Environment Reconstruction Based on Hybrid Aperture Sensing: Technologies and Applications

Jiajin Luo [1], Xiaohui Peng [1], Yang Yu [1], Ping Zhang [1], Baojian Zhou [1], Xiaoyan Bi [2], Yan Chen [2], Jianglei Ma [2], Peiying Zhu [2]
[1] Wireless Technology Lab
[2] Ottawa Advanced Wireless Technology Lab

Abstract

This paper proposes a hybrid aperture sensing (HAS) technology based on wireless networks, which utilizes the mobility of terminal nodes in the network to form virtual apertures and combines them with the real physical apertures of base stations to form hybrid aperture. By reusing the (uplink or downlink) air interface resources between base stations (real apertures) and terminals (virtual apertures), HAS significantly enhances the sensing resolution and enables the fusion of 3D point clouds from multiple network nodes, resulting in a digital twin environment model that corresponds to the physical world. The paper also explores the use of the reconstructed digital twin environment model for wireless communication channel prediction and clutter suppression in wireless sensing.

Keywords

hybrid aperture sensing (HAS), environment reconstruction, digital twin, sensing-assisted communication, sensing-assisted localization, 6G

Integrated sensing and communication (ISAC) as a highly promising feature for next-generation cellular networks, has been approved as one of the six major usage scenarios for IMT-2030 (6G) by ITU-R. In the scenario of ISAC, the cellular network infrastructures and devices not only serve as communication nodes for connectivity between people and things, but also as sensing nodes that help achieve environment imaging and reconstruction, high-precision target localization and tracking, gesture and motion recognition, and other sensing functions. Among them, environment reconstruction serves as an enabling technology for digital twins [1], and will play an important role in future scenarios such as digital reconstruction of a physical world, target localization, and autonomous driving. From the perspective of measurement, 3D environment reconstruction mainly has two technical routes: optical and microwave radar. Lidar-based 3D environment reconstruction solutions are known for their high precision, but are expensive, involve time-consuming deployment, and require a long data update cycle. Recently, the introduction of deep learning technology has made rapid progress in reconstructing 3D images based on 2D optical images using neural networks [2], which has lower equipment requirements but still needs to calibrate camera positions and separately training for different categories of objects. Its generalization capability and outdoor performance still need to be improved. Researchers in the radar imaging field have also conducted a lot of research on various imaging systems, such as vehicle-based and unmanned aerial vehicle-based 3D radar imaging systems [3–5], which are less vulnerable to ambient lighting and weather conditions compared to optical methods and have a larger coverage area. Traditional solutions for reconstructing environments are typically deployed and calibrated independently, which places high requirements on processing capabilities of devices. This results in high costs in terms of deployment, data update, and maintenance, failing to build ubiquitous digital twins in the future. To address this limitation, this paper proposes the hybrid aperture sensing (HAS) technology for next-generation wireless network, which fully utilizes the signal processing capabilities, data transmission capabilities, and positioning and synchronization capabilities of the network side, combines the terminal mobility, and does not require independent deployment and calibration, thus offering an on-demand, high-quality, and wide-coverage digital twin environment model. To further enhance the capabilities of wireless communication and localization, the impact of the environment on wireless channels [6] has also received more attention, and this paper further investigates channel prediction and clutter suppression by using the reconstructed environment model and evaluates its performance preliminarily. The following sections introduce the three parts of the article: sensing-based environment reconstruction, environment sensing-assisted communication, and environment sensing-assisted localization.

1 Sensing-based Environment Reconstruction

1.1 Hybrid Aperture Sensing

The sensing mode of a typical cellular network can be divided into two categories: base stations (BSs) only sensing and BS-terminal joint sensing. This paper proposes a HAS-based scheme, as shown in Figure 1, which adopts a joint bistatic sensing mode between BSs and mobile terminal. The BS side utilizes a real apertures antenna array to ensure the vertical resolution, while the mobile terminal, such as vehicles and UAVs, can form a virtual aperture to observe the environment, thereby achieving a larger field of view and higher horizontal resolution than BSs only sensing. The HAS-based solution also features reciprocity between uplink and downlink channels, allowing both uplink and downlink signals to be used for sensing. Considering that the BSs typically have stronger signal processing and transmission capabilities, this paper takes the uplink sensing, i.e., the BS receives and the mobile terminal transmits, as an example of a specific implementation of the HAS-based solution, and analyzes and demonstrates the performance of the cellular network in 3D environment reconstruction, as well as the use of the reconstructed environment model to predict communication channels and assist in target localization.

