TwinFocus: Autofocus for Handheld mmWave SAR Imaging via Physical and Digital Twin References

Yadong Li
yadongli@uw.edu
University of Washington
Seattle, WA, USA

Xinghua Sun
xinghua@uw.edu
University of Washington
Seattle, WA, USA

Qiancheng Li
qiancl2@uw.edu
University of Washington
Seattle, WA, USA

Akshay Gadre
gadre@uw.edu
University of Washington
Seattle, WA, USA

Abstract

Millimeter-wave (mmWave) imaging is increasingly being adopted across the supply chain and security industries to see through occlusions and detect hidden objects. Yet, most industrial mmWave imaging relies on a static, bulky, and expensive infrastructure to achieve precisely measured positions required for synthetic aperture imaging. In contrast, a handheld mmWave scanner can provide a compact, mobile and cost-effective alternative for high-resolution mmWave imaging, but it suffers from severe phase errors and image distortions due to motion inaccuracy.

This paper presents *TwinFocus*, a twin-reference-guided autofocus framework that overcomes the impact of these motion errors to enable practical handheld mmWave imaging. Our key insight is that motion-induced phase errors remain similar across objects in the scene. *TwinFocus* therefore leverages a reference object in the environment, whose synthetic mmWave response can be generated using lightweight digital twins, to estimate the phase errors affecting the target object. It then compensates for these impairments by aligning the defocused reference image with its twin representation via amplitude-domain image optimization. Extensive real-world evaluations and ablation studies across diverse object configurations, materials and occlusions demonstrate a 23.3% and 41.7% improvement in structural similarity using digital and physical twin references, respectively.

CCS Concepts

• **Computer systems organization → Sensors and actuators**.

Keywords

Millimeter-Wave Imaging, Synthetic Aperture Radar, Autofocus Digital Twin

ACM Reference Format:

Yadong Li, Xinghua Sun, Qiancheng Li, and Akshay Gadre. 2026. TwinFocus: Autofocus for Handheld mmWave SAR Imaging via Physical and Digital Twin References. In *The 24th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '26), June 21–25, 2026, Cambridge, United Kingdom.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3745756.3809185

MobiSys '26, June 21–25, 2026, Cambridge, United Kingdom

ACM ISBN 979-8-4007-2027-7/2026/06
https://doi.org/10.1145/3745756.3809185

1 Introduction

Radio Frequency (RF)-based imaging has seen widespread adoption across diverse real-world applications, including automotive sensing [10, 12], security screening [13, 14], non-destructive testing [32], and structural health monitoring [47]. Among RF modalities, millimeter-wave (mmWave) radar imaging stands out for its high spatial resolution and reasonable penetration capabilities, making it especially valuable for detailed object profiling through occlusions [21, 41] or packaging materials [5, 9, 32]. Typically, mmWave imaging systems are built upon synthetic aperture radar (SAR) techniques, which enable high-resolution imaging by synthesizing a large virtual antenna array through controlled radar movement [6]. However, due to its small wavelength, traditional mmWave SAR relies on bulky and expensive motion stages that move the radar with sub-millimeter accuracy to ensure coherent signal integration. Consequently, current mmWave imaging platforms remain unsuitable for mobile and ubiquitous deployment.

Recent advances in handheld mmWave SAR imaging seek to overcome this limitation by enabling mobility without sacrificing image quality. Prior work [49] compensates for the non-linear and non-uniform nature of handheld motion by integrating visual-inertial odometry (VIO) [25, 26] to estimate radar trajectories. Yet, the localization accuracy of existing VIO systems is inadequate for phase-sensitive mmWave SAR imaging. In fact, due to the small wavelength of mmWave signals, even sub-millimeter tracking errors (typical for VIO systems) introduce severe phase misalignments across the virtual aperture, leading to significantly defocused images [30]. While several solutions have proposed external calibration with corner reflectors [16] or accumulating radar returns over extended periods [49], they increase deployment constraints and user burden. Deep learning-based approaches [15, 17] have also been explored but require large training datasets and generalize poorly to unseen targets or environments.

In this paper, we introduce *TwinFocus*, a twin-reference-guided autofocus framework that compensates for these severe phase errors to enable practical handheld mmWave SAR imaging. Our key idea *infers and corrects motion-based phase errors for the scene by*

[0]Demo video : https://youtu.be/RMNm82wTOhg

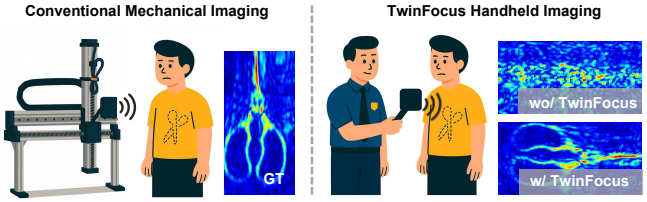

Figure 1: Conventional SAR imaging requires bulky and expensive motion stages while *TwinFocus* enables mobile and affordable mmWave imaging by overcoming severe phase errors caused by handheld motion.

minimizing the discrepancy observed from a reference object visible in the environment. Specifically, we identify two objects within the scene: a reference object, either manually deployed or naturally present in the environment, and a target object of interest that may lack a direct line of sight. *TwinFocus* starts with generating the ideal SAR response of the reference using either physical or digital twin, then develops a phase-error estimation framework that captures the phase misalignment between the defocused reference RF image and its well-focused counterpart. We then transfer this phase correction to the rest of the scene to compensate for phase errors in the target region to recover a well-focused image of the unknown target. Notably, unlike our prior work [16] that relies solely on corner reflectors that are rarely available in real-world scenes, the image-level alignment in *TwinFocus greatly relaxes the constraints on reference object geometry.*

We first perform a motivation study on handheld imaging by combining a mmWave radar [35] with a ZED 2i stereo camera [34] for VIO tracking. Our measurements (Sec. 2) show that although the tracking camera achieves sub-millimeter accuracy, this precision is still insufficient to guarantee high-quality handheld SAR images. Our theoretical analysis (Sec. 3) of the spatial distribution of phase errors further demonstrates that the phase errors remain similar across objects under certain conditions. This observation suggests that, instead of estimating phase errors for each target independently, we can recover them from a chosen reference and propagate the correction across the scene. Motivated by this opportunity, *TwinFocus* achieves handheld mmWave SAR by estimating motion-induced phase errors from a reference object and transferring the correction to the unknown target in two complementary scenarios:

1. Physical twin-guided autofocus for known references: In cooperative imaging that follow predefined procedures, such as security screening or industrial inspection, it is feasible to deploy a known reference object whose ideal mmWave response (i.e., its physical twin) can be pre-measured. Given this physical twin template, instead of directly matching the response in the unstable complex domain, *TwinFocus* performs robust optimization in the amplitude domain where phase errors manifest as observable blurring and distortion. We design an iterative, differentiable optimization process that estimates the phase errors by perturbing the phase profile across the synthetic array, progressively aligning the defocused reference amplitude with its ideal counterpart. The resulting phase correction is then transferred to the target region for high-fidelity reconstruction of unknown targets.

2. Digital twin-based autofocus for unknown references: In non-cooperative scenarios where the ground-truth reference SAR response is unavailable, *TwinFocus* extends its approach by generating a digital twin of the reference object through a lightweight visual-to-radar simulation pipeline. Specifically, we employ the tracking stereo camera to capture both the geometry of the unknown reference in addition to the coarse radar trajectory. *TwinFocus* efficiently simulates an aligned SAR image of the reference using a point-scattering model. Importantly, *TwinFocus* bypasses the need for phase-aligned simulation, which is extremely challenging due to the sub-millimeter precision required for modeling object geometry and radar arrays and instead utilizes the simulated SAR amplitude to guide phase error estimation. *TwinFocus* further employs a similarity-based optimization that prioritizes high-level visual consistency over strict pixel-wise alignment, thereby reducing sim-to-real discrepancies.

We implement *TwinFocus* using a Texas Instruments (TI) MMWCAS-RF-EVM four-chip cascaded radar [35], a ZED 2i stereo camera [34], and a 3D motion stage for ground-truth data collection. We develop a custom handheld platform that combines the tracking stereo camera with the mmWave radar for capturing real-world mmWave imaging using handheld trajectories. We perform extensive evaluations across various tracking error levels and object configurations, demonstrating robustness to variations in objects' shape, size, material, and environment under both LoS and NLoS scenarios. Using the structural similarity index measure (SSIM) as the evaluation metric, experimental results show that, *TwinFocus* outperforms the no-correction baseline by 41.7% with a physical twin for known references and by 23.3% with a digital twin for unknown references, while improving SSIM from 0.65 to 0.94 in the NLoS security case study.

Our contributions can be summarized as follows:

- We propose *TwinFocus*, a twin-reference-driven motion-induced phase error compensation framework that overcomes sub-millimeter tracking errors in handheld mmWave SAR imaging.
- We introduce an amplitude-domain alignment algorithm for phase error estimation of cooperative targets, along with a lightweight digital-twin pipeline to extend it for non-cooperative targets.
- We implement *TwinFocus* and validate its performance across diverse scenarios. Experiments demonstrate successful image reconstruction from heavily distorted signals and show the practical utility of *TwinFocus* in real-world tasks.[1]

Limitations: Despite the strong performance of *TwinFocus*, there are several opportunities for improvement (more details in Sec. 7): (1) extending robustness to larger VIO tracking errors, (2) better handling differences in reference and target size or orientation, and (3) improving support for composite reference objects through more accurate material-aware modeling.

