency division multiplexing (OFDM) system. This means that in an ISAC framework, we can transmit imaging-oriented pilots as in (1) on a single subcarrier, while using the remaining subcarriers for communication. Thus, the overhead of such form of imaging is very low, and can be easily plugged into any OFDM-based communication systems.

target, $\phi_{x,y}$ denotes the angle of departure (AoD) from the transmitter to the scatterer located in $[x, y]^T$ given by

$$\phi_{x,y} = \arctan\left(\frac{y_{\text{tx}} - y}{x_{\text{tx}} - x}\right) + \pi 1[x_{\text{tx}} < x], \quad (3)$$

$\theta_{x,y,k}$ denotes the AoA from the point scatter to the receiver k given by

$$\theta_{x,y,k} = \arctan\left(\frac{y_{\text{rx},k} - y}{x_{\text{rx},k} - x}\right) + \pi 1[x_{\text{rx},k} < x], \quad (4)$$

$\mathbf{a}(\phi_{x,y})$ and $\mathbf{b}(\theta_{x,y,k})$ denote the transmit steering vector towards $\phi_{x,y}$ and the receive steering vector towards $\theta_{x,y,k}$, respectively, given by

$$\mathbf{a}(\phi_{x,y}) = [1, e^{-j\pi \sin(\phi_{x,y})}, \dots, e^{-j\pi(N_{\text{tx}}-1)\sin(\phi_{x,y})}]^T, \quad (5)$$

$$\mathbf{b}(\theta_{x,y,k}) = [1, e^{-j\pi \sin(\theta_{x,y,k})}, \dots, e^{-j\pi(N_{\text{rx}}-1)\sin(\theta_{x,y,k})}]^T, \quad (6)$$

and $c_{x,y,k,m} = \alpha_{x,y,k} \bar{c}_{x,y,k,m}$ denotes the effective attenuation coefficient of the path from the transmitter to point $[x, y]^T$ and then to receiver k in frame m . The binary variable $\alpha_{x,y,k}$ equals to 1 if the signal reflected by point $[x, y]^T$ is received by the k -th receiver, and equals to 0 if it is not received due to occlusion effects or weak scattering at some sites [15]. The attenuation coefficient $\bar{c}_{x,y,k,m}$ is contributed by path loss and radar cross section (RCS). Following the Swerling-II target model [4], [26], we assume $\bar{c}_{x,y,k,m} \sim \mathcal{CN}(0, \gamma_{x,y,k})$, $\forall x, y, k, m$, which stays constant during the transmission of L symbols and changes i.i.d. in different frames. The variance $\gamma_{x,y,k}$ satisfies $\gamma_{x,y,k} = \bar{\gamma}_{r,x,y} \gamma_{\beta,x,y,k}$, where $\bar{\gamma}_{r,x,y}$ accounts for the scattering intensity and $\gamma_{\beta,x,y,k}$ denotes the path-loss. We adopt the normalization $\iint_{\mathcal{S}} \bar{\gamma}_{r,x,y} dx dy = 1$ [4].

The vectorized form of the received signal in (1) is given by

$$\mathbf{y}_{k,m} = \text{vec}(\mathbf{Y}_{k,m}) = \iint_{\mathcal{S}} \mathbf{v}_{x,y,k} c_{x,y,k,m} dx dy + \mathbf{z}_{k,m}, \quad \forall k, m, \quad (7)$$

where $\mathbf{v}_{x,y,k} = \mathbf{X}^T \mathbf{a}(\phi_{x,y}) \otimes \mathbf{b}(\theta_{x,y,k})$ and $\mathbf{z}_{k,m} = \text{vec}(\mathbf{Z}_{k,m})$. The concatenation of $\mathbf{y}_{k,m}$ over M frames is denoted as $\mathbf{Y}_k = [\mathbf{y}_{k,1}, \dots, \mathbf{y}_{k,M}]$. Since both attenuation components and noises are i.i.d. Gaussian distributed over the frames, the columns of $\mathbf{Y}_k$ are i.i.d. distributed according to $\mathcal{CN}(\mathbf{0}, \Sigma_{k,0})$ with the covariance matrix $\Sigma_{k,0}$ given by

$$\Sigma_{k,0} = \iint_{\mathcal{S}} \alpha_{x,y,k} \gamma_{x,y,k} \mathbf{v}_{x,y,k} \mathbf{v}_{x,y,k}^H dx dy + \sigma_z^2 \mathbf{I}_{LN_{\text{rx}}}, \quad \forall k. \quad (8)$$

Thus, the likelihood function of $\mathbf{Y}_k$ is given by

$$\begin{aligned} p(\mathbf{Y}_k | \{\alpha_{x,y,k}, \gamma_{x,y,k} : [x, y]^T \in \mathcal{S}\}) \\ = |\pi \Sigma_{k,0}|^{-M} e^{-\text{tr}(M \Sigma_{k,0}^{-1} \hat{\Sigma}_k)}, \quad \forall k, \end{aligned} \quad (9)$$

where $\hat{\Sigma}_k$ denotes the sample covariance matrix of the columns of the observation $\mathbf{Y}_k$ given by

$$\hat{\Sigma}_k = \frac{1}{M} \mathbf{Y}_k \mathbf{Y}_k^H, \quad \forall k. \quad (10)$$

III. GRID DISCRETIZATION AND PROBLEM STATEMENT

Different from point targets, extended targets usually comprise an infinite number of inseparable scattering points, rendering precise parameter estimation for individual scatterers impossible [4]. Consequently, the objective of imaging for extended targets is not to localize each single scatterer, but rather to create a high-resolution map of the reflectivity of the environment. To achieve this goal, we discretize the region of interest (RoI) into grids, thereby transforming the task of imaging into parameter estimation problem of each grid. Specifically, we discretize the rectangular RoI $\mathcal{A} = [x_1, x_2] \times [y_1, y_2]$ into Q grids, and approximate the scattering of extend targets as that contributed by a set of grid points within this region. We assume that each receiver constructs single-view image independently, and the grids adopted by different receivers can be distinct. Denote the position of the q -th grid point for the k -th receiver as $\mathbf{p}_{k,q} = [p_{x,k,q}, p_{y,k,q}]^T$. Then, the channel matrix $\mathbf{H}_{k,m}$ in (2) is approximated as $\tilde{\mathbf{H}}_{k,m}$ given by

$$\tilde{\mathbf{H}}_{k,m} = \sum_{q=1}^Q c_{k,q,m} \mathbf{b}(\theta_{k,q}) \mathbf{a}^T(\phi_{k,q}), \quad (11)$$

where $\phi_{k,q}$ and $\theta_{k,q}$ denote the AoD and AoA corresponding to the q -th grid at the k -th receiver obtained from (3) and (4) by replacing $[x, y]^T$ to $\mathbf{p}_{k,q}$, respectively, and $c_{k,q,m} = \alpha_{k,q} \bar{c}_{k,q,m}$ denotes the effective attenuation coefficient of the path from the transmitter to grid point q (particular to the set of grids adopted by receiver k) and then to receiver k . Here, the binary variable $\alpha_{k,q}$ equals to 1 if the q -th grid is in the field of view of receiver k , and equals to 0 otherwise, $\forall k, q$. Due to the central limit theorem when approximating the scattering from infinitely many points in a grid cell as that from a single point, we model the attenuation coefficient $\bar{c}_{k,q,m}$ as a CSCG variable with mean 0 and variance $\gamma_{k,q} = \bar{\gamma}_{r,k,q} \gamma_{\beta,k,q}$, $\forall k, q, m$, where $\bar{\gamma}_{r,k,q}$ accounts for the RCS part satisfying $\sum_{q \in [Q]} \bar{\gamma}_{r,k,q} = 1$ for normalization, and $\gamma_{\beta,k,q}$ accounts for the path-loss. Since the positions of grid points and BSs are known, we assume the path-losses $\gamma_{\beta,k,q}$'s are known in advance and denote $\gamma_{\beta,k} = [\gamma_{\beta,k,1}, \dots, \gamma_{\beta,k,Q}]^T$. For the k -th receiver, denote the effective scattering intensity of grid q as $\gamma_{r,k,q} = \alpha_{k,q} \bar{\gamma}_{r,k,q}$, which is to be estimated for determining the reflectivity map of the environment. We denote $\gamma_{r,k} = [\gamma_{r,k,1}, \dots, \gamma_{r,k,Q}]^T$.

The signal $\mathbf{Y}_{k,m}$ in (1) is approximated as $\tilde{\mathbf{Y}}_{k,m}$ given by

$$\tilde{\mathbf{Y}}_{k,m} = \tilde{\mathbf{H}}_{k,m} \mathbf{X} + \mathbf{Z}_{k,m}, \quad \forall k, m. \quad (12)$$

After vectorization, we have

$$\tilde{\mathbf{y}}_{k,m} = \text{vec}(\tilde{\mathbf{Y}}_{k,m}) = \mathbf{V}_k \mathbf{c}_{k,m} + \mathbf{z}_{k,m}, \quad \forall k, m, \quad (13)$$

where $\mathbf{c}_{k,m} = [c_{k,1,m}, \dots, c_{k,Q,m}]^T \in \mathbb{C}^Q$ and $\mathbf{V}_k = [\mathbf{v}_{k,1}, \dots, \mathbf{v}_{k,Q}] \in \mathbb{C}^{LN_{\text{rx}} \times Q}$ with

$$\mathbf{v}_{k,q} = \bar{\mathbf{a}}(\phi_{k,q}) \otimes \mathbf{b}(\theta_{k,q}), \quad \forall k, q, \quad (14)$$

$$\bar{\mathbf{a}}(\phi_{k,q}) = \mathbf{X}^T \mathbf{a}(\phi_{k,q}), \quad \forall k, q. \quad (15)$$

Then, the concatenation of $\tilde{\mathbf{y}}_{k,m}$ over M frames is given by

$$\tilde{\mathbf{Y}}_k = \mathbf{V}_k \mathbf{C}_k + \mathbf{Z}_k, \quad \forall k, \quad (16)$$

where $\tilde{\mathbf{Y}}_k = [\tilde{\mathbf{y}}_{k,1}, \dots, \tilde{\mathbf{y}}_{k,M}]$, $\mathbf{C}_k = [\mathbf{c}_{k,1}, \dots, \mathbf{c}_{k,M}]$, and $\mathbf{Z}_k = [\mathbf{z}_{k,1}, \dots, \mathbf{z}_{k,M}]$. Given the effective scattering intensity

$\gamma_{r,k}$ and grid position $\mathbf{P}_k = [\mathbf{p}_{k,1}, \dots, \mathbf{p}_{k,Q}]$, the columns of $\mathbf{Y}_k$ are i.i.d. distributed according to $\mathcal{CN}(\mathbf{0}, \Sigma_k)$, where the covariance matrix Σ_k is given by

$$\Sigma_k = \mathbf{V}_k \mathbf{\Gamma}_{r,k} \mathbf{\Gamma}_{\beta,k} \mathbf{V}_k^H + \sigma_z^2 \mathbf{I}_{LN_{rx}}, \quad \forall k, \quad (17)$$

with $\mathbf{\Gamma}_{r,k} = \text{diag}\{\gamma_{r,k}\}$ and $\mathbf{\Gamma}_{\beta,k} = \text{diag}\{\gamma_{\beta,k}\}$. By treating the quantized matrix Σ_k as the covariance matrix in the likelihood function of $\mathbf{Y}_k$ given in (9), we have

$$p(\mathbf{Y}_k | \gamma_{r,k}, \mathbf{P}_k) = |\pi \Sigma_k|^{-M} e^{-\text{tr}(M \Sigma_k^{-1} \hat{\Sigma}_k)}, \quad \forall k. \quad (18)$$

Based on the discretized model established above, we propose a two-phase approach to address the environment reconstruction challenge. In Phase I, each receiver recovers an individual image. Specifically, we formulate the imaging problem at each receiver k as a novel scattering intensity $\gamma_{r,k}$ and grid position $\mathbf{P}_k$ estimation problem, as will be introduced in Section IV. However, due to occlusion effects and weak RCS at some sites, each receiver has only a limited field of view about the target [6]–[8], as shown in Fig. 1. Even within its field of view, the performance is limited by the receiver's resolution. Thus, in Phase II, we fuse these individual results to construct a high-quality image of extended targets. Since the targets can be observed from different angular directions, significant diversity gain can be leveraged by fusing individual results, as will be introduced in Section V.