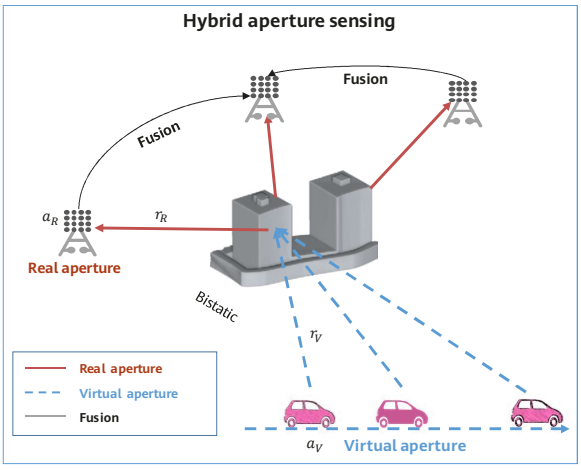


Figure 1 HAS-based solution

First, let's briefly review the basic principle of obtaining a scattered point cloud $C = \{ I(x, y, z)\}$ by using back projection (BP) algorithm. More algorithm details are available in [7, 8]. $I(x, y, z)$ represents the scattering intensity at a spatial target point $P(x, y, z)$, and can be calculated from (1).

$$I(x,y,z) = \iint s(r_H = r_V + r_R, a_V, a_R) e^{i\frac{2\pi}{\lambda} r_H} da_V da_R \quad (1)$$

where $s(r, a_V, a_R)$ represents the echo pulse compression result at a given virtual aperture position $a_V(x_V, y_V, z_V)$ and a given real aperture position $a_R(x_R, y_R, z_R)$. λ is the wavelength of the carrier. r_V represents the slant range between imaging target point P and the mobile node at position a_V. r_R represents the slant range between point P and real-aperture array element a_R of BSs. r_V and r_R can be expressed as

$$\begin{cases} r_V = \sqrt{(x_V - x)^2 + (y_V - y)^2 + (z_V - z)^2} \\ r_R = \sqrt{(x_R - x)^2 + (y_R - y)^2 + (z_R - z)^2} \end{cases}$$

Then, the obtained point cloud $I(x, y, z)$ with a specific scattering intensity is filtered to obtain a point cloud representing the geometric shape of targets. Finally, a 3D environment model can be reconstructed by triangulating these points using a surface reconstruction algorithm, such as Poisson surface reconstruction [9], Ball Pivoting [10], Occupancy Networks [11], or DeepSDF [12].

1.2 Simulation Evaluation Result

This section describes a simulation scenario of environment sensing by using BS-terminal hybrid apertures where BSs function as receivers, and a vehicle and a UAV function as transmitters. Figure 2 shows that four BSs (BS_1 to BS_4) are deployed around the factory building (in red). A vehicle drives around the factory, while a UAV flies above the factory parallel to the ground.

In this simulation, the normal direction of the antenna arrays of BSs is parallel to the ground, and the array size is 1 m x 0.5 m. The UAV forms a horizontal virtual aperture with an equivalent length of 3 m in the air. Similarly, the moving vehicle forms five horizontal virtual apertures in different directions, each with an equivalent length of 3 m. These apertures are labeled V_1 to V_5 in Figure 2. The channel information between the transmitter and receiver can be obtained by using ray tracing method, and more channel data can be found in our open dataset, Sensiverse [13].