[1]Code and data are available at https://github.com/leeyadong/TwinFocus

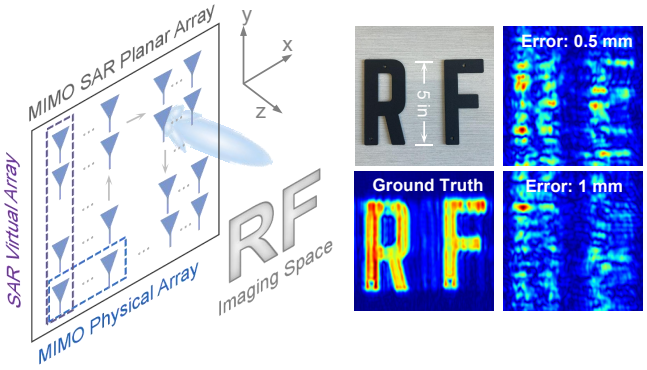

Figure 2: MIMO SAR imaging setup, where sub-mm tracking errors cause significant image distortion.

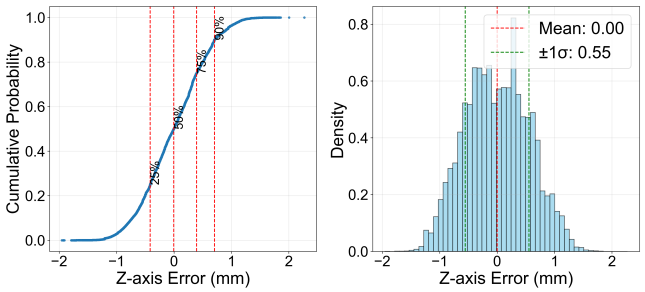

Figure 3: Error distribution analysis of ZED 2i camera z-axis tracking errors: ECDF (left) and histogram (right).

2 Challenges of Handheld Imaging

This section outlines the fundamentals of mmWave SAR imaging and explores the key sources of phase errors that degrade image quality significantly.

MIMO SAR Primer: 3D mmWave imaging requires a two-dimensional aperture to resolve scatterers in both azimuth and elevation. In this work, we adopt a MIMO–SAR configuration in which a linear multi-antenna array is scanned over a planar path to synthesize a large 2D virtual aperture efficiently (Fig. 2).

Signal Model. Let the x-axis denote azimuth, the y-axis elevation, and the z-axis depth. The positions of the physical transmit (TX) and receive (RX) antennas can be written as: $\mathbf{p}_T = (x_T,\ y_T,\ 0)$, $\mathbf{p}_R = (x_R,\ y_R,\ 0)$, and a voxel inside the target imaging volume O can be denoted as $\mathbf{r} = (x,\ y,\ z)$. The wavenumber is $k = 2\pi f/c$ where f is the signal frequency and c is the speed of propagation. Under the Born approximation [33], and assuming isotropic antennas, the multistatic baseband signal can be converted into an equivalent monostatic array [42]:

$$s(\mathbf{p}_c, k) = \iiint_O \sigma(\mathbf{r})\, e^{-j2kR(\mathbf{r};\mathbf{p}_c)}\, d\mathbf{r}, \tag{1}$$

where $\mathbf{p}_c$ is the phase center (i.e., midpoint) between a TX-RX pair, $\sigma(\mathbf{r})$ is the complex reflectivity of the target, and $R(\mathbf{r};\mathbf{p}_c)$ is the distance from the target voxel to the antenna.

Back-Projection Imaging. We aim to reconstruct the target voxel response $\sigma(\mathbf{r})$ from Eq. (1) by delineating the contribution of that voxel to the antenna array. Prior approaches have employed either time-domain methods [29] (e.g., back-projection) or frequency-domain methods [20] (e.g., range migration) to extract this response. Although frequency-domain approaches are more efficient via FFT, they require uniform, linear sampling conditions that are difficult to meet in handheld SAR. Hence, we adopt back-projection for its robustness to irregular trajectories. The target voxel response can then be reconstructed as:

$$\sigma(\mathbf{r}) = \iint_{\mathbf{p}_c}\int_k s(\mathbf{p}_c, k)\, e^{j2kR(\mathbf{r};\mathbf{p}_c)}\, dk\, d\mathbf{p}_c. \tag{2}$$

Eq. (2) coherently sums contributions from all virtual antenna positions and frequencies, yielding a focused 3D mmWave image.

Challenge of Motion Error: We empirically demonstrate how motion errors impact SAR image quality through a measurement study that visualizes the effects of varying tracking errors on image focus. Using the Texas Instruments (TI) MMWCAS mmWave radar [35] mounted on a 2D motion stage, we perform MIMO SAR imaging of 5-inch metallic letters. Implementation details are provided in Sec. 5.

The ground-truth image obtained with precise motion control is shown in Fig. 2. In contrast, handheld operation introduces non-uniform, non-linear motion, causing phase errors across the synthetic array and, in turn, incoherent voxel combination (Eq. 2). Due to the small wavelength of mmWave signals, even sub-wavelength Gaussian tracking error (0.5 mm) significantly blurs the reconstructed image (Fig. 2, right). Previous work [48, 51] mitigates this using high-precision motion capture systems with multiple reflective markers and cameras, but such setups cost ~$5000 [23] and require line-of-sight conditions.

We further evaluate a more portable and cost-effective tracking solution using a ZED 2i stereo camera [34] for visual-inertial odometry (VIO). Most distortions stem from errors along the z-axis (range direction), where even minor distance variations introduce large phase offsets [11, 16]. As shown in Fig. 3, the plotted z-error represents the deviation of the estimated z-position from a fitted 2D plane. This captures range-direction jitter even when no intentional z-motion is introduced. Our measurements of the camera's z-axis tracking error demonstrate that its accuracy is insufficient for high-quality SAR imaging. Therefore, it is essential to develop algorithms capable of correcting sub-wavelength tracking errors to enable practical handheld mmWave SAR imaging.

3 Understanding the Impact of Motion Error on SAR Images

This section explores why small motion errors lead to such significant degradation in SAR image quality and how these errors are related to phase offsets across objects.

Theoretical Model of Motion Error: We first identify the impact of the motion errors by quantifying the translation into phase errors and the amount of defocus they cause by extending the signal model from Sec. 2.

Acquisition with Position Error: As shown in Fig. 4 (left), let's say we only know the noisy locations of each capture $\tilde{\mathbf{p}}_c = \mathbf{p}_c + \Delta\mathbf{p}$ with position error $\Delta\mathbf{p} = (\Delta x,\ \Delta y,\ \Delta z)^\top$. Using this erroneous position

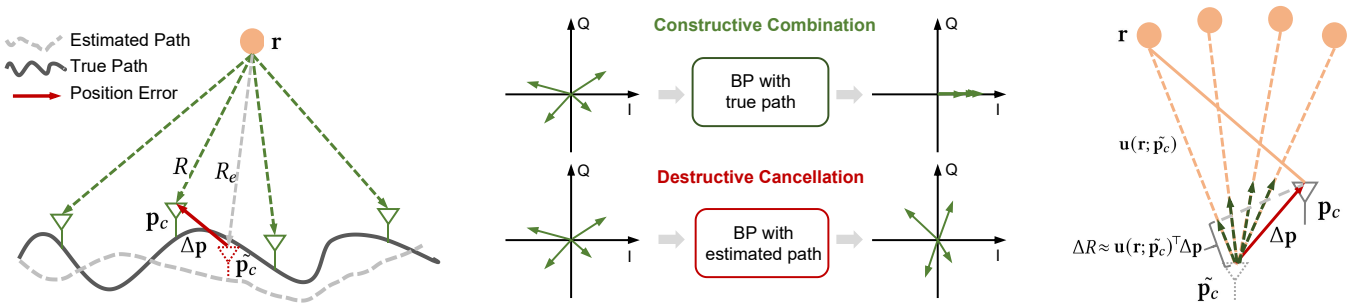

Figure 4: (1) Left: Signal acquisition with position error. (2) Middle: Position errors lead to incorrect phase alignment, causing the signals to cancel rather than reinforce each other. (3) Right: Approximating the phase error as the projection of the position error along the LoS direction reveals the similarity of phase errors across nearby objects.

in back-projection yields the distorted image:

$$\tilde{\sigma}(\mathbf{r}) = \iint_{\mathbf{p}_c} \int_k s(\mathbf{p}_c, k)\, e^{j2kR_e(\mathbf{r};\tilde{\mathbf{p}}_c)}\, dk\, d\mathbf{p}_c, \tag{3}$$

where $R_e(\mathbf{r};\tilde{\mathbf{p}}_c)$ is the distance between each erroneous antenna position $\tilde{\mathbf{p}}_c$ and the target point $\mathbf{r}$.

Residual Phase Term. Substituting the signal model of Eq. (1) into (3) and defining the slant range error $\Delta R(\mathbf{r};\mathbf{p}_c;\tilde{\mathbf{p}}_c) = R(\mathbf{r};\mathbf{p}_c) - R_e(\mathbf{r};\tilde{\mathbf{p}}_c)$, we obtain:

$$\tilde{\sigma}(\mathbf{r}) = \iiint \sigma(\mathbf{r}) \left[\iint_{\tilde{\mathbf{p}}_c} \int_k e^{-j2k\,\Delta R(\mathbf{r};\mathbf{p}_c;\tilde{\mathbf{p}}_c)}\, dk\, d\tilde{\mathbf{p}}_c \right] d\mathbf{r}. \tag{4}$$

Eq. (4) shows that motion error introduces a space-variant residual phase $e^{-j2k\Delta R}$. Consequently, the phase cannot be globally factorized and manifests as blurring and ghosting in the reconstructed image. Fig. 4 (middle) illustrates the received signals at different measurement locations as vectors in the IQ complex plane. We observe that using the true propagation path in back-propagation imaging enables accurate phase compensation across all measurement positions, aligning the signal phases and allowing constructive combination. However, in the presence of motion error, the residual phase terms prevent proper phase alignment, causing the signals to partially cancel rather than reinforce each other, which ultimately leads to image distortion. This analysis also reveals that *it is sufficient to compensate for the phase errors to reconstruct high-quality SAR images* rather than the actual motion errors. This is particularly true for *TwinFocus*, where the complementary stereo camera ensures sub-wavelength motion errors, making the resulting phase errors locally convex and avoiding phase wrapping.