IV. PHASE I: COVARIANCE-BASED SINGLE-VIEW IMAGING

This section addresses the single-view imaging problem at each receiver. First, based on the ML criterion and target geometric characteristics, we formulate the imaging problem as an estimation problem of grid positions and effective scattering intensities. This problem is then solved applying the alternating optimization (AO) framework. Subsequently, we theoretically reveal the importance of grid optimization in achieving high-quality imaging of extended targets.

A. Problem Formulation

Conventional imaging methods, such as FFT and CS-based schemes, face inherent performance limitations due to their requirement to estimate high-dimensional instantaneous channel matrix $\mathbf{C}_k$ for determining the set of grids with scatters. To overcome this challenge, we propose a covariance-based imaging scheme that aims to estimate the statistical effective scattering intensity $\gamma_{r,k}$ instead of $\mathbf{C}_k$, thereby dramatically reducing the number of parameters to be estimated. Moreover, considering that the number of resolvable grid points cannot be excessively increased due to limited resolution, we propose to dynamically adjust grid points to conform to the geometry of targets. As a result, we formulate the single-view imaging problem at each receiver k as a novel estimation problem of both scattering intensity $\gamma_{r,k}$ and grid position $\mathbf{P}_k$. Building upon the grid representation introduced earlier, this problem admits an ML estimation framework by minimizing the approximate negative log-likelihood cost function given by

$$-\frac{1}{M} \ln p(\mathbf{Y}_k | \gamma_{r,k}, \mathbf{P}_k) \propto \ln |\Sigma_k| + \text{tr}(\Sigma_k^{-1} \hat{\Sigma}_k). \quad (19)$$

Moreover, due to the continuous nature of physical extended targets, they have inherent spatial clustering properties. That is, extended targets exhibit spatially coherent structures rather than isolated points. To account for this physical prior, we introduce a regularization penalty term as follows

$$(\gamma_{r,k} \odot \gamma_{\beta,k})^T \mathbf{L}_k (\gamma_{r,k} \odot \gamma_{\beta,k}) = \sum_{(q,q') \in \mathcal{N}_k} \frac{1}{2} (\gamma_{r,k,q} \gamma_{\beta,k,q} - \gamma_{r,k,q'} \gamma_{\beta,k,q'})^2, \quad (20)$$

where $\mathbf{L}_k$ denotes the graph Laplacian matrix encoding grid adjacency relationships and the set $\mathcal{N}_k$ includes adjacent pairs after spatial discretization at the k -th receiver. By penalizing large energy differences between adjacent grids, it imposes smoothness that matches the continuous nature of extended targets while suppressing isolated artifacts.

Thus, the penalized ML cost function is given by

$$\mathcal{J}(\gamma_{r,k}, \mathbf{P}_k) = \ln |\Sigma_k| + \text{tr}(\Sigma_k^{-1} \hat{\Sigma}_k) + \eta (\gamma_{r,k} \odot \gamma_{\beta,k})^T \mathbf{L}_k (\gamma_{r,k} \odot \gamma_{\beta,k}), \quad (21)$$

where $\eta > 0$ denotes the penalty parameter used to enforce cluster. The optimization problem is then formulated as

$$\min_{\gamma_{r,k}, \mathbf{P}_k} \mathcal{J}(\gamma_{r,k}, \mathbf{P}_k) \quad (22a)$$

$$\text{s.t.} \quad 0 \leq \gamma_{r,k,q} \leq 1, \quad \forall q \quad (22b)$$

$$\mathbf{p}_{k,q} \in \mathcal{A}, \quad \forall q \quad (22c)$$

$$\|\mathbf{p}_{k,q} - \mathbf{p}_{k,q}^{(0)}\|_2 \leq d_{\max}, \quad \forall q. \quad (22d)$$

where the constraint (22b) is due to the normalization and the fact that the energy is non-negative; the condition (22c) restricts all grid points are within the RoI; and the constraint (22d) not only prevents the grid index ambiguity problem (e.g., swaps between (q, q') and (q', q) coordinates) inherent to permutation-invariant formulations, but also guarantees the separability of adjacent grid points provided that initial positions $\mathbf{p}_{k,q}^{(0)}$'s are not too close. In the formulated covariance-based optimization problem, the sample covariance matrix $\hat{\Sigma}_k$ of received signals provides sufficient information for imaging. In the following, we introduce an AO framework to iteratively refine $\gamma_{r,k}$ and $\mathbf{P}_k$, with the outline of the proposed algorithm provided in Algorithm 1.

B. Optimizing $\gamma_{r,k}$ with Fixed $\mathbf{P}_k$

For fixed $\mathbf{P}_k$, the optimization with respect to $\gamma_{r,k}$ is not convex considering that the objective function $\mathcal{J}(\gamma_{r,k}, \mathbf{P}_k)$ is the sum of a concave log-determinant term and two convex terms. To address this problem, we apply the coordinate descent algorithm to update $\gamma_{r,k}$. As shown in Algorithm 1, at each step of one iteration, we optimize $\mathcal{J}(\gamma_{r,k}, \mathbf{P}_k)$ with respect to one of the coordinates of $\gamma_{r,k}$. Specifically, given $\gamma_{r,k}$ and $\mathbf{P}_k$ obtained in the previous step, the coordinate optimization with respect to $\gamma_{r,k,q}$ is equivalent to the optimization of the increment d in $\gamma_{r,k,q}$. The resulting optimization problem to determine d can be expressed as

$$\min_{-\gamma_{r,k,q} \leq d \leq 1 - \gamma_{r,k,q}} \mathcal{J}(\gamma_{r,k} + d \mathbf{e}_q, \mathbf{P}_k), \quad (23)$$

where $\mathbf{e}_q$ denotes the q -th canonical basis vector with a single 1 at its q -th coordinate and zero elsewhere. Applying the well-known Sherman-Morrison rank-1 update identity [27], we have

$$\begin{aligned} \mathcal{J}(\gamma_{r,k} + d\mathbf{e}_q, \mathbf{P}_k) &\propto f_{k,q}(d) \\ &= \ln(1 + d\gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \mathbf{v}_{k,q}) - \frac{d\gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \hat{\Sigma}_k \Sigma_k^{-1} \mathbf{v}_{k,q}}{1 + d\gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \mathbf{v}_{k,q}} \\ &\quad + \frac{\eta}{2} \sum_{q' \in \mathcal{N}_{k,q}} (d\gamma_{\beta,k,q} + \gamma_{r,k,q} \gamma_{\beta,k,q} - \gamma_{r,k,q'} \gamma_{\beta,k,q'})^2, \end{aligned} \quad (24)$$

where the set $\mathcal{N}_{k,q}$ includes neighbor grids of the q -th grid observed at the k -th receiver.

The function $f_{k,q}(d)$ is well-defined only when $d > d_0 = -\frac{1}{\gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \mathbf{v}_{k,q}}$. Taking the derivative of $f_{k,q}(d)$ yields

$$f'_{k,q}(d) = \frac{c_3 d^3 + c_2 d^2 + c_1 d + c_0}{(1 + d a_{k,q})^2}, \quad (26)$$

where

$$c_3 = a_{k,q}^2 c_{k,q}, \quad (27)$$

$$c_2 = 2a_{k,q} c_{k,q} + a_{k,q}^2 e_{k,q}, \quad (28)$$

$$c_1 = a_{k,q}^2 + c_{k,q} + 2a_{k,q} e_{k,q}, \quad (29)$$

$$c_0 = a_{k,q} - b_{k,q} + e_{k,q}, \quad (30)$$

$$a_{k,q} = \gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \mathbf{v}_{k,q}, \quad (31)$$

$$b_{k,q} = \gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \hat{\Sigma}_k \Sigma_k^{-1} \mathbf{v}_{k,q}, \quad (32)$$

$$c_{k,q} = \eta |\mathcal{N}_{k,q}| \gamma_{\beta,k,q}^2, \quad (33)$$

$$e_{k,q} = \eta \gamma_{\beta,k,q} \sum_{q' \in \mathcal{N}_{k,q}} (\gamma_{r,k,q} \gamma_{\beta,k,q} - \gamma_{r,k,q'} \gamma_{\beta,k,q'}). \quad (34)$$

The derivative $f'_{k,q}(d)$ involves a cubic function, which admits closed-form roots. Denote the set of real roots in the feasible region as

$$\mathcal{D} = \{d : f'_{k,q}(d) = 0, \Im(d) = 0, \max\{-\gamma_{r,k,q}, d_0\} \leq \Re(d) \leq 1 - \gamma_{r,k,q}\}. \quad (35)$$

The optimal increment d is obtained by selecting $1 - \gamma_{r,k,q}$, $\max\{-\gamma_{r,k,q}, d_0\}$, or root in the set $\mathcal{D}$ with the minimum objective function value. Then, $\gamma_{r,k,q}$ can be updated as $\gamma_{r,k,q} \leftarrow \gamma_{r,k,q} + d$. Applying the Sherman-Morrison rank-1 update identity, the covariance matrix is updated as

$$\Sigma_k^{-1} \leftarrow \Sigma_k^{-1} - d \frac{\gamma_{\beta,k,q} \Sigma_k^{-1} \mathbf{v}_{k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1}}{1 + d \gamma_{\beta,k,q} \mathbf{v}_{k,q}^H \Sigma_k^{-1} \mathbf{v}_{k,q}}. \quad (36)$$

Following similar procedure, all coordinates of $\gamma_{r,k}$ are updated once in each iteration, as shown in Algorithm 1.

C. Optimizing $\mathbf{P}_k$ with Fixed $\gamma_{r,k}$

Given $\gamma_{r,k}$ obtained in the previous step, we aim to optimize grid positions in this part. Denote the set of grid indexes with positive scattering intensity estimates as $\mathcal{Q}_{a,k} = \{q \in [Q] : \gamma_{r,k,q} > 0\}$. Let $\psi = \text{vec}(\mathbf{P}_{a,k})$, where $\mathbf{P}_{a,k}$ is a submatrix of $\mathbf{P}_k$ with the column indices lying in $\mathcal{Q}_{a,k}$. In the following, we aim to update ψ , while keeping the grid parameters indexed by the set $[Q] \setminus \mathcal{Q}_{a,k}$ equal to those in the previous iteration. Specifically, given the effective scattering intensity $\gamma_{r,k}$ and

Algorithm 1 Covariance-based Single-View Imaging

Input: The sample covariance matrix $\hat{\Sigma}_k$, RoI $\mathcal{A}$, noise variance σ_z^2 , tolerances ϵ_1 and ϵ_2 , and iterations $\text{Iter}_{\max}$, Iter_1 , and Iter_2 .
Output: The resulting estimate $\gamma_{r,k}$ and $\mathbf{P}_k$.