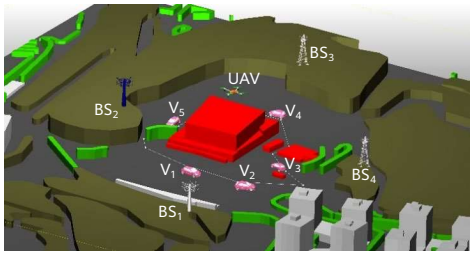
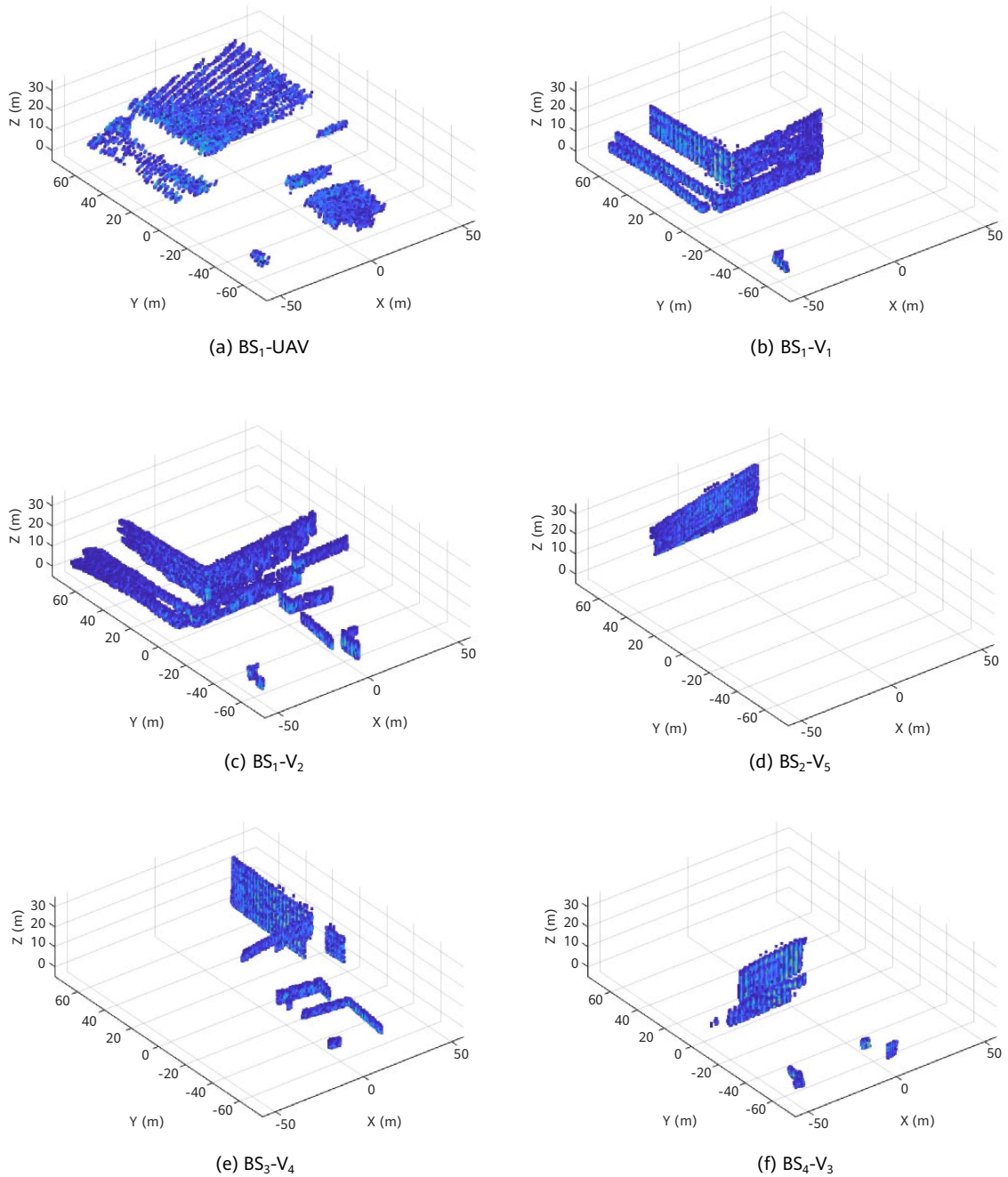


Figure 2 Simulation of HAS-based environment reconstruction

The BSs, vehicles, and UAV operate at a center frequency of 10 GHz with a bandwidth of 400 MHz. Transmitted waveforms such as OFDM, OTFS, and FMCW can be used. Echo data is processed by using imaging algorithms to obtain imaging results of the target building. Figure 3 shows the imaging results obtained through BS-mobile terminal pair sensing with the transmitted waveform set to OFDM. Due to visual angles and blocking, the coverage of each single pair of BS-mobile terminal sensing is limited, resulting in only partial imaging of the building. Therefore, the next step is to perform weighted fusion processing on the imaging results of each pair of BS-mobile terminal according to their coverage areas and corresponding theoretical resolutions.



(a) BS_1-UAV (b) BS_1-V_1

(c) BS_1-V_2 (d) BS_2-V_5

(e) BS_3-V_4 (f) BS_4-V_3

Figure 3 Imaging results based on BS-mobile terminal sensing

Figure 4 shows the fusion result of six groups of imaging point clouds. It can be seen that the fusion imaging result reflects the contour of the target building quite completely.