Opportunity of Phase Error Transfer. The above formulation, however, reveals a new opportunity to capture phase errors. By analyzing the spatial characteristics of phase error distribution, we show that in a scene with two targets, the phase error estimated from one reference target can serve as an accurate proxy for the others.

Spatial Distribution of Residual Phases. Since $||\Delta\mathbf{p}||$ is small (on the millimeter scale) compared to the slant range (tens of centimeters), we can use a first-order Taylor expansion to approximate the slant range error $\Delta R(\mathbf{r};\mathbf{p}_c;\tilde{\mathbf{p}}_c)$, as shown in Fig. 4 (right):

$$\Delta R(\mathbf{r};\mathbf{p}_c;\tilde{\mathbf{p}}_c) \;\approx\; \mathbf{u}(\mathbf{r};\tilde{\mathbf{p}}_c)^\top \Delta\mathbf{p}, \qquad \mathbf{u}(\mathbf{r};\tilde{\mathbf{p}}_c) = \frac{\mathbf{r} - \tilde{\mathbf{p}}_c}{\|\mathbf{r} - \tilde{\mathbf{p}}_c\|_2} \tag{5}$$

where $\mathbf{u}(\mathbf{r};\tilde{\mathbf{p}}_c)$ denotes the unit line-of-sight (LoS) vector between the erroneous antenna position and the target.

Let $\mathbf{r}_1$ and $\mathbf{r}_2$ be two scatterers observed at the same antenna position $\tilde{\mathbf{p}}_c$. Using (5), the difference between the phase errors of two scatters can be modeled as:

$$\Delta\phi_1 - \Delta\phi_2 \;=\; \frac{4\pi}{\lambda}\,(\mathbf{u}_1 - \mathbf{u}_2)^\top \Delta\mathbf{p}, \tag{6}$$

where $\mathbf{u}_i = (\mathbf{r}_i - \tilde{\mathbf{p}}_c)/\|\mathbf{r}_i - \tilde{\mathbf{p}}_c\|_2$ for $i \in \{1, 2\}$. Then we can obtain the upper bound for Eq. (6) by applying the Cauchy–Schwarz inequality:

$$\left|\Delta\phi_1 - \Delta\phi_2\right| \;\le\; \frac{4\pi}{\lambda}\,\|\mathbf{u}_1 - \mathbf{u}_2\|_2 \|\Delta\mathbf{p}\|_2 \tag{7}$$

Implication. Eq. (7) and Fig. 4 (right) indicate that the phase error difference between two targets is primarily determined by their angular separation and the magnitude of the motion error. Importantly, when the two targets lie within a small angular range, their phase errors become highly correlated. For $\lambda \approx 4$ mm (77 GHz) and a trajectory error of $\|\Delta\mathbf{p}\| \le 0.5$ mm of the VIO systems (Fig. 3), targets within $\theta \le 20^\circ$ exhibit a maximum phase mismatch of $\le 0.17\pi$, well within the tolerance of standard autofocus refinement. It is also worth noting that: (1) This value represents a worst-case upper bound and the practical error is often smaller. (2) Since mmWave SAR is dominated by specular reflections and only geometry-compatible surfaces produce strong returns, two strong scatterers that are both far apart and simultaneously visible are rare, making the above bound more conservative.

While prior work has also used reference targets to manipulate mmWave signal phases for sensing and imaging [16, 45], our approach differs fundamentally in both goal and methodology. For instance, Mobi2Sense [45] eliminates all motion-induced phase changes using static references, whereas our method selectively corrects only undesired motion while preserving the phase variation necessary for virtual aperture synthesis. Additionally, our previous study's [16] references are limited to idealized point targets (i.e. corner reflectors) that are often unavailable in real-world scenarios. In contrast, this paper establishes a theoretical foundation for phase error transfer and significantly relaxes the requirements on reference geometry, thus improving the practicality of handheld SAR imaging.

4 *TwinFocus* Design

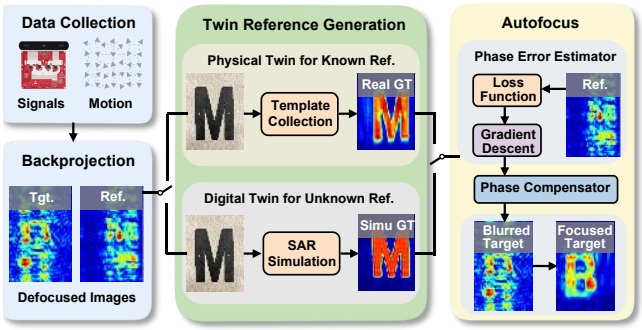

Figure 5: *TwinFocus* overview. It generates either a physical twin or a digital twin for phase error estimation by optimizing the defocused reference image and the template. The optimal phase vector is then applied to reconstruct the actual imaging target.

4.1 System Overview

The design of *TwinFocus* is grounded in the theoretical modeling discussed in Sec. 3, which shows that phase errors across neighboring spatial locations exhibit local similarity. This motivates the use of a specific reference target, an object whose SAR response is known, to infer the unknown trajectory-induced phase error, which can then be applied to correct the image of other targets scanned along the same trajectory. However, implementing this concept requires solving two fundamental challenges:

Phase Optimization: Even with a reference SAR template from prior data collection, phase-level optimization remains challenging due to its extreme sensitivity to sub-millimeter modeling errors and motion inaccuracies.

Unknown References: In many practical situations, the reference SAR response is not available in advance and thus remains unknown. How can we generate the reference SAR response on the fly for phase error calibration?

Fig. 5 shows how *TwinFocus* addresses these challenges: (1) when a real SAR template of the reference is available, instead of matching the exact complex-domain phase, which is highly sensitive to small distance variations, we estimate the motion-induced phase error using only the amplitude information, which is more robust; and (2) we introduce a lightweight digital twin model that simulates the SAR response of an unknown reference object in the environment, which can then be used for phase error correction.

4.2 Physical Twin Reference for Cooperative Autofocus

We first consider a cooperative SAR imaging scenario, where targets are scanned under predefined procedures. Such setups are common in applications such as security screening and industrial inspection. In these cases, a known object can be placed in the scene as a reference target. Because this reference is fixed and predefined, we can pre-collect its ideal SAR response (i.e., its mmWave physical twin) using a motion stage and store it as a template for phase error compensation during later scans.

Phase Optimization: Complex Domain vs. Amplitude Domain. Intuitively, if the true reflectivity of the reference target, including both amplitude and phase, were known, one could directly optimize the complex SAR response. In practice, however, measuring or modeling the phase response of arbitrary real-world objects is infeasible due to material properties, multipath effects, and calibration inconsistencies. Moreover, directly comparing complex-valued SAR images is highly unstable, as even minor geometric errors can cause large, unpredictable phase shifts. We overcome this by leveraging the fact that phase errors manifest as visible blurring and distortion in the amplitude domain. This motivates an amplitude-domain formulation where we avoid dealing with the unstable phase directly.

Let $\mathcal{B}(\boldsymbol{\phi}) \in \mathbb{C}^{H\times W}$ denote the back-projected complex SAR image reconstructed after applying a per-pulse phase correction $e^{-j\phi}$ to the raw data. The corresponding real-valued amplitude image is $\mathcal{A}(\boldsymbol{\phi}) = |\mathcal{B}(\boldsymbol{\phi})| \in \mathbb{R}^{H\times W}$. Given the amplitude of the reference template $\mathcal{A}_{\text{ref}}$, our goal is to find the phase error vector $\boldsymbol{\phi}$ that minimizes the discrepancy between the current defocused image $\mathcal{A}(\boldsymbol{\phi})$ and the reference amplitude template $\mathcal{A}_{\text{ref}}$. This optimization problem can be formulated using an image similarity loss $\mathcal{L}(\cdot,\cdot)$:

$$\min_{\boldsymbol{\phi}} \ \mathcal{L}\left(\mathcal{A}(\boldsymbol{\phi}), \mathcal{A}_{\text{ref}}\right), \tag{8}$$

Loss Function. We carefully select the loss function $\mathcal{L}(\cdot,\cdot)$ as it must reliably guide the optimizer to the correct motion-induced phase errors while being robust to the noise characteristics of SAR images. This can be achieved by measuring high-level similarity between the defocused and reference amplitude images. We adopt the Structural Similarity Index Measure (SSIM) as our loss function to balance robustness, interpretability, and differentiability. SSIM compares images based on luminance, contrast, and local structural patterns, and is well-known to be more robust to pixel-wise speckle noise and better reflects human-perceived visual similarity. Since SSIM ranges from 0 to 1, with higher values indicating greater similarity, the SSIM-based loss is defined as:

$$\mathcal{L}_{\text{SSIM}} = 1 - \text{SSIM}\left(\mathcal{A}(\boldsymbol{\phi}), \mathcal{A}_{\text{ref}}\right), \tag{9}$$

which is smooth, differentiable, and efficient for guiding optimization using solvers such as Adam.