- 1: **Initialize:** $\Sigma_k = \sigma_z^2 \mathbf{I}$, $\gamma_{r,k} = \mathbf{0}$, and uniform grids $\mathbf{p}_{k,q}^{(0)}, \forall q$.
- 2: **for** $j = 1, 2, \dots, \text{Iter}_{\max}$ **do**
- 3: **% Update of $\gamma_{r,k}$**
- 4: **for** $j_1 = 1, 2, \dots, \text{Iter}_1$ **do**
- 5: Select an index set $\mathcal{Q}$ from the random permutation of $[Q]$.
- 6: **for** $q \in \mathcal{Q}$ **do**
- 7: Solve the cubic function in (26) and obtain $\mathcal{D}$ (35).
- 8: Set $d = \arg \min_{d \in \mathcal{D} \cup \{\max\{-\gamma_{r,k,q}, d_0\}, 1 - \gamma_{r,k,q}\}} f_{k,q}(d)$.
- 9: Update $\gamma_{r,k,q} \leftarrow \gamma_{r,k,q} + d$.
- 10: Update Σ_k^{-1} applying (36).
- 11: **end for**
- 12: **end for**
- 13: **% Update of $\mathbf{P}_k$**
- 14: **for** $j_2 = 1, 2, \dots, \text{Iter}_2$ **do**
- 15: Calculate the gradient $\nabla_{\psi} f_{k,x,y}(\psi)$ applying (39) - (52).
- 16: Compute the stepsize ξ by Armijo rule.
- 17: Update ψ applying (38).
- 18: Apply the rectangle projection operator and the Euclidean ball projection operator in (53) for $\mathbf{P}_k$.
- 19: **end for**
- 20: If $\frac{\|\gamma_{r,k}^{(j)} - \gamma_{r,k}^{(j-1)}\|_2}{\|\gamma_{r,k}^{(j-1)}\|_2} \leq \epsilon_1$ and $\nabla_{\psi} f_{k,x,y}(\psi) \leq \epsilon_2$, stop.
- 21: **end for**

grid positions indexed by the set $q \in [Q] \setminus \mathcal{Q}_{a,k}$, the cost function with respect to ψ is expressed as

$$\mathcal{J}(\gamma_{r,k}, \mathbf{P}_k) \propto f_{k,x,y}(\psi) = \ln |\Sigma_k| + \text{tr}(\Sigma_k^{-1} \hat{\Sigma}_k). \quad (37)$$

It is challenging to find the optimal ψ that minimizes the cost function in (37) since it is non-convex with respect to ψ . Thus, a gradient descent approach is employed. Specifically, the grid parameters in ψ are updated as

$$\psi \leftarrow \psi - \xi \nabla_{\psi} f_{k,x,y}(\psi), \quad (38)$$

where the step size ξ is determined by the Armijo rule. The gradient $\nabla_{\psi} f_{k,x,y}(\psi)$ is given by

$$\nabla_{\psi} f_{k,x,y}(\psi) = \left[\frac{\partial f_{k,x,y}(\psi)}{\partial p_{x,k,\mathcal{Q}_{a,k}(1)}}, \frac{\partial f_{k,x,y}(\psi)}{\partial p_{y,k,\mathcal{Q}_{a,k}(1)}}, \dots, \frac{\partial f_{k,x,y}(\psi)}{\partial p_{x,k,\mathcal{Q}_{a,k}(|\mathcal{Q}_{a,k}|)}}, \frac{\partial f_{k,x,y}(\psi)}{\partial p_{y,k,\mathcal{Q}_{a,k}(|\mathcal{Q}_{a,k}|)}} \right]^T, \quad (39)$$

where for each $q \in \mathcal{Q}_{a,k}$, we have

$$\begin{aligned} \frac{\partial f_{k,x,y}(\psi)}{\partial p_{x,k,q}} &= \frac{2\gamma_{r,k,q} \gamma_{\beta,k,q} \Re\{\tilde{\mathbf{v}}_{k,q,x}^H \Sigma_{k,\setminus q}^{-1} (\mathbf{I}_{LN_{\text{Tx}}} - \hat{\Sigma}_k \Sigma_{k,\setminus q}^{-1}) \mathbf{v}_{k,q}\}}{1 + \gamma_{k,q} \mathbf{v}_{k,q}^H \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q}} \\ &\quad + \frac{2\gamma_{r,k,q}^2 \gamma_{\beta,k,q}^2 \mathbf{v}_{k,q}^H \Sigma_{k,\setminus q}^{-1} \hat{\Sigma}_k \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q} \Re\{\tilde{\mathbf{v}}_{k,q,x}^H \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q}\}}{(1 + \gamma_{k,q} \mathbf{v}_{k,q}^H \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q})^2}, \end{aligned} \quad (40)$$

$$\begin{aligned} \frac{\partial f_{k,x,y}(\psi)}{\partial p_{y,k,q}} &= \frac{2\gamma_{r,k,q} \gamma_{\beta,k,q} \Re\{\tilde{\mathbf{v}}_{k,q,y}^H \Sigma_{k,\setminus q}^{-1} (\mathbf{I}_{LN_{\text{Tx}}} - \hat{\Sigma}_k \Sigma_{k,\setminus q}^{-1}) \mathbf{v}_{k,q}\}}{1 + \gamma_{k,q} \mathbf{v}_{k,q}^H \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q}} \\ &\quad + \frac{2\gamma_{r,k,q}^2 \gamma_{\beta,k,q}^2 \mathbf{v}_{k,q}^H \Sigma_{k,\setminus q}^{-1} \hat{\Sigma}_k \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q} \Re\{\tilde{\mathbf{v}}_{k,q,y}^H \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q}\}}{(1 + \gamma_{k,q} \mathbf{v}_{k,q}^H \Sigma_{k,\setminus q}^{-1} \mathbf{v}_{k,q})^2}, \end{aligned} \quad (41)$$

$$\Sigma_{k,\setminus q}^{-1} = (\Sigma_k - \gamma_{k,q} \mathbf{v}_{k,q} \mathbf{v}_{k,q}^H)^{-1}, \quad (42)$$

$$\tilde{\mathbf{v}}_{k,q,x} = (\mathbf{X}^T \otimes \mathbf{I}_{N_{\text{rx}}}) \bar{\mathbf{A}}_{k,q,x} (\mathbf{a}(\phi_{k,q}) \otimes \mathbf{b}(\theta_{k,q})), \quad (43)$$

$$\tilde{\mathbf{v}}_{k,q,y} = (\mathbf{X}^T \otimes \mathbf{I}_{N_{\text{rx}}}) \bar{\mathbf{A}}_{k,q,y} (\mathbf{a}(\phi_{k,q}) \otimes \mathbf{b}(\theta_{k,q})), \quad (44)$$

$$\bar{\mathbf{A}}_{k,q,x} = a_{k,q,x} \mathbf{A}_{N_{\text{tx}}} \otimes \mathbf{I}_{N_{\text{rx}}} + b_{k,q,x} \mathbf{I}_{N_{\text{tx}}} \otimes \mathbf{A}_{N_{\text{rx}}}, \quad (45)$$

$$\bar{\mathbf{A}}_{k,q,y} = a_{k,q,y} \mathbf{A}_{N_{\text{tx}}} \otimes \mathbf{I}_{N_{\text{rx}}} + b_{k,q,y} \mathbf{I}_{N_{\text{tx}}} \otimes \mathbf{A}_{N_{\text{rx}}}, \quad (46)$$

$$\mathbf{A}_{N_{\text{tx}}} = -j\pi \text{diag}\{0, 1, \dots, N_{\text{tx}} - 1\}, \quad (47)$$

$$\mathbf{A}_{N_{\text{rx}}} = -j\pi \text{diag}\{0, 1, \dots, N_{\text{rx}} - 1\}, \quad (48)$$

$$a_{k,q,x} = \frac{(x_{\text{tx}} - p_{x,k,q})(y_{\text{tx}} - p_{y,k,q})}{((x_{\text{tx}} - p_{x,k,q})^2 + (y_{\text{tx}} - p_{y,k,q})^2)^{\frac{3}{2}}}, \quad (49)$$

$$b_{k,q,x} = \frac{(x_{\text{rx},k} - p_{x,k,q})(y_{\text{rx},k} - p_{y,k,q})}{((x_{\text{rx},k} - p_{x,k,q})^2 + (y_{\text{rx},k} - p_{y,k,q})^2)^{\frac{3}{2}}}, \quad (50)$$

$$a_{k,q,y} = \frac{-(x_{\text{tx}} - p_{x,k,q})^2}{((x_{\text{tx},k} - p_{x,k,q})^2 + (y_{\text{rx},k} - p_{y,k,q})^2)^{\frac{3}{2}}}, \quad (51)$$

$$b_{k,q,y} = \frac{-(x_{\text{rx},k} - p_{x,k,q})^2}{((x_{\text{rx},k} - p_{x,k,q})^2 + (y_{\text{rx},k} - p_{y,k,q})^2)^{\frac{3}{2}}}. \quad (52)$$

After obtaining ψ , we project grid positions onto the feasible set defined by the constraints (22c) and (22d). We adopt the alternating projection method since both the rectangular region $\mathcal{A}$ and the Euclidean ball are convex. Specifically, we first project $\mathbf{p}_{k,q}$ onto the rectangle by performing $p_{k,q,x} \leftarrow \min(\max(p_{k,q,x}, x_1), x_2)$ and $p_{k,q,y} \leftarrow \min(\max(p_{k,q,y}, y_1), y_2)$, and then project it onto the ball via

$$\mathbf{p}_{k,q} \leftarrow \begin{cases} \mathbf{p}_{k,q}, & \|\mathbf{p}_{k,q} - \mathbf{p}_{k,q}^{(0)}\|_2 \leq d_{\text{max}}, \\ \mathbf{p}_{k,q}^{(0)} + \frac{d_{\text{max}}(\mathbf{p}_{k,q} - \mathbf{p}_{k,q}^{(0)})}{\|\mathbf{p}_{k,q} - \mathbf{p}_{k,q}^{(0)}\|_2}, & \|\mathbf{p}_{k,q} - \mathbf{p}_{k,q}^{(0)}\|_2 > d_{\text{max}}. \end{cases} \quad (53)$$

The above projections iterate for a few times until convergence to ensure both constraints are satisfied. We choose uniform grids in $\mathcal{A}$ as initial grids. By ensuring that each optimized grid point is in an Euclidean ball centered by the corresponding initial grid, we can prevent the grid index ambiguity problem while making sure adjacent points are not too close.