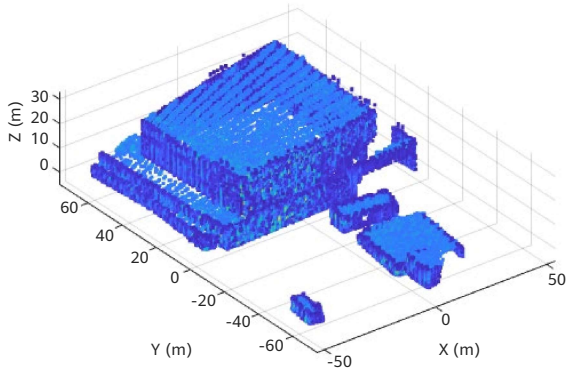


Figure 4 Fusion result of HAS-based imaging point clouds

It should be noted that while HAS can achieve higher-resolution sensing results than BS only sensing, the time and frequency synchronization errors between transmitting and receiving nodes will affect the sensing performance. This is a problem that needs to be further explored and studied in the future.

1.3 Point Cloud Modeling and Performance Evaluation

Based on the environment imaging results of the previous section, we can reconstruct a precise geometric model (shown in Figure 5) that is easy to express, process, and transmit and contains physical characteristics. This model is crucial for various applications, including three-dimensional modeling in indoor and outdoor scenarios, wireless channel estimation, mixed reality, and digital twin.

The problem of reconstructing a geometric model on a point cloud can be described as follows: given a set of unordered three-dimensional points $\mathrm{P} = \{p_1, p_2, \ldots, p_n\}$, the goal is to reconstruct a mesh with clear geometric topology structure, $\mathrm{Mesh} = \{V, E\}$, using a surface reconstruction algorithm, where V and E represent the coordinates of the mesh vertices and the connectivity relationships between them, respectively.

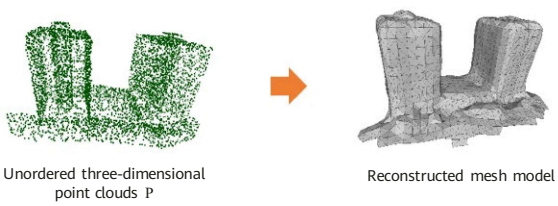


Figure 5 Reconstructed mesh model based on point clouds

Quality evaluation of the reconstructed mesh model typically involves comparing the difference between the sampling points in the reconstructed mesh model and the true value of scattering points. Common evaluation metrics include Chamfer Distance [14] and F-score [15], which are defined as follows:

$$CD\left(S_1, S_2\right) = \frac{1}{|S_1|}\sum_{x \in S_1} \min_{y \in S_2} \|x-y\|_2^2 + \frac{1}{|S_2|}\sum_{x \in S_2} \min_{y \in S_1} \|x-y\|_2^2 \quad (2)$$

S_1 and S_2 respectively represent the sampling points of the reconstructed model and the true value of scattering points. To use the length dimension, this paper uses the square root of Chamfer Distance, i.e., $\sqrt{\text{Chamfer Distance}}$ as one of the metrics for measuring the reconstruction quality. And

$$F-score = \frac{2P\left(d\right) \cdot R\left(d\right)}{P\left(d\right) + R\left(d\right)} \quad (3)$$

where $P(d)$ and $R(d)$ respectively represent the matching degree between the reconstructed model and the true value of scattering points with respect to a given distance threshold d, and are defined as follows:

$$P\left(d\right) = \frac{100}{|S_1|}\sum_{x \in S_1}\left[\min_{y \in S_2} \|x-y\|_2 < d\right]$$

$$R\left(d\right) = \frac{100}{|S_2|}\sum_{x \in S_2}\left[\min_{y \in S_1} \|x-y\|_2 < d\right]$$

By leveraging the environment imaging results from the previous section as input and using the Poisson surface reconstruction algorithm, a geometric model is reconstructed based on the input point clouds, as shown in Figure 6. The reconstructed model is evaluated quantitatively using the $\sqrt{\text{Chamfer Distance}}$ and F-score. As presented in Table 1, the reconstructed mesh model exhibits $\sqrt{\text{Chamfer Distance}}$ of 0.96 m, and an F-score with a distance threshold of 1.5 m is 96. In contrast, the original input point clouds exhibit $\sqrt{\text{Chamfer Distance}}$ of 1.18 m and an F-score of 94 using the same distance threshold. By comparing the quantitative error evaluation result in Table 1, it is evident that the reconstructed geometric model is more representative of the environment than the original point cloud, and the $\sqrt{\text{Chamfer Distance}}$ and F-score values are better than those of imaging point clouds.