System Workflow. As shown in Fig. 5, for cooperative scenes where the reference object is available in advance, *TwinFocus* proceeds as follows: (1)*Reference Template Acquisition:* Capture and store a clean SAR image $\mathcal{A}_{\text{ref}}$ of a reference target using a motion stage. (2) *Handheld Scanning:* Collect SAR data under freehand motion, which introduces unknown phase errors. (3) *Phase Error Estimation:* Estimate $\widehat{\boldsymbol{\phi}}$ by aligning the defocused reference image to $\mathcal{A}_{\text{ref}}$ using the SSIM loss. (4) *Phase Correction and Imaging:* Apply $\widehat{\boldsymbol{\phi}}$ to reconstruct a focused image of the unknown target.

4.3 Digital Twin Reference for Non-Cooperative Autofocus

Cooperative reference targets, whose SAR responses can be pre-captured, may not be available in all application scenarios. To extend

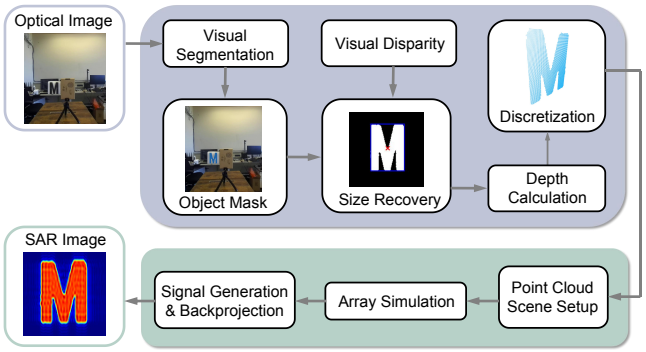

Figure 6: *TwinFocus*' digital twin framework that converts the visual image of the reference to its corresponding simulated mmWave SAR image.

the practical deployment of our system, we further propose using non-cooperative objects in the environment as reference targets, even when their SAR responses are unknown.

Key idea. Instead of requiring a pre-scanned reference object, we simulate its SAR response using a lightweight visual-to-radar digital twin pipeline, as shown in Fig. 6. Specifically, our amplitude-domain optimization framework allows us to bypass the need for phase-accurate simulation, which would otherwise demand sub-millimeter geometric accuracy and prohibitively complex electromagnetic modeling. Hence, *TwinFocus* only needs to replicate the high-level structure of the reference object rather than its exact geometry or scattering properties.

Step 1: Visual 3D reconstruction. We begin by identifying a reference object that is visually distinct and in line-of-sight (LoS). While this object must be visible to the stereo camera, the actual imaging target may be in non-line-of-sight (NLoS). As shown in Fig. 6, we first use a stereo camera to capture RGB image pairs of the scene. Then, we apply the Segment Anything Model (SAM) 2 [27], a general-purpose vision foundation model, to extract a mask of the chosen reference object, without requiring manual annotation or training. We next use stereo disparity to compute per-pixel depth and reconstruct a 3D point cloud of the masked object.

Step 2: Signal simulation using point-scatter model. Let $\mathbf{p}_c^i = (x_c^i, y_c^i, 0)$ denote the i-th virtual antenna (phase center), and $\{\mathbf{r}_j\}$ be the 3D point cloud. For each antenna–point pair, we compute their distance as $R_{ij} = \|\mathbf{r}_j - \mathbf{p}_c^i\|_2$ Then, the simulated radar return at antenna i is the coherent sum of backscatter from every point in the extracted 3D point cloud:

$$\hat{s}_i(k) = \sum_j \alpha_j e^{-j2kR_{ij}} \tag{10}$$

where α_j is the reflectivity of point j (default to 1). After obtaining the simulated baseband signal, we apply back-projection to generate a simulated template image to serve as the ground truth for phase optimization. We pass both simulated and real images through a sigmoid function to smooth amplitude values while preserving object structure.

Step 3: Radar-Camera Alignment. Since SSIM is sensitive to spatial misalignment, we align the simulated SAR image of the reference object with its corresponding region in the real SAR image, as the former is generated in the camera frame while the latter is captured in the radar frame. We estimate the reference object's 3D position using a SAM2 mask and stereo point cloud, then map it to the radar frame using a one-time calibration of the translational offset between simulated and real SAR images. Since the radar and camera are co-located and remain fixed, this offset is reusable across scans. We further compute SSIM only on the cropped object region, reducing sensitivity to residual shifts and scale differences.

Design choice: Why point scatter model instead of ray tracing with full mesh? Ray tracing can capture complex multipath interactions but is computationally prohibitive for real-time handheld use, especially when simulating tens of thousands of antenna elements in our case. In contrast, the point-scatterer model, though less precise, is orders of magnitude faster and sufficient for amplitude simulation. While the simulated image may differ in fine-grained details due to noise or visual mismatch, the SSIM loss remains robust by focusing on structural similarity. Empirically, we find that even imperfectly simulated templates provide reliable guidance for phase correction.

5 Implementation

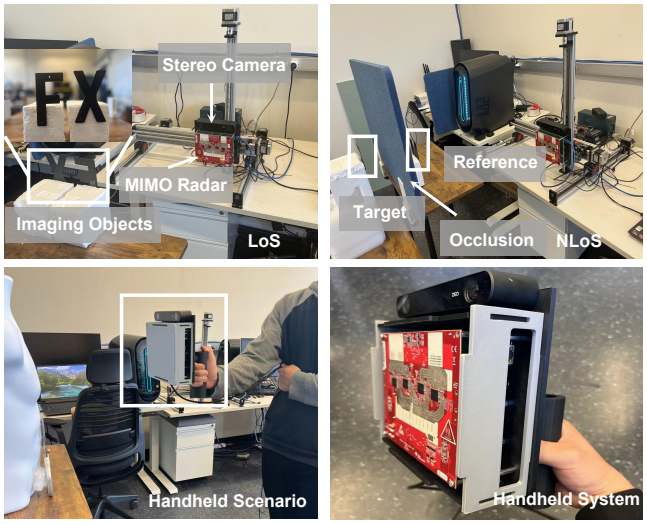

Figure 7: *TwinFocus* setup for ground-truth data collection in both LoS (left) and NLoS (right) scenarios.

Hardware Setup. As shown in Fig. 7, our handheld SAR imaging system is built around the Texas Instruments (TI) MMWCAS four-chip cascaded mmWave radar [35], paired with an MMWDSP evaluation board for raw ADC data acquisition. The radar operates with 12 TX and 16 RX, forming a linear virtual array of 86 non-overlapping elements. The chirp parameters are configured as follows: a sampling rate of 8 Msps, ADC samples of 256, chirp duration of 40 μs, and frequency slope of 120 MHz/μs, resulting in a total bandwidth of 4.8 GHz. The radar uses a 100 ms frame interval, with each frame containing 12 chirps from 12 antennas.

We obtain a coarse estimate of the scanning trajectory and capture the image for simulating reference responses by incorporating

a ZED 2i stereo camera [34] and co-locating it with the radar. The ZED 2i stereo camera records depth and 6-DoF pose data at 60 Hz. To collect ground-truth SAR images, we employ a 3D high-precision motion stage equipped with linear rails along the x, y, and z axes. We also build a handheld setup as shown in Fig. 7 (bottom) to collect real handheld trajectories and validate the imaging performance.

Data Collection: Our dataset includes over 18 objects and 38 reference–target metal-letter pairs captured under both LoS and NLoS conditions, with a 10° angular separation unless otherwise specified. The synthetic planar array measures about 480 mm × 260 mm, with antenna spacing of 1 mm along the x-axis and 2 mm along the y-axis. To evaluate tracking errors, we inject Gaussian motion errors ($\mu = 0, \sigma = 0.5mm$), matching the error profile of the ZED 2i camera, into GT data. We also collect real handheld trajectories using the portable setup shown in Fig. 7 (bottom) to examine the effect of 3D handheld scanning. Users follow a zig-zag handheld scanning trajectory over an area of about 0.5 × 0.5 m at a typical speed of 0.2 m/s.

Implementation Detail: We implement *TwinFocus* using PyTorch. The optimization employs the Adam optimizer with a learning rate of 5×10^{-2} for 300 epochs, combined with a CosineAnnealingLR learning rate scheduler. The image size is set to 100 × 100. We initialize the phase error vector to zero and iteratively update it to approximate the ground truth. Since our MIMO radar has 86 array elements, we assume that all elements share the same motion error at each time step, a condition that has been experimentally verified in [16]. Consequently, the dimensionality of the phase error vector to be optimized equals the number of distinct radar positions, which in our implementation is 780.

6 Evaluation

6.1 Baselines and Metrics

Baselines. We evaluate both variants of *TwinFocus*: ***TwinFocus*-Phys.**, which uses a physical twin, and ***TwinFocus*-Dig.**, which uses a digital twin. We compare them against three baselines: **Without Focus**, which performs no phase correction; **Sharpness Optimization** [8, 22, 46], a classical autofocus method that maximizes image sharpness; and **IFNet** [17], a state-of-the-art deep learning approach for handheld mmWave imaging. We do not compare with [16] because it is limited to corner reflectors and 1D scanning, making it incompatible with our handheld 2D setting and arbitrary reference objects.

Evaluation Metrics. We use *SSIM* as the primary image-quality metric, following prior work [17]. We also report *semantic similarity*, computed as the cosine similarity between ResNet features extracted from the reconstructed and ground-truth RGB SAR images.

6.2 Comparison with Baseline

We evaluate the effectiveness of our method by comparing *TwinFocus* against existing baselines under controlled conditions. As shown in Fig. 7 (upper left), we use six metallic letters (A, C, F, R, X, Z), each 5 inches tall, and collect all pairwise combinations so that each letter serves as both reference and target, yielding 30 unique pairs. For each pair, we inject Gaussian motion errors five times, resulting in 150 data samples. The radar–target distance is fixed at 60 cm. We refer to this dataset as the *TwinFocus* baseline dataset, which is also used in subsequent evaluations.