D. Role of Dynamic Grids

In this part, we establish the theoretical foundation of adopting dynamic grids for achieving high-quality imaging. The cost function in (37) can be rewritten as

$$\ln|\Sigma_k| + \text{tr}(\Sigma_k^{-1} \hat{\Sigma}_k) = \ln|\mathbf{T}_k| + \text{tr}(\mathbf{T}_k^{-1}) + \ln|\hat{\Sigma}_k| \quad (54)$$

$$= \sum_{n=1}^N \left(\ln \tau_n + \frac{1}{\tau_n} \right) + \ln|\hat{\Sigma}_k|, \quad (55)$$

where $N = LN_{\text{rx}}$ and τ_n denotes the eigenvalue of $\mathbf{T}_k = \hat{\Sigma}_k^{-\frac{1}{2}} \Sigma_k \hat{\Sigma}_k^{-\frac{1}{2}}$. The cost function (55) is minimized when $\tau_n \rightarrow 1, \forall n$. The deviation from optimality is bounded by

$$\sum_{n=1}^N (\tau_n - 1)^2 = \|\hat{\Sigma}_k^{-\frac{1}{2}} (\Sigma_k - \hat{\Sigma}_k) \hat{\Sigma}_k^{-\frac{1}{2}}\|_F^2 \quad (56)$$

$$\leq \|\hat{\Sigma}_k^{-\frac{1}{2}}\|_F^4 \|\Sigma_k - \hat{\Sigma}_k\|_F^2 \quad (57)$$

$$\leq \|\hat{\Sigma}_k^{-\frac{1}{2}}\|_F^4 (\|\Sigma_{k,0} - \hat{\Sigma}_k\|_F + \|\Sigma_k - \Sigma_{k,0}\|_F)^2. \quad (58)$$

where (57) follows from the sub-multiplicative property of Frobenius norm and (58) follows from the triangle inequality.

The term $\|\Sigma_{k,0} - \hat{\Sigma}_k\|_F$ quantifies the gap between the sample covariance matrix and the true one, which can be reduced by increasing the number M of frames based on the law of large numbers. The term $\|\Sigma_k - \Sigma_{k,0}\|_F$ quantifies discretization error. Minimizing this term requires careful grid design.

It is evident that we can reduce $\|\Sigma_k - \Sigma_{k,0}\|_F$ by increasing the number Q of grids, but this introduces a critical limitation in the distinguishability of adjacent grids. Specifically, the matrix coherence of $\mathbf{V}_k$, defined as the maximum inner product between its normalized columns, is given by

$$\mu(\mathbf{V}_k) \triangleq \max_{q' \neq q} \frac{|\mathbf{v}_{k,q'}^H \mathbf{v}_{k,q}|}{\|\mathbf{v}_{k,q'}\|_2 \|\mathbf{v}_{k,q}\|_2} \quad (59)$$

$$= \max_{q' \neq q} \frac{|\bar{\mathbf{a}}^H(\phi_{k,q'}) \bar{\mathbf{a}}(\phi_{k,q})|}{L} \frac{|\mathbf{b}^H(\theta_{k,q'}) \mathbf{b}(\theta_{k,q})|}{N_{\text{rx}}}. \quad (60)$$

Generally speaking, smaller coherence $\mu(\mathbf{V}_k)$ provides stronger recovery guarantees. However, as Q increases, $\mu(\mathbf{V}_k)$ approaches 1, thereby violating the distinguishability of adjacent grids. Taking $\frac{|\mathbf{b}^H(\theta_{k,q'}) \mathbf{b}(\theta_{k,q})|}{N_{\text{rx}}}$ as an example, we have

$$\frac{|\mathbf{b}^H(\theta_{k,q'}) \mathbf{b}(\theta_{k,q})|}{N_{\text{rx}}} = \frac{1 - e^{j\pi N_{\text{rx}} \Delta_{\theta,q,q'}}}{N_{\text{rx}} |1 - e^{j\pi \Delta_{\theta,q,q'}}|} = \left| \frac{\sin(\frac{\pi N_{\text{rx}}}{2} \Delta_{\theta,q,q'})}{N_{\text{rx}} \sin(\frac{\pi}{2} \Delta_{\theta,q,q'})} \right|, \quad (61)$$

where $\Delta_{\theta,q,q'} = \sin \theta_{k,q'} - \sin \theta_{k,q}$. In the case of $|\Delta_{\theta,q,q'}| \leq \frac{1}{N_{\text{rx}}}$, it is evident that $\frac{|\mathbf{b}^H(\theta_{k,q'}) \mathbf{b}(\theta_{k,q})|}{N_{\text{rx}}}$ increases as $|\Delta_{\theta,q,q'}|$ decreases. Once $|\Delta_{\theta,q,q'}| \ll \frac{1}{N_{\text{tx}}}$, we have $\frac{|\mathbf{b}^H(\theta_{k,q'}) \mathbf{b}(\theta_{k,q})|}{N_{\text{rx}}} \rightarrow 1$. Similar convergence can be observed for the DoA term. Thus, for fixed numbers of transmit and receive antennas, we have $\mu(\mathbf{V}_k) \rightarrow 1$ as the number Q of grids increases, thereby violating the incoherence requirements for stable recovery.

To avoid this issue, we propose to use dynamic grids to reduce the discretization error, instead of excessively increasing the number of grids. Specifically, we have

$$\|\Sigma_k - \Sigma_{k,0}\|_F^2 = \|\Psi_k - \Psi_{k,0}\|_F^2 \quad (62)$$

$$= \|\Psi_k - \mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k} - \mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp\|_F^2 \quad (63)$$

$$= \|\Psi_k - \mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}\|_F^2 + \|\mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp\|_F^2 \quad (64)$$

$$\leq \|\Psi_k - \mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}\|_F^2 + 2\|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0}\|_F^2, \quad (65)$$

where $\Psi_{k,0} = \iint_{\mathcal{S}} \alpha_{x,y,k} \gamma_{x,y,k} \mathbf{v}_{x,y,k} \mathbf{v}_{x,y,k}^H dx dy$, $\Psi_k = \mathbf{V}_k \Gamma_{r,k} \Gamma_{\beta,k} \mathbf{V}_k^H$, $\bar{\mathbf{V}}_k$ is a submatrix of $\mathbf{V}_k$ with the column indices lying in the set $\mathcal{R}_k = \{q \in [Q] : \gamma_{r,k,q} > 0\}$, $\mathcal{P}_{\bar{\mathbf{V}}_k}$ denotes the projection matrix onto the subspace spanned by the columns of $\bar{\mathbf{V}}_k$, $\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp = \mathbf{I} - \mathcal{P}_{\bar{\mathbf{V}}_k}$, (64) follows because $\|\mathbf{A} + \mathbf{B}\|_F^2 = \|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2$ in the case of $\Re(\text{tr}(\mathbf{A}^H \mathbf{B})) = 0$, and (65) follows because

$$\|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}\|_F^2 + \|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp\|_F^2 = \text{tr}(\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0}) + \text{tr}(\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0}) \quad (66)$$

$$= \text{tr}(\Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0}) \quad (67)$$

$$= \|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0}\|_F^2, \quad (68)$$

and $\|\mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp\|_F^2 + \|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}\|_F^2 = \|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0}\|_F^2$ obtained in a similar way.

The upper bound in (65) consists of two terms. The first term $\|\Psi_k - \mathcal{P}_{\bar{\mathbf{V}}_k} \Psi_{k,0} \mathcal{P}_{\bar{\mathbf{V}}_k}^\perp\|_F^2$ quantifies how well the discretized matrix Ψ_k approximates the projection of the true matrix $\Psi_{k,0}$ onto the subspace spanned by the columns of $\bar{\mathbf{V}}_k$. For fixed grids, this term can be reduced by optimizing the effective scattering intensity $\gamma_{r,k}$. The second term $2\|\mathcal{P}_{\bar{\mathbf{V}}_k}^\perp \Psi_{k,0}\|_F^2$ measures the energy of $\Psi_{k,0}$ that is orthogonal to the subspace spanned by the columns of $\bar{\mathbf{V}}_k$. It arises due to subspace mismatch. For fixed scattering intensity, we can use grid optimization to align $\bar{\mathbf{V}}_k$ with the dominant eigenvectors of $\Psi_{k,0}$, thereby reducing the subspace mismatch effect. This process helps to generate a coordinate system intrinsically aligned with the target, where grid positions are adjusted to conform to target geometries, leading to better imaging performance than optimizing the scattering intensity alone.

V. PHASE II: MULTI-VIEW FUSION

Each receiver obtains individual images applying the method proposed in Section IV. This section addresses the challenge of fusing individual images to produce high-quality environment reconstruction.

A. Edge-Preserving Interpolation

The number of resolvable grid points is constrained by available resources, such as the numbers of transmit and receive antennas. This limited number of grids will inevitably result in non-smooth contours. Moreover, receivers operate on heterogeneous grids tailored to their respective spatial positions. Such topological distinction makes it challenging for fusion. To address these issues, we introduce some cross-disciplinary concepts and approaches from computer vision to the field of imaging for the first time. Based on these techniques, we interpolate each receiver's coarse image onto a set of common and finer grids, thereby establishing a unified foundation for the subsequent fusion process.

While traditional space-invariant interpolation methods, such as nearest-neighbor, bi-linear, bi-cubic, and Gaussian techniques, are computationally efficient, their fundamental limitation lies in determining weights solely by geometric distance to the interpolation point. These methods overlook local image structures by employing isotropic kernels to forcibly average intensities across physical boundaries, thereby inevitably suffering from edge blurring. This motivates us to explore space-variant interpolation, in which interpolation weights are dynamically adjusted to exploit image structural features. However, some existing space-variant interpolation methods, such as new edge-directed interpolation [28], are constrained to regular grids, failing to address the irregular grid positions inherent to our imaging problem. To this end, we propose an EP-NNI method that simultaneously preserves edge features and accommodates irregular grid points.

Specifically, we aim to interpolate $\gamma_{r,k}$ corresponding to Q irregular points to $\gamma'_{r,k}$ corresponding to $Q' = Q'_1 \times Q'_2$ regular grids shared by K receivers while preserving

edge features. The core idea of EP-NNI lies in integrating gradient-based geometric features into natural neighbor interpolation (NNI) weights, thereby effectively inhibiting smoothing across edges while maintaining smoothness along edges. The implementation of this idea proceeds as follows. First, we estimate the gradient at each irregular point using local plane fitting, which operates based on the assumption that within a small neighborhood around a point $q \in [Q]$, the scattering intensity can be modeled as a linear function $\gamma_{r,k,j} \approx c_{k,q,1}p_{x,k,j} + c_{k,q,2}p_{y,k,j} + c_{k,q,3}$, where $j \in \mathcal{N}_{k,q}$ and the set $\mathcal{N}_{k,q} \subset [Q]$ includes irregular points in the neighborhood of point q [29]. Based on the least square criterion, $\mathbf{c}_{k,q} = [c_{k,q,1}, c_{k,q,2}, c_{k,q,3}]^T$ is estimated as

$$\mathbf{c}_{k,q} = \arg \min_{\mathbf{c}_{k,q}} \|\mathbf{r}_{k,q} - \mathbf{A}_{k,q} \mathbf{c}_{k,q}\|_2^2 \quad (69)$$

$$= (\mathbf{A}_{k,q}^T \mathbf{A}_{k,q})^{-1} \mathbf{A}_{k,q}^T \mathbf{r}_{k,q}, \quad (70)$$

where $\mathbf{r}_{k,q} \in \mathbb{R}^{|\mathcal{N}_{k,q}| \times 1}$ is formed by stacking the elements $\gamma_{r,k,j}, \forall j \in \mathcal{N}_{k,q}$ into a vector, and the matrix $\mathbf{A}_{k,q} \in \mathbb{R}^{|\mathcal{N}_{k,q}| \times 3}$ is constructed by stacking $p_{x,k,j}, \forall j \in \mathcal{N}_{k,q}$ in its first column, stacking $p_{y,k,j}, \forall j \in \mathcal{N}_{k,q}$ in its second column, and placing an all-one vector in its third column. Then, the gradient at the irregular grid $q \in [Q]$ can be approximated as

$$\nabla \gamma_{r,k,q} = \begin{bmatrix} \frac{\partial \gamma_{r,k,q}}{\partial p_{x,k,q}} & \frac{\partial \gamma_{r,k,q}}{\partial p_{y,k,q}} \end{bmatrix}^T \approx [c_{k,q,1} \quad c_{k,q,2}]^T.$$