Figure 6 3D environment reconstruction result

Table 1 Evaluating the environment reconstruction error

Performance	Imaging Point Cloud	Reconstructed Model
$\sqrt{\text{Chamfer Distance}}$	1.18 m	0.96 m
F-score (d = 1.5 m)	94	96

2 Environment Sensing-assisted Communication: Channel Prediction

A 3D reconstruction of the surrounding environment is obtained through 3D imaging and meshing. Let's assume that the objects are made of concrete. For a UE at a given location in the environment, multipath parameters $\{\tau_n, A_n, \text{AOD}_n, \text{ZOD}_n, \text{AOA}_n, \text{ZOA}_n\}$ between a BS and the UE can be obtained by using wireless channel prediction methods such as electromagnetic calculation and ray tracing. These multipath parameters represent the delay, amplitude, azimuth angle of departure (AOD), zenith angle of departure (ZOD), azimuth angle of arrival (AOA), and zenith angle of arrival (ZOA) of the nth multipath, respectively. Based on these parameters, the frequency domain response of the channel can be further calculated.

In channel prediction based on the reconstructed environment, errors are inevitable due to inaccuracies in environment reconstruction and UE locations. Specifically, errors in environment reconstruction are primarily geometric errors of a 3D environment and errors in the electromagnetic parameters of materials. Errors in environment reconstruction can occur due to (1) a failure in perfectly reconstructing the environment and (2) a failure in real-time imaging for dynamic targets (such as vehicles or pedestrians) in the environment. UE locations in the environment are subject to dynamic changes, which presents challenges for accurate UE localization. In practice, only estimated UE locations can be used for prediction. This paper primarily focuses on examining the impact of errors in environment reconstruction.

To quantitatively evaluate the impact of errors in environment reconstruction on channels, let's consider a frequency of 10 GHz as an example. Figure 7 depicts one BS located at a specific location and 500 UEs randomly distributed. Using the ray tracing method, we obtain multipath parameters for each BS–UE pair based on an ideal model and a reconstructed model. For statistical analysis, we only consider UEs with attenuation of the strongest path less than 110 dB (indicated by red points). Among these UEs, we reserve the multipaths for which energy attenuation is less than 20 dB when compared with the strongest path for each UE. This enables us to use the Chamfer Distance and compare the errors in multipath parameters for the two models. Figure 8 shows the multipath parameters obtained separately based on the two models, and the Chamfer Distance is used to compare the errors in these parameters.

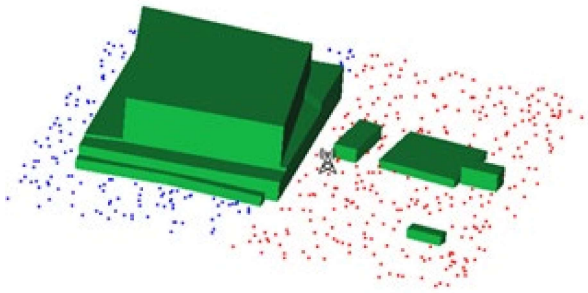


Figure 7 Scenario configuration for analysis of errors in channel multipath prediction

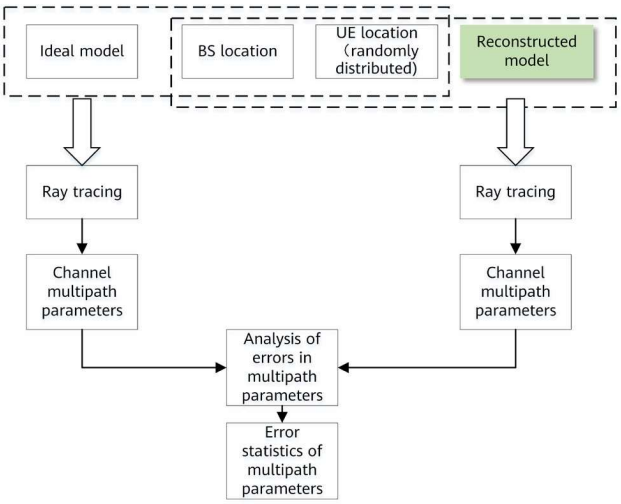


Figure 8 Process for analyzing the channel prediction errors

Figure 9 shows the statistical results of the error analysis for the multi-path parameter prediction based on the reconstructed model. From the statistical results, it can be seen that the time delay error is within 3.5 ns for 90% of the cases, the AOA error is within 3 degrees for 90% of the cases, the AOD error is within 3 degrees for 90% of the cases, the ZOA error is within 1 degree for 90% of the cases, and the ZOD error is within 1 degree for 90% of the cases. The channel prediction result can be used to assist beam sweeping in communications, reducing the beam sweeping overhead.