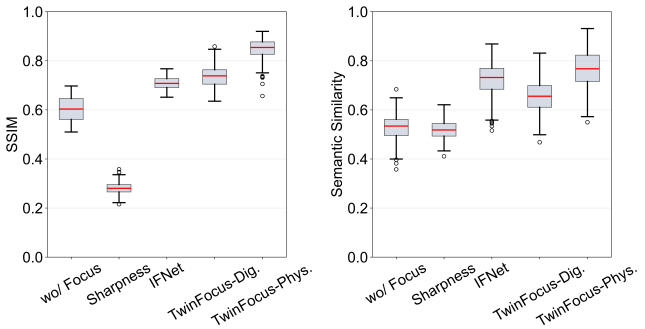

Figure 8: Quantitative comparison with baselines. Although IFNet [17] achieves reasonably good quantitative scores, its qualitative results reveal limited generalization ability.

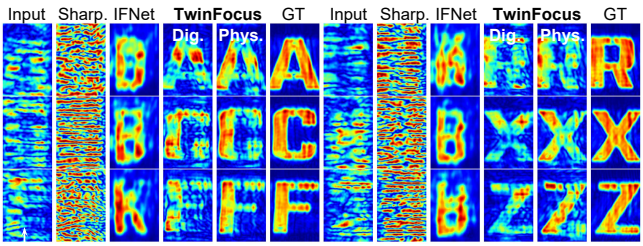

Figure 9: Qualitative comparison with baselines.

Results: Fig. 8 reports the mean SSIM and semantic similarity of all methods. *TwinFocus*-Dig. and *TwinFocus*-Phys. achieve average SSIM values of 0.74 and 0.85, corresponding to 23.3% and 41.7% improvements over the no-correction baseline, respectively. The same trend is observed for semantic similarity. In contrast, the model-driven Sharpness Optimization method [22] further degrades image quality, likely because it is designed for 2D remote sensing settings with milder motion perturbations and fewer antennas. Although IFNet [17] achieves reasonably good quantitative scores after fine-tuning on our dataset, its qualitative results reveal limited generalization. As shown in Fig. 9, IFNet reconstructs unseen letters as training-set letters such as B and K, whereas *TwinFocus* substantially improves image quality and produces clearly recognizable reconstructions. Based on these results, we compare primarily against the no-correction baseline in the following sections.

6.3 Validation with Handheld Motion

To evaluate *TwinFocus* under realistic handheld usage, we built a freehand imaging setup, as shown in Fig. 7 (bottom). A user was instructed to move the radar in different zig-zag patterns while a ZED 2i stereo camera recorded various motion trajectories (Fig. 10). We then reproduced these captured 3D trajectories using a motion stage to obtain ground-truth data for performance evaluation. *TwinFocus* was applied to estimate and correct motion errors based on

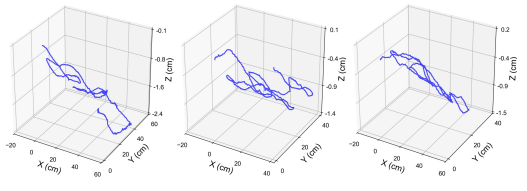

Figure 10: Motion trajectories of various 3D handheld scanning.

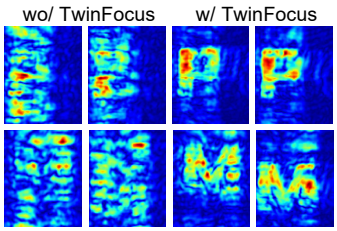

Figure 11: Imaging results of radar with handheld motion.

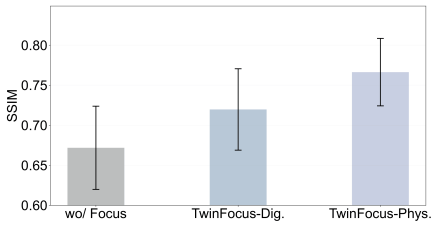

Figure 12: Quantitative results of radar with handheld motions.

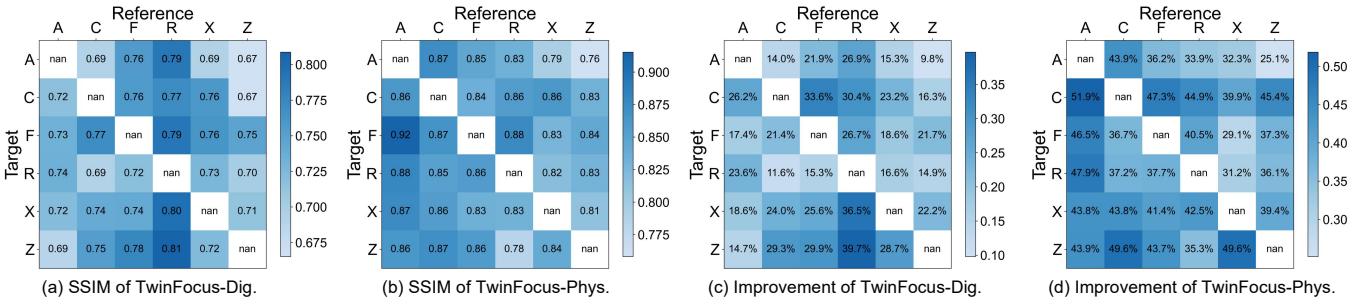

Figure 13: *TwinFocus*'s generalization across different pairs of reference and target objects.

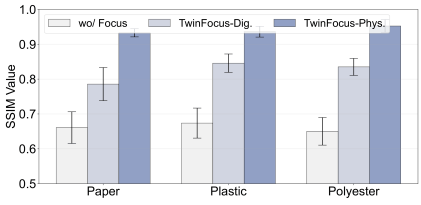

Figure 14: Robustness to different NLoS occlusions.

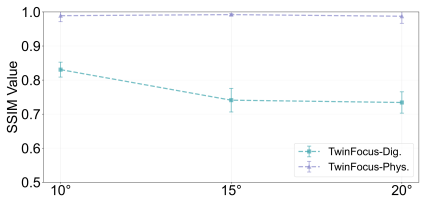

Figure 15: Robustness to reference-target angle separations.

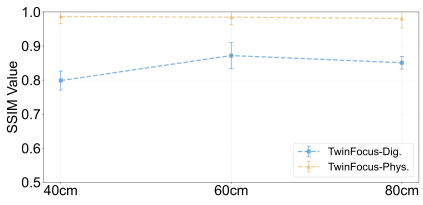

Figure 16: Robustness to object-radar distances.

the trajectory distribution measured by the ZED camera. In this experiment, the letter K was used as the reference, and letters P and M served as target objects. As shown in Fig. 11, the qualitative results before and after applying *TwinFocus*-Dig. and *TwinFocus*-Phys. indicate that our method produces sharper, more focused images with clearly recognizable shapes. The quantitative results in Fig. 12 further show that *TwinFocus* achieves significant improvement in SSIM for radar signals collected under handheld motion.

6.4 *TwinFocus* Robustness

This subsection evaluates the robustness of *TwinFocus* across various challenging real-world scenarios.

Different Object Shapes: Quantifying shape complexity and exhaustively evaluating all geometries is difficult. We use letters here because they provide diverse geometric primitives, such as circular, rectangular, and irregular structures, in a controlled setting. Using a comprehensive dataset of 30 reference–target letter pairs and 150 samples with Gaussian motion errors, we group results by object shape and analyze the SSIM of the autofocused images. This experiment shows that *TwinFocus* generalizes across diverse shapes rather than overfitting to specific geometries.

Fig. 13(a) and (b) show the SSIM quality generated by both *TwinFocus*-Dig. and *TwinFocus*-Phys. Fig. 13(c) and (d) show the percentage improvement in SSIM from the results obtained without phase correction. The median improvement observed across the object pairs is 22.04% and 40.93% by *TwinFocus*-Dig. and *TwinFocus*-Phys.

Non-Line-of-Sight (NLoS): A fundamental benefit of RF-imaging systems is that wireless signals can pass through materials for NLoS imaging. Real-world applications often require imaging through occluding materials, such as in security screening, non-destructive testing, and through-wall imaging. Hence, we evaluate *TwinFocus* in NLoS scenarios to demonstrate its practical utility when direct line-of-sight to targets is unavailable. As shown in Fig.7 (upper right), the reference target remains in LoS while the main target is fully occluded. We test three occlusion materials, including polyester, paper, and plastic, with 4-inch metal letters placed 60 cm (target) and 76 cm (reference) from the radar. For each material, we use four

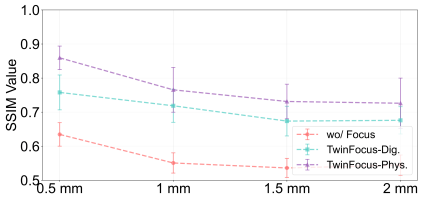

Figure 17: Robustness to different tracking errors.

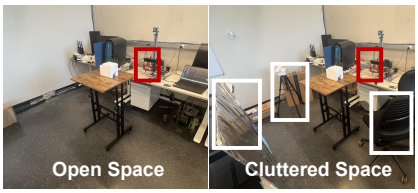

Figure 18: Evaluation setups across environments.

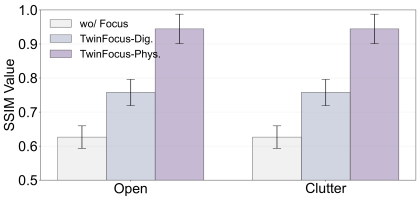

Figure 19: Robustness to different environments.