Next, based on the gradients at irregular grids, we compute the structure tensor at each of the Q' regular grid points to guide the anisotropic interpolation. As a powerful concept from computer vision, the structure tensor characterizes the local geometric structure by forming a smoothed second-moment matrix of the gradients [30]. For point $q' \in [Q']$, the structure tensor is given by

$$\mathbf{J}_{q'} = \sum_{q \in \mathcal{N}'_{k,q'}} w_{q,q'}^{\text{sib}} \nabla \gamma_{r,k,q} (\nabla \gamma_{r,k,q})^T \quad (71)$$

$$\approx \sum_{q \in \mathcal{N}'_{k,q'}} w_{q,q'}^{\text{sib}} \begin{bmatrix} c_{k,q,1}^2 & c_{k,q,1} c_{k,q,2} \\ c_{k,q,1} c_{k,q,2} & c_{k,q,2}^2 \end{bmatrix}, \quad (72)$$

where the set $\mathcal{N}'_{k,q'} \subset [Q]$ includes irregular grid points near the regular grid point q' , and $w_{q,q'}^{\text{sib}}$ denotes the Sibson natural neighbor weight determined by two Voronoi diagrams [31]. In the first diagram, the RoI $\mathcal{A}$ is partitioned into Q cells centered by the original Q irregular grid points, with the Voronoi cell $\text{Vor}(q)$ containing all points in the plane $\mathcal{A}$ closer to the q -th grid than to any other grids. The second diagram is constructed based on the target Q' regular grid points, with $\text{Vor}(q')$ representing the Voronoi cell corresponding to the grid $q' \in [Q']$. Applying Sibson's method, the weight $w_{q,q'}^{\text{sib}}$ is proportional to the area overlap between the Voronoi cells $\text{Vor}(q')$ and $\text{Vor}(q)$, given by [31]

$$w_{q,q'}^{\text{sib}} = \frac{\text{Area}(\text{Vor}(q) \cap \text{Vor}(q'))}{\text{Area}(\text{Vor}(q'))}. \quad (73)$$

Based on the structure tensor in (72) and the steering kernel in [32], we design an edge-preserving weight to penalize cross-edge interpolation, given by

$$\tilde{w}_{q,q'} = w_{q,q'}^{\text{sib}} \exp \left\{ - \frac{\mathbf{d}_{q,q'}^T \mathbf{J}_{q'} \mathbf{d}_{q,q'}}{2\sigma_{\text{EP}}^2} \right\}, \quad (74)$$

where $\mathbf{d}_{q,q'} = \mathbf{p}_{k,q} - \mathbf{p}_{k,q'}$ and σ_{EP}^2 serves as an edge-sensitive decay parameter. Applying singular value decomposition (SVD), the structure tensor satisfies $\mathbf{J}_{q'} = \mathbf{F}_{q'} \mathbf{\Lambda}_{q'} \mathbf{F}_{q'}^T$,

where $\Lambda_{q'} = \text{diag}\{\lambda_{1,q'}, \lambda_{2,q'}\}$ with $\lambda_{1,q'} \geq \lambda_{2,q'}$ and $\mathbf{F}_{q'} = [\mathbf{f}_{1,q'} \ \mathbf{f}_{2,q'}]$. Here, $\mathbf{f}_{1,q'}$ represents the dominant edge normal direction corresponding to the maximum intensity variation, while its orthogonal counterpart $\mathbf{f}_{2,q'}$ indicates the edge tangent direction along minimal variation [30]. Then, the quadratic term in the exponent of (74) can be expressed as $\mathbf{d}_{q,q'}^T \mathbf{J}_{q'} \mathbf{d}_{q,q'} = \lambda_{1,q'} (\mathbf{d}_{q,q'}^T \mathbf{f}_{1,q'})^2 + \lambda_{2,q'} (\mathbf{d}_{q,q'}^T \mathbf{f}_{2,q'})^2$. As we can see, the anisotropic weight $\tilde{w}_{q,q'}$ imposes maximal penalty when $\mathbf{d}_{q,q'}$ is parallel to $\mathbf{f}_{1,q'}$ and minimal penalty when orthogonal to $\mathbf{f}_{1,q'}$, thereby preserving sharp edge structures.

By adaptively weighting contributions from neighbor observation points, the scattering intensity at the q' -th interpolation point is given by

$$\gamma'_{r,k,q'} = \sum_{q \in \mathcal{N}'_{k,q'}} \frac{\tilde{w}_{q,q'}}{\sum_{q'' \in \mathcal{N}'_{k,q'}} \tilde{w}_{q'',q'}} \gamma_{r,k,q}. \quad (75)$$

Then, for each receiver k , the scattering intensity corresponding to Q' interpolation points is denoted as $\gamma'_{r,k} = [\gamma'_{r,k,1}, \dots, \gamma'_{r,k,Q'}]^T \in \mathbb{R}^{Q'}$.

B. Fusion

In this section, we aim to aggregate the interpolated scattering intensities at K receivers to obtain high-accuracy environment reconstruction. We formulate this task as a novel optimization problem that integrates a weighted least squares term, a sparsity promoting term, and a gradient sparsity promoting term, given by

$$\min_{\gamma'_r, \Lambda} f_{\text{WLS}}(\gamma'_r, \Lambda) + \mu f_{\text{sparse}}(\gamma'_r) + \eta f_{\text{TV}}(\gamma'_r) \quad (76a)$$

$$\text{s.t.} \quad 0 \leq \gamma'_{r,q} \leq 1, \quad \forall q \quad (76b)$$

$$\lambda_{k,q} \in \{0, 1\}, \quad \forall k, q, \quad (76c)$$

where $\mu > 0$ and $\eta > 0$ are penalty parameters. In this problem, the optimization variables are the fused scattering intensity $\gamma'_r = [\gamma'_{r,1}, \dots, \gamma'_{r,Q'}]^T \in \mathbb{R}^{Q'}$ and the binary matrix $\Lambda \in \{0, 1\}^{K \times Q'}$ with its (k, q) -th element $\lambda_{k,q}$ indicating whether grid q lies within the field of view of receiver k . The inputs of this problem are the interpolated scattering intensities $\{\gamma'_{r,k} : k \in [K]\}$ over a common set of Q' grids from K receivers obtained in the previous subsection. The three terms in the objective function (76a) are introduced in the following.

The weighted least squares term in (76a) is given by

$$f_{\text{WLS}}(\gamma'_r, \Lambda) = \sum_{k=1}^K \sum_{q=1}^{Q'} (\lambda_{k,q} \gamma_{\beta,k,q} (\gamma'_{r,k,q} - \gamma'_{r,q})^2 + (1 - \lambda_{k,q}) \gamma_{\beta,k,q} (\gamma'_{r,k,q})^2). \quad (77)$$

The large-scale fading coefficient $\gamma_{\beta,k,q}$ serves as a weighting factor. Larger values indicate better channel conditions and thus higher confidence in the corresponding estimate $\gamma'_{r,k,q}$. Moreover, as introduced before, each receiver only has a limited field of view about the target, and this field is unknown in advance. The binary variable $\lambda_{k,q}$ acts as an adaptive selector, excluding receivers that receive weak scattering signals from grid q and only fusing estimates from receivers whose fields of view likely contains this grid. Specifically, when $\lambda_{k,q} = 0$, grid q is presumed to lie outside the field of view of receiver k , and thus we only penalize $\gamma_{\beta,k,q} (\gamma'_{r,k,q})^2$ because $\gamma'_{r,k,q}$ is

supposed to approach 0 in this case. For valid estimate $\gamma'_{r,k,q}$ with $\lambda_{k,q} = 1, \forall k$, (77) penalizes their squared deviation from the fusion target $\gamma'_{r,q}$, ensuring the final fused value represents a channel-weighted consensus of trustworthy estimates. Thus, the criterion in (77) allows us to estimate effective observation zones without prior knowledge of receivers' fields of view, and ensures only the estimates from receivers with strong scattering intensity are aggregated.

To incorporate essential target geometric features into the fusion process, we propose to introduce two regularization terms into the objective function. Specifically, to exploit the fact that targets occupy a limited number of grids, we employ the following sparsity-promoting term

$$f_{\text{sparse}}(\gamma'_r) = \|\gamma'_r\|_1. \quad (78)$$

This term promotes a solution where only a small number of grids exhibit nonzero scattering intensities. Moreover, considering that extended targets usually occupy multiple adjacent grids, we introduce the l_1 -based anisotropic total variation regularization term, which is defined as [33]

$$f_{\text{TV}}(\gamma'_r) = \|\mathbf{D}\gamma'_r\|_1, \quad (79)$$

where $\mathbf{D} = [\mathbf{D}_x^T, \mathbf{D}_y^T]^T$ with $\mathbf{D}_x \in \{-1, 1\}^{Q' \times Q'}$ and $\mathbf{D}_y \in \{-1, 1\}^{Q' \times Q'}$ denoting the discrete gradient operator along x - and y -axis. The discrete gradient exhibits high values only when scattering intensities of adjacent grids are significantly different, and approaches zero in regions where the scattering intensity remains approximately constant. For large extended targets, transitions between empty grids and scatterer-occupied grids are sparse. By promoting gradient sparsity, one can generate image with enhanced spatial continuity [14], [34].

In the following, we adopt the AO framework to iteratively optimize γ'_r and Λ . When Λ is fixed, by introducing auxiliary variable $\mathbf{z}_r = \mathbf{D}\gamma'_r$, we can obtain the following problem

$$\min_{\gamma'_r, \mathbf{z}_r} f_{\text{WLS}}(\gamma'_r, \Lambda) + \mu \|\gamma'_r\|_1 + \eta \|\mathbf{z}_r\|_1 \quad (80a)$$

$$\text{s.t.} \quad \mathbf{z}_r = \mathbf{D}\gamma'_r, \quad (80b)$$

$$0 \leq \gamma'_{r,q} \leq 1, \quad \forall q, \quad (80c)$$

The augmented Lagrangian is expressed as

$$\begin{aligned} \mathcal{L}(\gamma'_r, \Lambda, \mathbf{z}_r, \mathbf{u}_r) &= f_{\text{WLS}}(\gamma'_r, \Lambda) + \mu \|\gamma'_r\|_1 + \eta \|\mathbf{z}_r\|_1 \\ &\quad + \frac{\rho}{2} \|\mathbf{z}_r - \mathbf{D}\gamma'_r + \mathbf{u}_r\|_2^2, \end{aligned} \quad (81)$$

where ρ denotes the penalty factor and $\mathbf{u}_r$ denotes the scaled dual variable.