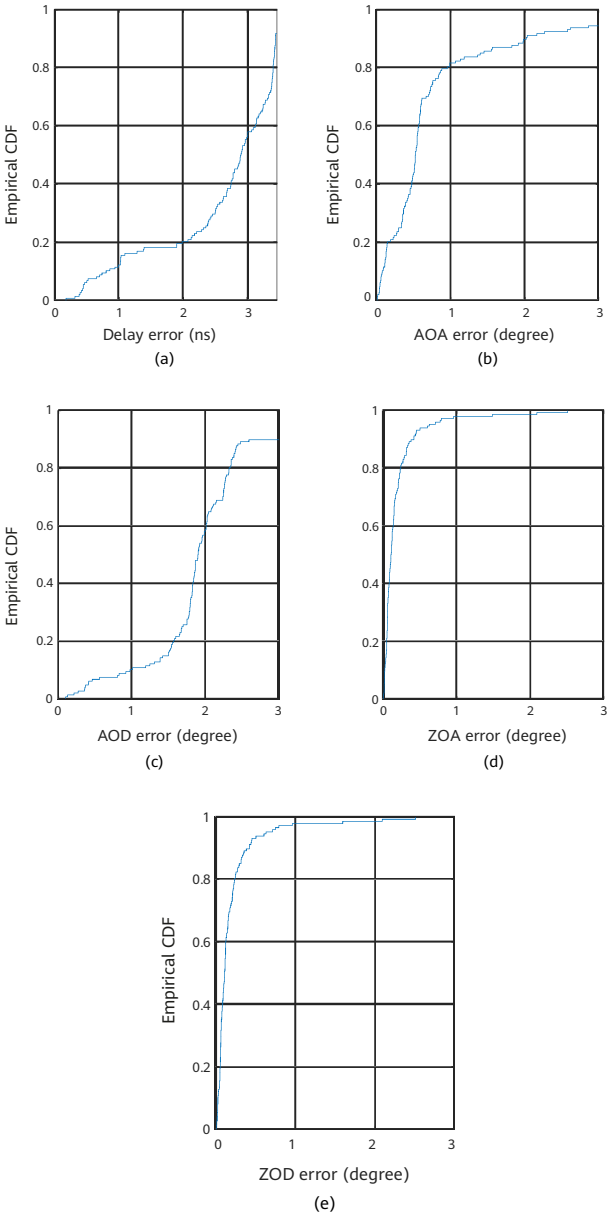


Figure 9 (a) Delay errors; (b) AOA errors; (c) AOD errors; (d) ZOA errors; (e) ZOD errors

3 Environment Sensing-assisted Localization: Clutter Suppression

In existing cellular networks, the primary focus of positioning and tracking of targets is on active devices, such as smartphones, vehicles, and UAVs. In the future, with the introduction of sensing functions, 6G networks will be able to provide positioning and tracking not only for active devices but also for passive targets. This will open up a range of new application scenarios, including but not limited to flight tracking of UAVs, UAV intrusion detection, localization and tracking of vehicles, detection of pedestrians and animals on highways, and communications assisted by sensing results of moving targets. This section takes passive targets as an example for analysis, and the mobile targets mentioned below refer to passive targets unless specified otherwise.

Targets to be sensed and observation nodes are usually located on the ground or in a low-altitude environment, which makes ground clutter (echoes from the ground and surrounding buildings) a significant factor that affects the system sensing performance. This is especially true for detecting, localizing, and tracking small targets that move slowly in a low-altitude environment. The main problems faced include:

1. Both the reflected signals of the targets to be detected and the clutter reflected from surrounding buildings and the ground exist in the environment. If coherent accumulation is performed without any additional processing, strong side-lobe clutter may obscure targets, significantly reducing the target detection capability of BSs. Therefore, measures must be taken to suppress clutter and minimize its impact on target detection.
2. Targets that move at low speeds and have weak reflected energy. Moving targets in the environment mainly include vehicles, pedestrians, and UAVs. Among these, pedestrians move at a relatively low speed and have a small radar cross section (RCS), making them particularly susceptible to static side-lobe clutter. Consequently, detecting such targets requires a higher level of interference suppression capability.

Next, we will illustrate the performance of mobile target sensing based on a monostatic BS using a simulation example.