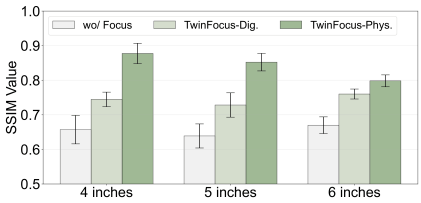

Figure 20: Robustness to different object sizes.

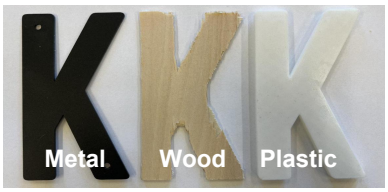

Figure 21: Evaluation targets across materials.

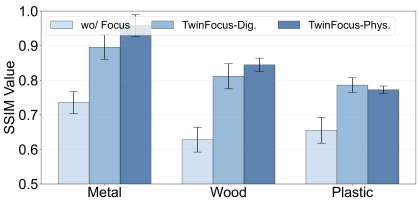

Figure 22: Robustness to different materials of the target.

letter combinations (M, K, P) and apply positional perturbations, generating 120 defocused SAR samples.

Fig. 14 shows SSIM results under each occlusion. Across all materials, *TwinFocus* consistently improves image quality for both the physics-based and digital-twin autofocus methods. Even with the digital-twin approach, average SSIM improvements reach 27.7%, 18.9%, and 25.5% for polyester, paper, and plastic, respectively. This is because our signal model does not rely on any assumptions about the occluding object; as long as the reference remains visible, target phase offsets can still be accurately corrected.

Different Ranges and Angles. We evaluate the impact of object–radar distance (40, 60, and 80 cm) and reference–target angular separation (10°, 15°, and 20°) to characterize the performance boundary of *TwinFocus*. We use letter K as the reference and letter P as the target, with 40 samples under Gaussian motion errors. Fig. 15 shows that, with the physical twin, performance remains stable across all angles. In contrast, the digital twin degrades at larger separations, since larger view differences make the approximation more challenging and require a more accurate template. As shown in Fig. 16, *TwinFocus* maintains strong performance across all tested distances, with slightly better results at 60 cm.

Different Tracking Errors. Different mobile tracking systems may introduce varying levels of positional error. We validate *TwinFocus*'s tolerance to tracking inaccuracies, using the *TwinFocus* baseline dataset and systematically add simulated Gaussian tracking errors with standard deviations ranging from 0.5 mm to 2 mm to the camera trajectory. This controlled experiment allows us to isolate the impact of tracking accuracy on correction performance and identify the maximum correctable error threshold. Fig. 17 shows how *TwinFocus* outperforms the baseline image even at relatively larger motion errors (Note that the standard deviation of the motion error in the ZED 2i stereo imaging system is 0.5 mm, as shown in Fig. 3.)

Different Multipath Environments: Environmental conditions, particularly multipath reflections, can significantly affect radar signal phase. This experiment evaluates *TwinFocus*'s robustness under different environments and examines the transferability of reference data across them. We use a pair of letters, "U" (reference) and "G" (target), and collect measurements in both open and cluttered settings (Fig. 18). Motion errors are introduced to generate 60 SAR samples. We then use the reference collected in one environment as the template for phase error estimation in the other.

Results in Fig. 19 shows comparable SSIM improvements in both environments, indicating that *TwinFocus*'s performance is unaffected by surrounding conditions. This robustness stems from our autofocus design, which optimizes primarily around a small region of interest, the reference object, rather than the broader environment.

Different Object Sizes: Real-world scenarios involve objects of varying sizes, and correction performance may depend on the relative sizes of the target and reference objects. This experiment evaluates size-related effects and validates the scalability of our approach across different object dimensions. We use a 3-inch letter "N" as the reference and letters "B", "X", and "H", measuring 4 in., 5 in., and 6 in. respectively, as the targets. We then introduce Gaussian positional errors with a standard deviation of 0.5 mm and generate 60 SAR samples containing these errors.

Fig. 20 shows *TwinFocus*'s performance across the differently sized objects. We see a general trend of reduction in imaging quality as the objects become larger. This is because larger objects span a larger angle-of-arrival for the radar, and the phase error similarity degrades as the angular distance increases (as predicted in Sec. 3). Despite this, *TwinFocus* continues to improve the quality of image

Image Size	50 × 50	75 × 75	100 × 100
Optimization Time (s)	4.30	8.85	15.05
Array Size	200 × 86	400 × 86	600 × 86
Optimization Time (s)	4.54	8.48	12.51
Digital Twin Operation			Time (s)
Visual Segmentation			0.48
Point Cloud Extraction			0.07
Array Signal Simulation			28.35

Table 1: Computational cost of *TwinFocus* for image phase optimization and digital twin operations.

by 13.4% and 17.9% using digital and physical twins, respectively, for the largest target.

Different Object Material: We next evaluate the performance of *TwinFocus* across different object materials to highlight the practical utility of *TwinFocus* in real-world scenarios where the properties of unknown objects may vary widely. In this experiment, the reference is a metal letter "M," while the target is a letter "K" made of various materials, including metal, wood, and plastic, as shown in Fig. 21. We run the experiments using 20 samples with Gaussian motion errors for each material.

Fig. 22 shows the SSIM achieved by both *TwinFocus*-Dig. and *TwinFocus*-Phys., demonstrating that both the simulated and physical templates yield notable improvements across different materials. We primarily use metal objects as references because their simulated responses more closely match real measurements due to stronger reflectivity. However, our method generalizes well across different reference–target material combinations, with only a slight degradation when the target material differs significantly.

Computational Efficiency: We evaluate the runtime of *TwinFocus* on an NVIDIA RTX 4090 GPU. As shown in Table 1, to examine scalability, we vary both image resolution and array size, fixing the array at 780 × 86 for image-size experiments and the image at 100 × 100 for array-size experiments. The reported runtime also assumes SAR backprojection is performed after data collection. In practice, backprojection can be computed on-the-fly during acquisition, further reducing the effective post-scan latency. We also report the digital-twin overhead separately in Table 1. Since this cost is incurred only once per reference scan, it can be amortized over multiple target scans in the same environment and partially hidden by running in the background during scanning.

6.5 Security Screening Case Study

We present a case study to demonstrate the practical potential of *TwinFocus* for portable security screening. We use real-world targets and generic-shaped references that differ significantly in geometry, size, and structural complexity. As shown in Fig. 23, we simulate this scenario using a mannequin[2] wearing a cotton T-shirt, with concealed scissors beneath the shirt and a LoS knife as the targets. We use generic-shaped reference objects, including a metal card and a circular reflector, and incorporate typical motion errors from the ZED 2i stereo tracking system. As shown in Fig. 23, without phase correction, the concealed objects are difficult to distinguish, whereas *TwinFocus* substantially improves image quality. For the scissors, SSIM increases from 0.65 to 0.94. These results suggest that *TwinFocus* has the potential to support handheld high-resolution imaging in real-world tasks.

[2]Although human screening may involve other body motion errors, similar applications, such as imaging objects inside containers or packages, can still benefit from the same portable workflow.

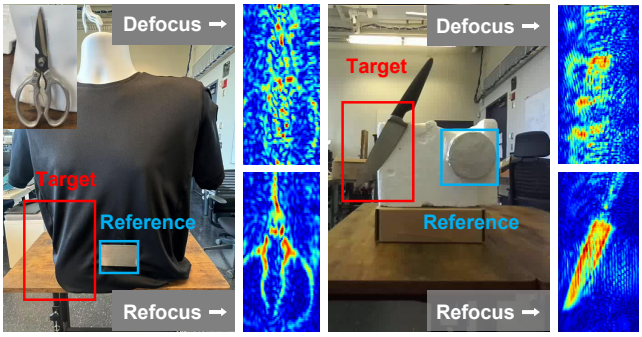

Figure 23: Security screening case study with real-world targets and generic-shaped references.

7 Discussion and Limitations

VIO Tracking Accuracy. *TwinFocus* currently assumes sub-wavelength pose accuracy to avoid phase ambiguity during autofocus. For mmWave SAR, phase errors become ambiguous when motion exceeds roughly half the wavelength, so millimeter-level tracking is sufficient in our setup. This assumption is supported by the ZED 2i specification (±1 mm under favorable conditions) [1] and by our empirical measurements, where most observed range-direction errors remain within the required regime. Nevertheless, this assumption may not hold for lower-quality VIO systems or challenging visual conditions. Future work can explore more robust phase optimization and motion compensation methods that can tolerate larger tracking errors.

Reference Object Constraints. *TwinFocus* does not eliminate the need for a reference, but relaxes its geometric constraints. It replaces specialized calibration hardware with an in-scene reference, making handheld imaging more practical, flexible, and cost-effective. Additionally, *TwinFocus* performs best when the reference and target have similar spatial extent and viewing geometry. As object size increases, correction quality gradually declines because different parts of a larger target experience less similar phase errors relative to the reference. Reference orientation also matters: in our experiments, objects are approximately perpendicular to the radar array, which is common in mmWave imaging due to specular reflection. Large tilts or viewpoint changes can weaken or remove returns, producing incomplete reconstructions. This limitation arises from the sensing physics rather than the autofocus algorithm itself.