The problem (80) can be solved via alternating direction method of multipliers (ADMM) by updating the set of variables $\{\gamma'_r, \mathbf{z}_r, \mathbf{u}_r\}$ in an alternating manner. The scattering intensity γ'_r is updated by

$$\begin{aligned} \gamma'_r \leftarrow \arg \min_{\gamma'_r} &\sum_{k=1}^K \sum_{q=1}^{Q'} \lambda_{k,q} \gamma_{\beta,k,q} (\gamma'_{r,k,q} - \gamma'_{r,q})^2 \\ &+ \mu \sum_{q=1}^{Q'} \gamma'_{r,q} + \frac{\rho}{2} \|\mathbf{z}_r - \mathbf{D}\gamma'_r + \mathbf{u}_r\|_2^2. \end{aligned} \quad (82)$$

By setting the gradient of the objective function in (82) to be zero, we have

$$\begin{aligned} & \left(2 \sum_{k=1}^K \mathbf{B}_k \mathbf{\Lambda}_k + \rho \mathbf{D}^T \mathbf{D} \right) \gamma'_r \\ &= 2 \sum_{k=1}^K \mathbf{B}_k \mathbf{\Lambda}_k \gamma_{r,k} + \rho \mathbf{D}^T (\mathbf{z}_r + \mathbf{u}_r) - \mu \mathbf{1}_{Q'}, \end{aligned} \quad (83)$$

where $\mathbf{B}_k = \text{diag}\{\gamma_{\beta,k,q} : q \in [Q']\}$, $\mathbf{\Lambda}_k = \text{diag}\{\lambda_{k,q} : q \in [Q']\}$, and $\mathbf{1}_{Q'}$ denotes an all-one vector with dimension Q' . To avoid the computational burden of obtaining the inverse of the matrix $2 \sum_{k=1}^K \mathbf{B}_k \mathbf{\Lambda}_k + \rho \mathbf{D}^T \mathbf{D}$, we adopt the conjugate gradient method to find the solution γ'_r . Then, we project each element of γ'_r to $[0, 1]$ as follows

$$\gamma'_{r,q} \leftarrow \min(\max(\gamma'_{r,q}, 0), 1), \quad \forall q \in [Q']. \quad (84)$$

The subproblem corresponding to $\mathbf{z}_r$ is given by

$$\min_{\mathbf{z}_r} \eta \|\mathbf{z}_r\|_1 + \frac{\rho}{2} \|\mathbf{z}_r - \mathbf{D} \gamma'_r + \mathbf{u}_r\|_2^2. \quad (85)$$

Apply proximal operator for ℓ_1 -norm, we have

$$\mathbf{z}_r \leftarrow \text{prox}_{\eta/\rho \|\cdot\|_1}(\mathbf{D} \gamma'_r - \mathbf{u}_r) \quad (86)$$

$$= \text{sgn}(\mathbf{D} \gamma'_r - \mathbf{u}_r) \max\left(|\mathbf{D} \gamma'_r - \mathbf{u}_r| - \frac{\eta}{\rho}, 0\right), \quad (87)$$

where $\text{sgn}(\cdot)$ denotes the sign function, defined as $\text{sgn}(s) = -1$ for $s \leq 0$ and $\text{sgn}(s) = 1$ otherwise. The dual variable $\mathbf{u}_r$ is updated in the following way:

$$\mathbf{u}_r \leftarrow \mathbf{u}_r + (\mathbf{z}_r - \mathbf{D} \gamma'_r). \quad (88)$$

When γ'_r is fixed, the optimal $\lambda_{k,q}, \forall k, q$ admits a closed-form as follows

$$\begin{aligned} \lambda_{k,q} \leftarrow \arg \min_{\lambda_{k,q}} & \lambda_{k,q} \gamma_{\beta,k,q} (\gamma'_{r,k,q} - \gamma'_{r,q})^2 \\ & + (1 - \lambda_{k,q}) \gamma_{\beta,k,q} (\gamma'_{r,k,q})^2 \end{aligned} \quad (89)$$

$$= \begin{cases} 0, & (\gamma'_{r,q})^2 - 2\gamma'_{r,q} \gamma'_{r,k,q} > 0, \\ 1, & (\gamma'_{r,q})^2 - 2\gamma'_{r,q} \gamma'_{r,k,q} \leq 0. \end{cases} \quad (90)$$

This criterion ensures that the estimate $\gamma'_{r,k,q}$ is incorporated into the fusion only when it exhibits sufficient strength relative to $\gamma'_{r,q}$; otherwise, it is excluded to prevent invalid data from distorting the fusion result. The overall multi-view fusion algorithm is shown in Algorithm 2.

VI. SIMULATION RESULTS

In this section, we provide numerical results to verify the effectiveness of the proposed imaging scheme. The system setup is as follows. The transmitter is located at $[-3, 7.5]^T$ in meter. We assume that there are $K = 3$ receivers, located at $[18, 7.5]^T$, $[7.5, 18]^T$, and $[7.5, -3]^T$ in meter, respectively. Extended targets are located in the RoI $\mathcal{A}$ with $x_1 = y_1 = 0$ and $x_2 = y_2 = 15$. The power spectral density of the AWGN at receivers is -169 dBm/Hz and the channel bandwidth is 1 MHz. The path-loss of the path from the transmitter to the point scatter located in $[x, y]^T$ and then to the receiver k is given by $\gamma_{\beta,x,y,k} = \beta_0^2 d_{\text{tx},x,y}^{-2} d_{\text{rx},k,x,y}^{-2}$, where $d_{\text{tx},x,y}$ and $d_{\text{rx},k,x,y}$ denote the distance between the transmitter to the point scatter and the distance between the point scatter and

Algorithm 2 Multi-View Fusion

Input: The estimates $\gamma_{r,k}, \forall k$ at K receivers, grid positions $\mathbf{P}_k, \forall k$, RoI $\mathcal{A}$, dimension $Q' = Q'_1 \times Q'_2$, tolerances ϵ_1 and ϵ_2 , and iterations $\text{Iter}_{\max}$ and Iter_1 .

Output: The aggregated result γ'_r .

- 1: $\forall k$: Compute the gradient at grid $q \in [Q]$ based on (69).
- 2: $\forall k$: Compute the structure tensor at interpolation grid $q' \in [Q']$ based on (72) and perform SVD.
- 3: $\forall k$: Compute the edge-preserving weight in (74).
- 4: $\forall k$: Obtain EP-NNI result $\gamma'_{r,k}$ applying (75).
- 5: Initialize $\gamma'_r = \sum_k \gamma'_{r,k}$, $\mathbf{z}_r = \mathbf{D} \gamma'_r$, and $\mathbf{u}_r = \mathbf{0}$.
- 6: **for** $j = 1, 2, \dots, \text{Iter}_{\max}$ **do**
- 7: Update $\mathbf{\Lambda}$ according to (90).
- 8: **for** $j_1 = 1, 2, \dots, \text{Iter}_1$ **do**
- 9: Update γ'_r by solving (83) via conjugate gradient method and performing projection as in (84).
- 10: Update $\mathbf{z}_r$ according to (86).
- 11: Update $\mathbf{u}_r$ according to (88).
- 12: **end for**
- 13: If $\frac{\|\gamma_r^{(j)} - \gamma_r^{(j-1)}\|_2^2}{\|\gamma_r^{(j-1)}\|_2^2} \leq \epsilon_1$ and $\frac{\|\mathbf{\Lambda}^{(j)} - \mathbf{\Lambda}^{(j-1)}\|_F^2}{\|\mathbf{\Lambda}^{(j-1)}\|_F^2} \leq \epsilon_2$, stop.
- 14: **end for**

the k -th receiver, respectively, and $\beta_0 = -35$ dB measures the path-loss at the reference distance.

In Fig. 2, we present the single-view images at the first receiver (located at $[18, 7.5]^T$) applying the covariance-based scheme, in the cases with and without the cluster-promoting penalty, grid optimization, and the edge-preserving weight. We assume that both the transmitter and receiver are equipped with 16 antennas, the length of orthogonal pilot is $L = 16$, the number of frames is $M = 20$, and the transmit power is $P = 10$ dBm. We choose a triangle and a circle as extended targets with uniform scattering intensity across all scatterers. The contours of extended targets are plotted with red solid lines in Fig. 2. Without loss of generality, we assume that the RoI is within the field of view of the receiver to verify the single-view imaging performance. In Fig. 2, the detected grids with high scattering intensity are displayed, which are selected by sorting all detected grids according to their intensity and retaining the highest-intensity ones until their cumulative intensity reaches 95% of the total recovered intensity. In Fig. 2(a), we divide the RoI into $Q = 400$ uniform grids, and optimize the effective scattering intensity $\gamma_{r,k}$ with fixed grids based on the ML criterion (19). In Fig. 2(b), the cluster-promoting penalty (20) is taken into consideration for the optimization of $\gamma_{r,k}$. In Fig. 2(c) and Fig. 2(d), we optimize both grid positions and scattering intensity, using only the ML criterion and adding the cluster-promoting penalty, respectively. It is evident that by introducing the cluster-promoting penalty, which leverages the prior knowledge that extended targets typically occupy multiple contiguous grids, the recovered extended targets exhibit enhanced continuity and reduced internal voids. Moreover, grid optimization sharply concentrates the detected energy within the targets, significantly reducing sidelobes and yielding boundaries that closely align with the true shape. Fig. 2(e) and Fig. 2(f) are obtained by interpolating the imaging result in Fig. 2(d) onto finer $Q' = 60 \times 60$ grids via NNI and EP-NNI techniques, respectively. As we can see, in contrast to standard NNI, the edge-preserving weight in EP-NNI significantly

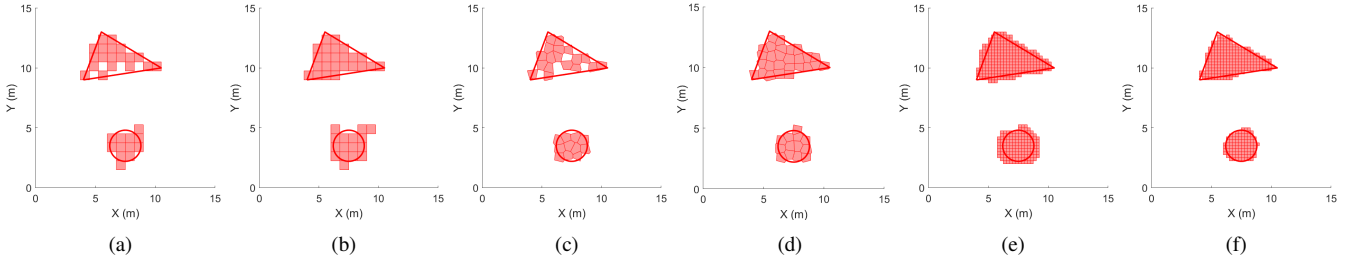


Fig. 2. Covariance-based single-view images at receiver 1 with $N_{tx} = 16, N_{rx} = 16, L = 16, M = 20, P = 10$ dBm, and $Q = 400$: (a) Fixed grid without penalty (20); (b) Fixed grid with penalty (20); (c) Optimized grid without penalty (20); (d) Optimized grid with penalty (20); (e) NNI; (f) EP-NNI.