In the simulation, the carrier frequency is set to 10 GHz, and the bandwidth is set to 100 MHz. The BS use angle and range measurements to locate the target, with a single transmitting element and 1024 receiving elements configured in a two-dimensional uniform planar array. The receiving elements are spaced at a distance of $\lambda/2$, where λ is the wavelength of the signal, and the configuration can be equivalent to two mutually perpendicular 32-element receive uniform linear arrays using MIMO technology, without affecting the analysis results. The 32 x 32 element array is placed on the YZ plane, with its normal direction along (1,0,0) to cover the moving target. Based on monostatic BS sensing, the target's angle, range, and speed information can be obtained.

The target moves along a straight line at a constant speed, with a trajectory length of 30 m. In this example, we use a cylindrical target with a radius and height of 3 m as an extended target. The signal processing flowchart is shown in Figure 10, and it should be noted that an environment-assisted module is added compared to traditional methods (such as ECA, extensive cancellation algorithm). The reflected signal in the frequency domain can be represented as $Y(f) = S(f)*H(f)$, where $S(f)$ is the Fourier transform of the transmitted waveform $s(t)$, and $H(f)$ is the frequency-domain response of the channel. After obtaining the echo data, the target angle, range, and Doppler information can be obtained through signal processing, and finally, the target localization information can be obtained.

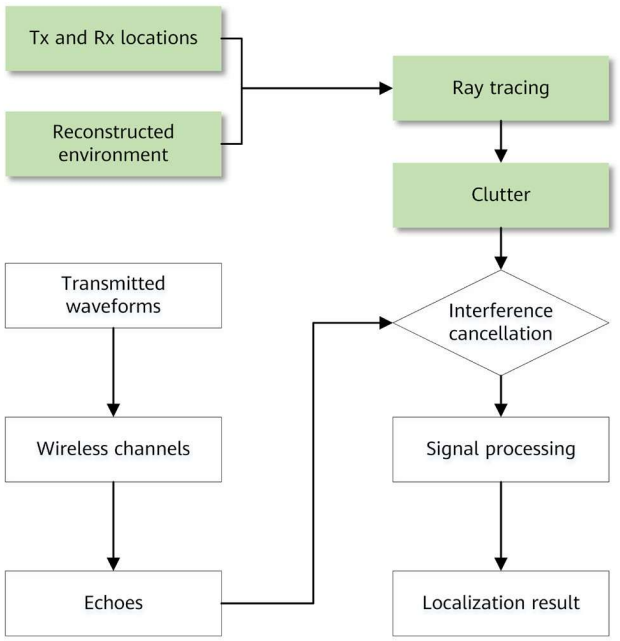


Figure 10 Flow of signal processing for detecting moving targets

As previously mentioned, wireless sensing can be used to obtain an environment reconstruction result. This result and the transceiver location can be used to obtain multipath information by using channel prediction algorithms such as electromagnetic calculation and ray tracing. By relying on multipath estimation for the static environment, background clutter can be suppressed based on the environment information.

As shown in Figure 11a and Figure 11b, we use $\sqrt{\text{Chamfer Distance}}$ to quantitatively evaluate the accuracy of target localization. Assuming that the true scattering point coordinates of the target are S1 (as shown in red line in Figure 11), and the detected target localization coordinates are S2 (as shown in blue point in Figure 11), the error of the target localization is 1.73 m when not using the reconstructed environment prior information, and 1.07 m when using the reconstructed environment prior information. Figure 11c shows the corresponding detection results. It can be seen that using the prior information of the reconstructed environment can detect more points. Without using the prior information of the reconstructed environment, some target points will be buried by environmental echoes.