What Makes a Good Reference. A good reference object should be easy to segment, geometrically stable, and imaged reliably by the radar. In our experiments, shape has a limited impact as long as the object has clear boundaries that support robust reconstruction and alignment. Homogeneous materials such as metal, wood, and

Project	Autofocus	2D Scan	Arb. Ref.	Training -free
MilliPCD [2]	✘	✔	–	✘
SquiggleMilli [28]	✘	✔	–	✘
TIM [50]	✘	✔	–	✘
PSF [16]	✔	✘	✘	✔
IFNet [17]	✔	✘	–	✘
TwinFocus	✔	✔	✔	✔

Table 2: Comparison of *TwinFocus* with prior handheld mmWave imaging work. *TwinFocus* supports autofocus, 2D scanning, and arbitrary references without machine learning model training.

plastic work well because *TwinFocus* mainly relies on phase consistency rather than exact amplitude matching. Composite objects are more challenging, since strong reflectors can dominate weaker components and reduce the match between the simulated reference and measured SAR response. More accurate material-aware simulation may help address this limitation. Cross-scenario reference reuse can also work if the reference geometry and imaging quality remain consistent.

8 Related Work

Traditional mmWave SAR implementations [10, 40, 43] rely on bulky, high-precision motion stages to maintain coherence. To enable mobile and ubiquitous applications, recent research explores handheld mmWave SAR, focusing on overcoming sub-mm motion errors that degrade imaging quality.

Motion Compensation: Motion compensation aims to accurately measure the radar's trajectory and directly correct the motion errors to preserve coherence [4]. However, at 77 GHz, even sub-mm errors can cause severe phase misalignment and image defocusing. Initial works [48, 51] corrected motion errors using precise motion capture systems [23] with multiple cameras and markers. Although effective, these systems are expensive and impractical for real-world deployment. To reduce cost and improve usability, subsequent works [30, 49, 50] incorporate mobile tracking solutions based on VIO. However, they fail to meet the stringent tracking requirements, causing the results to suffer from deblurring and artifacts. Other efforts [2, 28] also adopt this VIO setup but focus on challenges like sparse sampling and specular reflections, rather than correcting for tracking errors.

TwinFocus also employs a cost-effective stereo-IMU tracking system for coarse trajectory estimation. However, we estimate and correct phase errors directly from radar data, eliminating the need for precise tracking and enabling effective phase correction with minimal hardware overhead.

Signal Processing-based Autofocus: Signal processing-based autofocus techniques estimate phase errors directly from the radar signals when tracking is unreliable. Classic approaches in remote sensing, including phase gradient autofocus [7, 37], map-drift autofocus [3], and metric-based optimization [8, 22, 38, 44, 46], have been widely studied. However, they are primarily designed for 2D far-field imaging with lower-frequency radar and perform poorly when extended to near-field, high-frequency 3D mmWave SAR. Recent methods [16] attempt to address this by introducing known point targets (i.e., corner reflectors) as references, but such targets are rarely available in practice and are typically limited to 1D scanning.

TwinFocus differs in greatly relaxing the constraints on the reference geometry and generalizes to 2D scanning trajectories. By reconstructing the SAR response of the reference object, we perform image-level autofocus through gradient-based phase optimization that enables robust phase correction using various references.

Machine Learning-based Autofocus: Machine learning-based methods take defocused images as input and use well-focused images as ground truth for training [19, 24, 39]. Researchers have designed various network architectures, including convolutional neural networks (CNNs) [15, 36], generative adversarial networks (GANs) [15, 31], and deep unfolding networks [17, 18], for recovering SAR images with handheld phase errors. Despite promising results on certain datasets, these approaches require large-scale training data, and models trained on specific targets often fail to adapt to unseen objects or environments.

In contrast, *TwinFocus* requires no model training. It estimates phase errors directly from a single SAR measurement of a reference object and transfers them to correct defocus across the scene. This makes *TwinFocus* fully interpretable and generalizable to different objects and environments.

9 Conclusion

This paper presented *TwinFocus*, a twin reference-guided differentiable autofocus framework designed to address the critical challenge of phase error correction in handheld mmWave SAR imaging. Unlike prior work that either relies on precise mechanical scanning, large training datasets, or simplified point target assumptions, *TwinFocus* leverages reference objects with relaxed geometrical constraints and minimizes the discrepancy between real and simulated SAR images to achieve high-fidelity image focusing. By aligning images in the amplitude domain, *TwinFocus* enables robust phase optimization even with unknown reference geometries while requiring only lightweight simulation support. Extensive experiments across various object shapes, sizes, materials, environments, motion errors, and both LoS and NLoS scenarios demonstrate that *TwinFocus* consistently improves SAR image quality. Looking forward, we envision *TwinFocus* as a critical step toward enabling compact, mobile, and high-resolution mmWave SAR systems suitable for ubiquitous sensing scenarios.

References

[1] [n. d.]. ZED Doc. https://support.stereolabs.com/hc/en-us/articles/1500008652841-What-is-the-expected-accuracy-of-the-position-x-y-z-for-the-positional-tracking-functionality-of-the-ZED-cameras-i-e-0-01-0-1-0-1-in-meters-for-x-y-z.

[2] Pingping Cai and Sanjib Sur. 2023. MilliPCD: Beyond Traditional Vision Indoor Point Cloud Generation via Handheld Millimeter-Wave Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 160 (Jan. 2023), 24 pages. https://doi.org/10.1145/3569497

[3] T.M. Calloway and G.W. Donohoe. 1994. Subaperture autofocus for synthetic aperture radar. *IEEE Trans. Aerospace Electron. Systems* 30, 2 (1994), 617–621. https://doi.org/10.1109/7.272285

[4] Jianlai Chen, Mengdao Xing, Hanwen Yu, Buge Liang, Jian Peng, and Guang-Cai Sun. 2022. Motion Compensation/Autofocus in Airborne Synthetic Aperture Radar: A Review. *IEEE Geoscience and Remote Sensing Magazine* 10, 1 (2022), 185–206. https://doi.org/10.1109/MGRS.2021.3113982

[5] Laura Dodds, Tara Boroushaki, Kaichen Zhou, and Fadel Adib. 2025. *Non-Line-of-Sight 3D Object Reconstruction via mmWave Surface Normal Estimation.* Association for Computing Machinery, New York, NY, USA, 445–458. https://doi.org/10.1145/3711875.3729138

[6] Laura Dodds, Hailan Shanbhag, Junfeng Guan, Saurabh Gupta, and Haitham Hassanieh. 2024. Around the Corner mmWave Imaging in Practical Environments. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking* (Washington D.C., DC, USA) *(ACM MobiCom '24)*. Association for Computing Machinery, New York, NY, USA, 953–967. https://doi.org/10.1145/3636534.3690671

[7] Aaron Evers and Julie Ann Jackson. 2019. A Generalized Phase Gradient Autofocus Algorithm. *IEEE Transactions on Computational Imaging* 5, 4 (2019), 606–619. https://doi.org/10.1109/TCI.2019.2899453

[8] J. R. Fienup. 2000. Synthetic-aperture radar autofocus by maximizing sharpness. *Opt. Lett.* 25, 4 (Feb 2000), 221–223. https://doi.org/10.1364/OL.25.000221

[9] Akshay Gadre, Deepak Vasisht, Nikunj Raghuvanshi, Bodhi Priyantha, Manikanta Kotaru, Swarun Kumar, and Ranveer Chandra. 2022. MiLTOn: Sensing product integrity without opening the box using non-invasive acoustic vibrometry. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 390–402.

[10] Xiangyu Gao, Sumit Roy, and Guanbin Xing. 2021. MIMO-SAR: A hierarchical high-resolution imaging algorithm for mmWave FMCW radar in autonomous driving. *IEEE Transactions on Vehicular Technology* 70, 8 (2021), 7322–7334.

[11] Yuan Gao, Mohammad Tayeb Ghasr, and Reza Zoughi. 2020. Effects of and Compensation for Translational Position Error in Microwave Synthetic Aperture Radar Imaging Systems. *IEEE Transactions on Instrumentation and Measurement* 69, 4 (2020), 1205–1212. https://doi.org/10.1109/TIM.2019.2910340

[12] Ruixu Geng, Yadong Li, Dongheng Zhang, Jincheng Wu, Yating Gao, Yang Hu, and Yan Chen. 2024. DREAM-PCD: Deep Reconstruction and Enhancement of mmWave Radar Pointcloud. *IEEE Transactions on Image Processing* 33 (2024), 6774–6789. https://doi.org/10.1109/TIP.2024.3512356

[13] Ruixu Geng, Dongheng Zhang, Yadong Li, Zhi Wu, Jiamu Li, Qi Chen, Yang Hu, and Yan Chen. 2025. Attacking mmWave Imaging With Neural Meta-Material Rendering. *IEEE Transactions on Information Forensics and Security* 20 (2025), 5340–5355. https://doi.org/10.1109/TIFS.2025.3571667

[14] Mike Golden, Peter Pietra, Hugo Teufel III, and Chief Privacy Officer. 2008. TSA Whole Body Imaging. (2008).