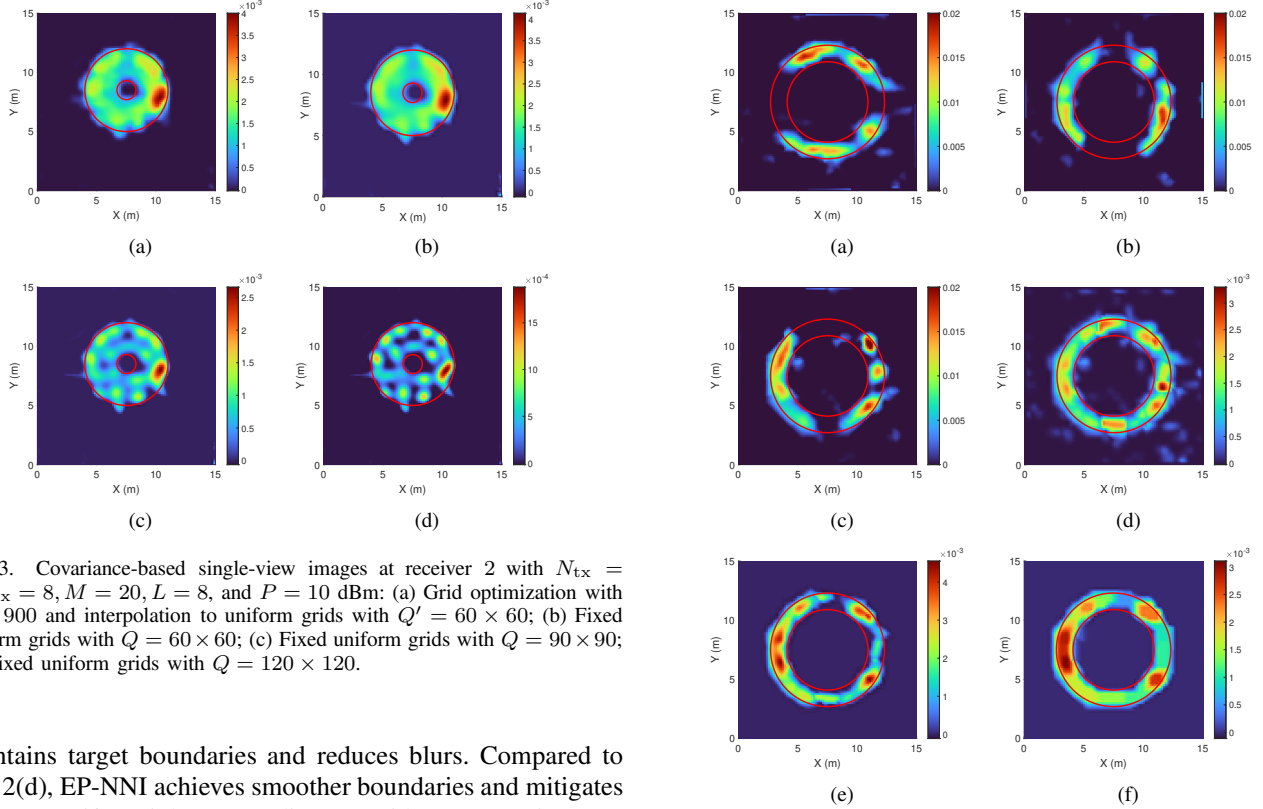


Fig. 3. Covariance-based single-view images at receiver 2 with $N_{tx} = 8, N_{rx} = 8, M = 20, L = 8$, and $P = 10$ dBm: (a) Grid optimization with $Q = 900$ and interpolation to uniform grids with $Q' = 60 \times 60$; (b) Fixed uniform grids with $Q = 60 \times 60$; (c) Fixed uniform grids with $Q = 90 \times 90$; (d) Fixed uniform grids with $Q = 120 \times 120$.

maintains target boundaries and reduces blurs. Compared to Fig. 2(d), EP-NNI achieves smoother boundaries and mitigates staircase artifacts inherent to discrete grid representations.

In Fig. 3, we present covariance-based single-view images at receiver 2 (located at $[7.5, 18]^T$) with $N_{tx} = N_{rx} = L = 8, M = 20$, and $P = 10$ dBm. We choose an annulus as the extended target. Fig. 3(a) is obtained by jointly estimating effective scattering intensity and grid positions with $Q = 900$ based on the criterion in (22), followed by EP-NNI onto finer uniform grids with $Q' = 60 \times 60$. In comparison, Figs. 3(b), 3(c), and 3(d) present the images using fixed uniform grids with $Q = 60 \times 60$, $Q = 90 \times 90$, and $Q = 120 \times 120$, respectively, where only the scattering intensity is optimized. As we can see, Figs. 3(a) and 3(b) present similar performance, i.e., the approach with dynamic but a small number of grids achieves comparable performance to fixed-grid reconstruction with a slightly higher grid density ($Q = 60 \times 60$), both clearly recovering the annular target. However, considering that the distance between grid points that can be distinguished is constrained by the numbers of antennas, when the number of grids is further increased to $Q = 90 \times 90$ and $Q = 120 \times 120$, the reconstructed images exhibit multiple hollow regions be-

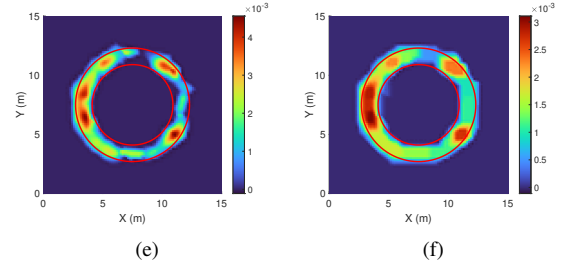


Fig. 4. Covariance-based single-view images and multi-view fusion results with $N_{tx} = N_{rx} = L = 8, M = 20, P = 0$ dBm, $Q = 900$, and $Q' = 60 \times 60$: (a) Image at receiver 1; (b) Image at receiver 2; (c) Image at receiver 3; (d) Multi-view fusion with WLS penalty; (e) Multi-view fusion with WLS and sparse penalties; (f) Multi-view fusion with WLS, sparse, and TV penalties.

yond the true inner ring, making it impossible to identify target structure. This demonstrates the significance of adopting dynamic grids instead of arbitrarily increasing grid density.

Fig. 4 shows single-view images at three receive BSs and fused images at the CPU, in the scenario with $N_{tx} = N_{rx} = L = 8, M = 20$, and $P = 0$ dBm. We assume that each receiver has a limited field of view, with a $\pi/8$ obstacle-induced angular blind sector centered on the boresight direction of the array. The covariance-based imaging results at three receivers are provided in Figs. 4(a), 4(b), and 4(c), respectively, each of which is obtained by jointly optimizing effective scattering intensity and grid positions. We can observe that each receiver captures only a part of the extended

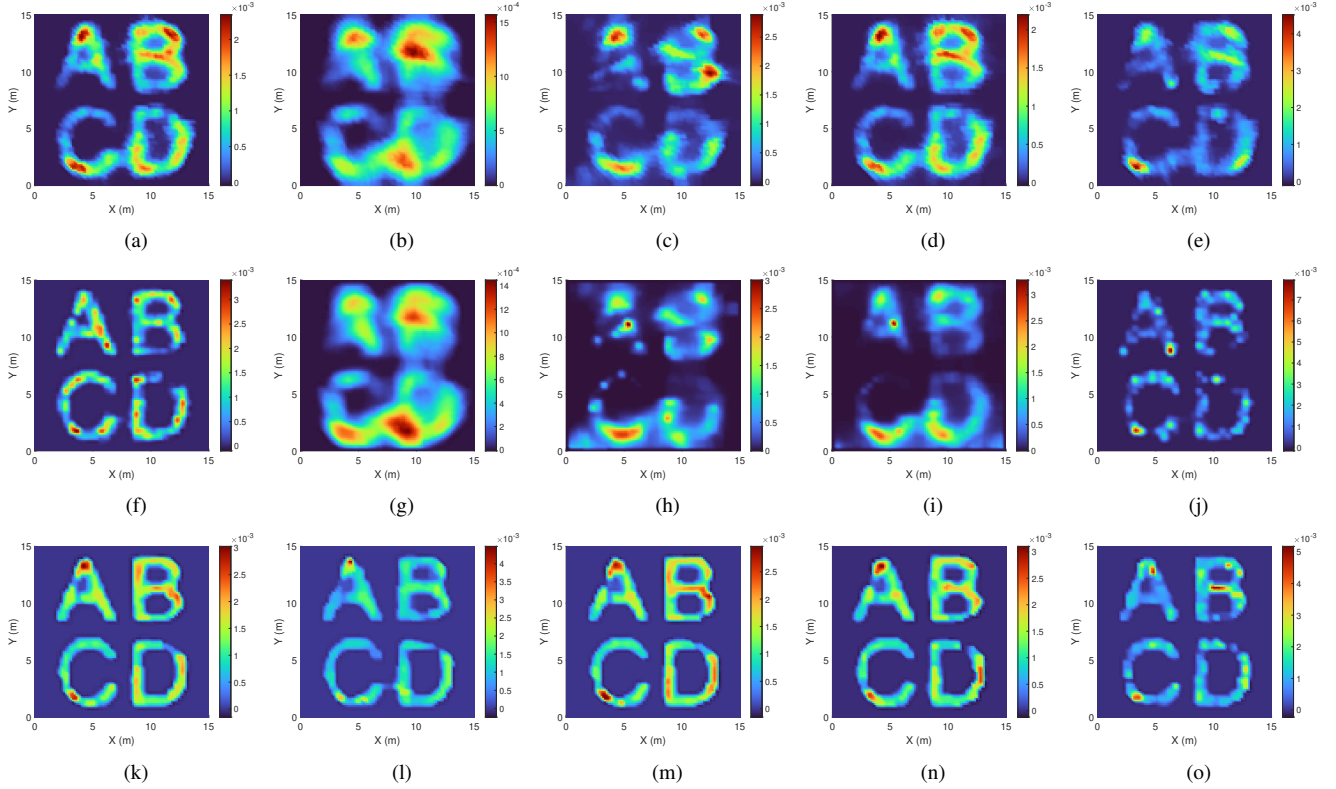


Fig. 5. Imaging performance comparison between different schemes with $K = 3$, $Q = 900$, and $Q' = 60 \times 60$. The first, the second, and the third rows present the images of benchmark scheme I, benchmark scheme II, and our proposed scheme, respectively. The first column is obtained assuming $N_{tx} = N_{rx} = L = 24$, $M = 20$, and $P = 10$ dBm. The second column is obtained assuming $N_{tx} = N_{rx} = L = 12$, $M = 20$, and $P = 10$ dBm. The third column is obtained assuming $N_{tx} = N_{rx} = L = 12$, $M = 20$, and $P = -10$ dBm. The fourth column is obtained assuming $N_{tx} = N_{rx} = L = 24$, $M = 20$, and $P = -10$ dBm. The fifth column is obtained assuming $N_{tx} = N_{rx} = L = 24$, $M = 5$, and $P = 10$ dBm.