[15] Jaime Laviada, Guillermo Álvarez Narciandi, and Fernando Las-Heras. 2022. Artifact Mitigation for High-Resolution Near-Field SAR Images by Means of Conditional Generative Adversarial Networks. *IEEE Transactions on Instrumentation and Measurement* 71 (2022), 1–11. https://doi.org/10.1109/TIM.2022.3200107

[16] Yadong Li, Dongheng Zhang, Ruixu Geng, Zhi Lu, Zhi Wu, Yang Hu, Qibin Sun, and Yan Chen. 2024. A high-resolution handheld millimeter-wave imaging system with phase error estimation and compensation. *Communications Engineering* 3, 1 (05 Jan 2024), 4. https://doi.org/10.1038/s44172-023-00156-2

[17] Yadong Li, Dongheng Zhang, Ruixu Geng, Jincheng Wu, Yang Hu, Qibin Sun, and Yan Chen. 2024. IFNet: Deep Imaging and Focusing for Handheld SAR With Millimeter-Wave Signals. *IEEE Transactions on Mobile Computing* (2024), 1–16. https://doi.org/10.1109/TMC.2024.3489641

[18] Yadong Li, Dongheng Zhang, Ruixu Geng, Jincheng Wu, Yang Hu, Qibin Sun, and Yan Chen. 2024. IFNet: Imaging and Focusing Network for handheld mmWave Devices. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2415–2419. https://doi.org/10.1109/ICASSP48485.2024.10447461

[19] Zhi Liu, Shuyuan Yang, Quanwei Gao, Zhixi Feng, Min Wang, and Licheng Jiao. 2022. AFnet and PAFnet: Fast and Accurate SAR Autofocus Based on Deep Learning. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–13. https://doi.org/10.1109/TGRS.2022.3217063

[20] J.M. Lopez-Sanchez and J. Fortuny-Guasch. 2000. 3-D radar imaging using range migration techniques. *IEEE Transactions on Antennas and Propagation* 48, 5 (2000), 728–737. https://doi.org/10.1109/8.855491

[21] Nishant Mehrotra, Divyanshu Pandey, Akarsh Prabhakara, Yawen Liu, Swarun Kumar, and Ashutosh Sabharwal. 2024. Hydra: Exploiting Multi-Bounce Scattering for Beyond-Field-of-View mmWave Radar. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking* (Washington D.C., DC, USA) *(ACM MobiCom '24)*. Association for Computing Machinery, New York, NY, USA, 1545–1559. https://doi.org/10.1145/3636534.3690710

[22] Robert L. Morrison, Minh N. Do, and David C. Munson. 2007. SAR Image Autofocus By Sharpness Optimization: A Theoretical Study. *IEEE Transactions on Image Processing* 16, 9 (2007), 2309–2321. https://doi.org/10.1109/TIP.2007.903252

[23] NaturalPoint, Inc. 2025. OptiTrack Motion Capture Systems. https://optitrack.com/. Accessed: 2025-05-29.

[24] Wei Pu. 2022. SAE-Net: A Deep Neural Network for SAR Autofocus. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–14. https://doi.org/10.1109/TGRS.2021.3139914

[25] Tong Qin, Shaozu Cao, Jie Pan, and Shaojie Shen. 2019. A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors. arXiv:1901.03642 [cs.CV] https://arxiv.org/abs/1901.03642

[26] Tong Qin, Peiliang Li, and Shaojie Shen. 2018. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics* 34, 4 (2018), 1004–1020. https://doi.org/10.1109/TRO.2018.2853729

[27] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. 2024. SAM 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714* (2024). https://arxiv.org/abs/2408.00714

[28] Hem Regmi, Moh Sabbir Saadat, Sanjib Sur, and Srihari Nelakuditi. 2021. SquiggleMilli: Approximating SAR Imaging on Mobile Millimeter-Wave Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 125 (Sept. 2021), 26 pages. https://doi.org/10.1145/3478113

[29] Angel Ribalta. 2011. Time-Domain Reconstruction Algorithms for FMCW-SAR. *IEEE Geoscience and Remote Sensing Letters* 8, 3 (2011), 396–400. https://doi.org/10.1109/LGRS.2010.2078486

[30] Moh Sabbir Saadat, Sanjib Sur, Srihari Nelakuditi, and Parmesh Ramanathan. 2020. MilliCam: Hand-held Millimeter-Wave Imaging. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)*. 1–9. https://doi.org/10.1109/ICCCN49398.2020.9209710

[31] Jacqueline M Schellberg, Hem Regmi, and Sanjib Sur. 2023. mmSight: Towards Robust Millimeter-Wave Imaging on Handheld Devices. In *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 117–126. https://doi.org/10.1109/WoWMoM57956.2023.00026

[32] Hailan Shanbhag, Sohrab Madani, Akhil Isanaka, Deepak Nair, Saurabh Gupta, and Haitham Hassanieh. 2023. Contactless material identification with millimeter wave vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. 475–488.

[33] M. Slaney, A.C. Kak, and L.E. Larsen. 1984. Limitations of Imaging with First-Order Diffraction Tomography. *IEEE Transactions on Microwave Theory and Techniques* 32, 8 (1984), 860–874. https://doi.org/10.1109/TMTT.1984.1132783

[34] Stereolabs Inc. 2025. ZED 2: Versatile stereo camera for spatial perception. https://www.stereolabs.com/products/zed-2. Accessed: 30 June 2025.

[35] Texas Instruments. 2025. MMWCAS-RF-EVM: mmWave cascade imaging radar RF evaluation module. https://www.ti.com/tool/MMWCAS-RF-EVM. Accessed: 30 June 2025.

[36] Christos Vasileiou, Josiah Smith, Shiva Thiagarajan, Matthew Nigh, Yiorgos Makris, and Murat Torlak. 2022. Efficient CNN-Based Super Resolution Algorithms for Mmwave Mobile Radar Imaging. In *2022 IEEE International Conference on Image Processing (ICIP)*. 3803–3807. https://doi.org/10.1109/ICIP46576.2022.9897190

[37] D.E. Wahl, P.H. Eichel, D.C. Ghiglia, and C.V. Jakowatz. 1994. Phase gradient autofocus-a robust tool for high resolution SAR phase correction. *IEEE Trans. Aerospace Electron. Systems* 30, 3 (1994), 827–835. https://doi.org/10.1109/7.303752

[38] J. Wang and X. Liu. 2006. SAR Minimum-Entropy Autofocus Using an Adaptive-Order Polynomial Model. *IEEE Geoscience and Remote Sensing Letters* 3, 4 (2006), 512–516. https://doi.org/10.1109/LGRS.2006.878446

[39] Mou Wang, Shunjun Wei, Zichen Zhou, Jun Shi, Xiaoling Zhang, and Yongxin Guo. 2022. 3-D SAR Autofocusing With Learned Sparsity. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–18. https://doi.org/10.1109/TGRS.2022.3210547

[40] Claire M Watts, Patrick Lancaster, Andreas Pedross-Engel, Joshua R Smith, and Matthew S Reynolds. 2016. 2D and 3D millimeter-wave synthetic aperture radar imaging on a PR2 platform. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4304–4310.

[41] Shunjun Wei, Jinshan Wei, Xinyuan Liu, Mou Wang, Shan Liu, Fan Fan, Xiaoling Zhang, Jun Shi, and Guolong Cui. 2021. Nonline-of-sight 3-D imaging using millimeter-wave radar. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), 1–18.

[42] Muhammet Emin Yanik, Dan Wang, and Murat Torlak. 2020. Development and Demonstration of MIMO-SAR mmWave Imaging Testbeds. *IEEE Access* 8 (2020), 126019–126038. https://doi.org/10.1109/ACCESS.2020.3007877

[43] Muhammet Emin Yanik, Dan Wang, and Murat Torlak. 2020. Development and demonstration of MIMO-SAR mmWave imaging testbeds. *IEEE Access* 8 (2020), 126019–126038.

[44] T. Zeng, R. Wang, and F. Li. 2013. SAR Image Autofocus Utilizing Minimum-Entropy Criterion. *IEEE Geoscience and Remote Sensing Letters* 10, 6 (2013), 1552–1556. https://doi.org/10.1109/LGRS.2013.2261975

[45] Fusang Zhang, Jie Xiong, Zhaoxin Chang, Junqi Ma, and Daqing Zhang. 2022. Mobi2Sense: empowering wireless sensing with mobility. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking* (Sydney, NSW, Australia) *(MobiCom '22)*. Association for Computing Machinery, New York, NY, USA, 268–281. https://doi.org/10.1145/3495243.3560518

[46] Tao Zhang, Guisheng Liao, Yachao Li, Tong Gu, Tinghao Zhang, and Yongjun Liu. 2022. A Two-Stage Time-Domain Autofocus Method Based on Generalized Sharpness Metrics and AFBP. *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), 1–13. https://doi.org/10.1109/TGRS.2021.3068789

[47] Xiaoxuan Zhang, Jie Liang, Nan Wang, Tianying Chang, Qijia Guo, and Hong-Liang Cui. 2019. Broadband millimeter-wave imaging radar-based 3-D holographic reconstruction for nondestructive testing. *IEEE Transactions on Microwave Theory and Techniques* 68, 3 (2019), 1074–1085.

[48] Guillermo Álvarez Narciandi, Jaime Laviada, and Fernando Las-Heras. 2021. Freehand mm-Wave Imaging With a Compact MIMO Radar. *IEEE Transactions on Antennas and Propagation* 69, 2 (2021), 1224–1229. https://doi.org/10.1109/TAP.2020.3013745

[49] Guillermo Álvarez Narciandi, Jaime Laviada, and Fernando Las-Heras. 2021. Towards Turning Smartphones Into mmWave Scanners. *IEEE Access* 9 (2021), 45147–45154. https://doi.org/10.1109/ACCESS.2021.3067458

[50] Guillermo Álvarez Narciandi, Jaime Laviada, and Fernando Las-Heras. 2024. Short-Time Coherent–Incoherent Synthetic Aperture Radar Processing for a Handheld Imaging System. *IEEE Transactions on Instrumentation and Measurement* 73 (2024), 1–9. https://doi.org/10.1109/TIM.2024.3352706

[51] Guillermo Álvarez Narciandi, Miguel López-Portugués, Fernando Las-Heras, and Jaime Laviada. 2019. Freehand, Agile, and High-Resolution Imaging With Compact mm-Wave Radar. *IEEE Access* 7 (2019), 95516–95526. https://doi.org/10.1109/ACCESS.2019.2929522