TABLE I
IMAGING PERFORMANCE COMPARISON

		Column 1	Column 2	Column 3	Column 4	Column 5
Benchmark Scheme I	P-ISLR (dB)	-3.61	0.86	-1.62	-2.49	-3.60
	IoU	0.59	0.37	0.44	0.51	0.56
Benchmark Scheme II	P-ISLR (dB)	-9.05	0.79	-1.97	-2.54	-7.65
	IoU	0.81	0.39	0.41	0.46	0.70
Proposed Scheme	P-ISLR (dB)	-11.37	-7.77	-11.20	-7.65	-9.81
	IoU	0.89	0.82	0.88	0.82	0.84

target with poor quality. Fig. 4(d) presents the fused image based on the WLS criterion in (77), exhibiting serious spurious detections outside the extended target. Thus, we introduce the sparsity-promoting penalty (78), which effectively suppresses false alarms, as shown in Fig. 4(e). However, this comes at the cost of inducing structural fragmentation in the reconstruction. To address this problem, we additionally incorporate the TV penalty (79) to enhance the structural continuity by promoting gradient sparsity. As we can see from Fig. 4(f), this combined cost function yields substantially improved image quality. By leveraging diversity gain from multi-view observations, data fusion contributes to superior reconstruction performance compared to single-view images.

In Fig. 5, we compare the multi-view imaging performance of the proposed scheme with the following benchmark schemes:

- Benchmark Scheme I: Under this scheme, we first mul-

tiply the received signal $\mathbf{Y}_{k,m}$ by $\mathbf{X}^H$ to obtain virtual array signal, and then apply IFFT to transform this signal into the wavenumber domain. The single-view image is formed based on the angular information, followed by multi-frame accumulation to enhance the signal-to-noise ratio. The final image is obtained by applying the fusing algorithm proposed in Section V-B.

- Benchmark Scheme II: Based on the model in (13), we solve the CS problem applying the successive linear approximation variational Bayesian inference (SLA-VBI) algorithm proposed in [35], and obtain the estimation of both the sparse signals $\mathbf{c}_{k,m}$ and dynamic grid parameters $\mathbf{P}_k$. The final image is formed by accumulating $|\mathbf{c}_{k,q,m}|^2$ among M frames and applying the fusing algorithm proposed in Algorithm 2.

We choose four capital letters as extended targets. We assume that each receiver has a limited field of view, with a $\pi/5$

obstacle-induced angular blind sector centered on the line-of-sight direction between the transmitter and receiver. The first, the second, and the third rows of Fig. 5 present the multi-view images of benchmark scheme I, benchmark scheme II, and our proposed scheme, respectively. The first column of Fig. 5 is obtained assuming $N_{\text{tx}} = N_{\text{rx}} = L = 24$, $M = 20$, and $P = 10$ dBm. In the second column, we reduce N_{tx} , N_{rx} , and L to 12. In the third column, we consider the scenario with $N_{\text{tx}} = N_{\text{rx}} = 24$ and $L = 12$, where the non-orthogonal pilots are randomly selected from the sphere. The fourth column is obtained with lower transmit power $P = -10$ dBm. The fifth column is obtained using a smaller number of frames with $M = 5$. To quantitatively compare different schemes in different settings, Table I presents pseudo integrated sidelobe ratio (P-ISLR) and intersection over union (IoU) of the images in Fig. 5. P-ISLR is defined as [20]

$$\text{P-ISLR} = 10 \log_{10} \frac{\sum_{i \in \text{sidelobe}} \gamma'_{r,i}}{\sum_{i \in \text{mainlobe}} \gamma'_{r,i}}, \quad (91)$$

where all the grid cells within the true target are considered as the main lobe and those outside are considered as side-lobes. The IoU metric measures how well the target shape is reconstructed, defined as [14]

$$\text{IoU} = \frac{\text{Area}(S \cap \hat{S})}{\text{Area}(S \cup \hat{S})}, \quad (92)$$

where S denotes the spatial region occupied by extended targets, and $\hat{S}$ denotes its estimate obtained by sorting all detected grid cells according to corresponding intensities and retaining the highest-intensity ones until their cumulative intensity reaches 95% of the total recovered intensity. As we can see, the letters are well-reconstructed by the proposed scheme in all considered regimes, whereas benchmark schemes perform poorly, especially in the scenario with limited resources, such as a small number of transmit/receive antennas, low transmit power, or limited pilot length. Specifically, as shown in the second column, benchmark results exhibit severe sidelobes due to limited spatial resolution; in contrast, our proposed imaging scheme shows sharp target boundaries and enhanced sidelobe suppression, achieving at least 8 dB reduction in P-ISLR and twofold increase in IoU compared to benchmarks. As shown in columns 3 and 4, in the cases with limited pilot length and power, compared with benchmark methods which show blurred boundaries, our approach substantially reduces P-ISLR by nearly 10 dB and 5 dB, respectively, and doubles the IoU at the same time. In the scenario with limited frames, our scheme maintains superior performance as shown in the last column. The imaging performance comparison shown in Fig. 5 and Table I demonstrates the substantial performance improvement of the proposed scheme over benchmark schemes. The performance gain can be interpreted as the ability of our scheme to efficiently reduce the number of parameters to be estimated by exploiting channel statistical properties.

VII. CONCLUSION

This work investigated the problem of multi-view imaging for extended targets in the environment. An advanced two-phase approach was developed to address this problem. In Phase I, a novel covariance-based imaging framework

was proposed to generate single-view image by jointly estimating effective scattering intensity and grid positions at each receiver. Compared to FFT and CS based schemes that aim to estimate high-dimensional instantaneous channels, our covariance-based formulation dramatically reduced the number of variables to be estimated, thereby enabling high-quality imaging even in the scenario with narrowband and small aperture. Moreover, after grid optimization, the grid topology is better aligned with target geometries. In Phase II, we developed a multi-view fusion strategy to aggregate individual images from all receivers. Specifically, we designed an EP-NNI method to map heterogeneous images onto a set of common and finer grids, and formulated an optimization problem to estimate the fused scattering intensity and “informative” subset of receivers. Extensive numerical results were provided to verify the effectiveness of the proposed scheme, demonstrating its potential to support high-quality imaging in future 6G networks.

REFERENCES

- [1] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, “Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond,” *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.
- [2] L. Liu, S. Zhang, and S. Cui, “Leveraging a variety of anchors in cellular network for ubiquitous sensing,” *IEEE Commun. Mag.*, vol. 62, no. 9, pp. 98–104, Sep. 2024.
- [3] W. Zhu, S. Zhang, and L. Liu, “Joint transmission and compression optimization for networked sensing with limited-capacity fronthaul links,” to appear in *IEEE Trans. Wireless Commun.*, 2025.
- [4] E. Fishler, A. Haimovich, R. S. Blum, L. J. Cimini, D. Chizhik, and R. A. Valenzuela, “Spatial diversity in radars—Models and detection performance,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 823–838, 2006.
- [5] S. Hou, K. Solna, and H. Zhao, “A direct imaging algorithm for extended targets,” *Inverse Problems*, vol. 22, no. 4, pp. 1151–1178, Jun. 2006.
- [6] W. Yi, G. Li, and G. Battistelli, “Distributed multi-sensor fusion of PHD filters with different sensor fields of view,” *IEEE Trans. Signal Process.*, vol. 68, pp. 5204–5218, 2020.
- [7] Z. Tian, Z. Zhang, and L. Hanzo, “Distributed multi-view sparse vector recovery,” *IEEE Trans. Signal Process.*, vol. 71, pp. 1448–1463, 2023.
- [8] S. K. Dehkordi, L. Pucci, P. Jung, A. Giorgetti, E. Paolini, and G. Caire, “Multistatic parameter estimation in the near/far field for integrated sensing and communication,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 17 929–17 944, Dec. 2024.
- [9] L. M. Ulander, H. Hellsten, and G. Stenstrom, “Synthetic-aperture radar processing using fast factorized back-projection,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 760–776, Jul. 2003.
- [10] S. Patole and M. Torlak, “Two dimensional array imaging with beam steered data,” *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5181–5189, Dec. 2013.
- [11] Y. Huang, J. Yang, W. Tang, C.-K. Wen, and S. Jin, “Fourier transform-based wavenumber domain 3D imaging in RIS-aided communication systems,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13 872–13 888, Oct. 2024.
- [12] X. Tong, Z. Zhang, J. Wang, C. Huang, and M. Debbah, “Joint multi-user communication and sensing exploiting both signal and environment sparsity,” *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1409–1422, Nov. 2021.
- [13] A. Wu, Y. Wu, Y. Jin, Y. Wang, and Z. Guo, “ $L_{(1/2)}$ regularization for ISAR imaging and target enhancement of complex image,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5200410.
- [14] I. Roldan, F. Fioranelli, and A. Yarovsky, “Total variation compressive sensing for 3D shape estimation in short-range imaging radars,” *IEEE Trans. Radar Systems*, vol. 1, pp. 583–592, 2023.
- [15] K. Zhi, T. Yang, S. Li, Y. Song, A. Rezaei, and G. Caire, “Near-field integrated imaging and communication in distributed MIMO networks,” *arXiv preprint arXiv:2508.17526*, 2025.

- [16] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 806–811, Aug. 1985.
- [17] Q. Shi, L. Liu, S. Zhang, and S. Cui, "Device-free sensing in OFDM cellular network," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1838–1853, Jun. 2022.
- [18] Q. Shi and L. Liu, "Joint LOS identification and data association for 6G-enabled networked device-free sensing," *IEEE Trans. Commun.*, vol. 72, no. 8, pp. 5117–5129, Aug. 2024.
- [19] J. Gao, W. Zhu, S. Zhang, Y. Wu, J. Cao, G. Caire, and L. Liu, "Integrated massive communication and target localization in 6G cell-free networks," *arXiv preprint arXiv:2510.14281*, 2025.
- [20] M. Manzoni, D. Tagliaferri, S. Tebaldini, M. Mizmizi, A. V. Monti-Guarnieri, C. M. Prati, and U. Spagnolini, "Wavefield networked sensing: Principles, algorithms and applications," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 181–197, Jan. 2025.
- [21] D. Tagliaferri, M. Manzoni, M. Mizmizi, S. Tebaldini, A. V. Monti-Guarnieri, C. M. Prati, and U. Spagnolini, "Cooperative coherent multi-static imaging and phase synchronization in networked sensing," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 10, pp. 2905–2921, Oct. 2024.
- [22] Q. Shi, S. Zhang, and L. Liu, "A 6G-based multi-view reconstruction approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Milan, Italy, 2025.
- [23] S. Mandelli, M. Henninger, and J. Du, "Sampling and reconstructing angular domains with uniform arrays," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3628–3642, Jun. 2022.
- [24] Z. Li, A. Dubey, S. Shen, N. K. Kundu, J. Rao, and R. Murch, "Radio tomographic imaging with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 15 784–15 797, Nov. 2024.
- [25] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*. New York: McGraw-Hill, 1953.
- [26] M. Skolnik, *Introduction to Radar Systems*. McGraw-hill, 2001.
- [27] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124–127, 1950.
- [28] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image process.*, vol. 10, no. 10, pp. 1521–1527, 2001.
- [29] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. Boca Raton, FL, USA: Routledge, 2018.
- [30] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner Stuttgart, 1998, vol. 1.
- [31] R. Sibson, "A brief description of natural neighbour interpolation," *Interpreting multivariate data*, pp. 21–36, 1981.
- [32] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, 2007.
- [33] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
- [34] C. Li, W. Yin, H. Jiang, and Y. Zhang, "An efficient augmented Lagrangian method with applications to total variation minimization," *Comput. Optim. Appl.*, vol. 56, no. 3, pp. 507–530, Dec. 2013.
- [35] W. Xu, A. Liu, B. Zhou, and M.-j. Zhao, "Successive linear approximation VBI for joint sparse signal recovery and dynamic grid parameters estimation," *to appear in IEEE Trans. Wireless Commun.*, 2025.