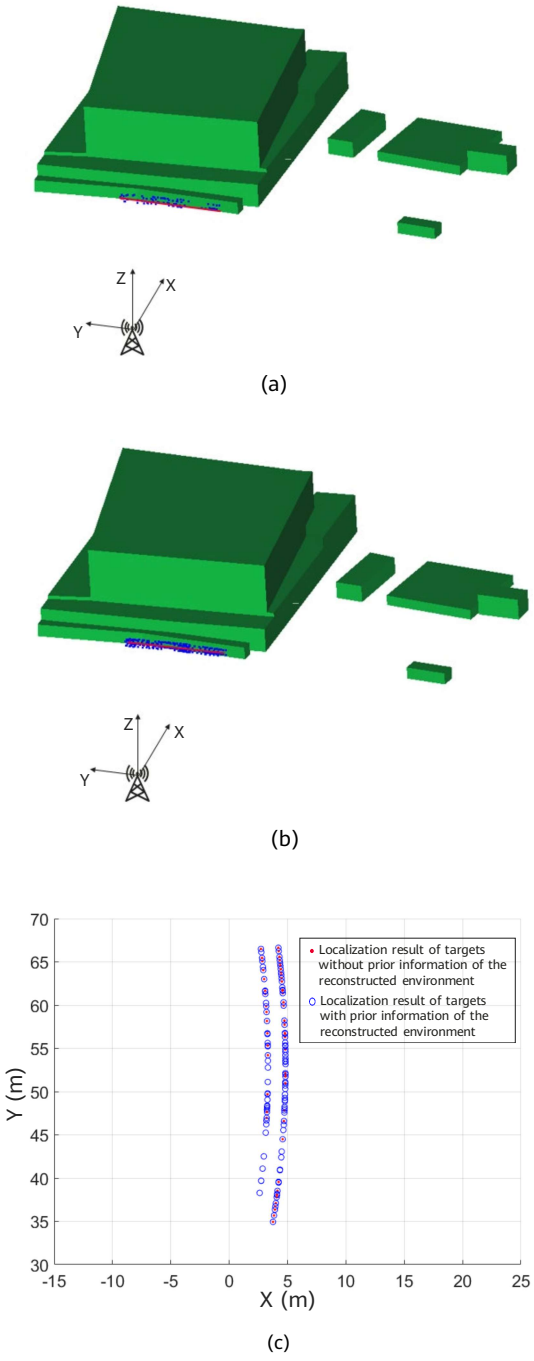


Figure 11 (a) Detection result without using prior information of the reconstructed environment; (b) Detection result using prior information of the reconstructed environment; (c) Projection of the localization results on a two-dimensional plane

4 Conclusions

This paper outlines an end-to-end solution covering HAS, digital twin environment reconstruction, and digital twin-assisted communication and localization. We evaluated the performance in environment reconstruction, parameter estimation, and target localization by simulation. The simulation results demonstrate that HAS enables submeter-level high-precision environment reconstruction in large urban scenarios on 10 GHz (centimeter waves). Based on the reconstructed environment model, ray tracing is used to predict the channels of UEs at different locations. In addition, we provided the cumulative distribution functions (CDFs) for errors in terms of delay and angles. Furthermore, the reconstructed environment model is used to enhance the detection of extended moving targets. By using environment-assisted clutter suppression, we improved the localization accuracy by approximately 40% compared to the traditional ECA approach. It should be noted that non-ideal factors, such as bistatic synchronization errors and changes in surface materials of buildings, were not considered in the simulation presented in this paper. To ensure high precision in the digital twin, several technical aspects, such as mitigating the impact of non-ideal factors and extracting the electromagnetic scattering coefficients of object surfaces, need to be further explored.

References

[1] A. Bayesteh, J. He, Y. Chen, P. Zhu, J. Ma, A. W. Shaban, Z. Yu , Y. Zhang, Z. Zhou, and G. Wang, "Integrated Sensing and Communication (ISAC) — From Concept to Practice," 11 2022. [OL]. Available: https://www.huawei.com/en/huaweitech/future-technologies/integrated-sensing-communication-concept-practice

[2] J. Gao, T. Shen, Z. Wang, W. Chen, K. Yin, D. Li, O. Litany, Z. Gojcic, and S. Fidler, "Get3D: A generative model of high quality 3D textured shapes learned from images," *Advances In Neural Information Processing Systems 35,* pp. 31841–31854, 2022.

[3] Q. Xiaolan, J. Zekun, Y. Zhenli, C. Yao, L. Bei, L. Yitong, W. Wei, D. Yongwei, Z. Liangjiang, and D. Chibiao, "Key technology and preliminary progress of microwave vision 3D SAR experimental system," *Journal of Radars,* vol. 11, p. 1–19, 2022.

[4] K. Qian, Z. He and X. Zhang, "3D point cloud generation with millimeter-wave radar," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, p. 1–23, 2020.

[5] 3GPP, "TR 22.837 Feasibility study on integrated sensing and communication," [OL]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4044

[6] Y. Yang, F. Gao, X. Tao, G. Liu, and C. Pan, "Environment Semantics Aided Wireless Communications: A Case Study of mmWave Beam Prediction and Blockage Prediction," *IEEE Journal on Selected Areas in Communications*, 2023.

[7] A. F. Yegulalp, "Fast backprojection algorithm for synthetic aperture radar," *Proceedings of the 1999 IEEE Radar Conference. Radar into the Next Millennium (Cat. No. 99CH36249)*, 1999.

[8] M. I. Duersch, "Backprojection for synthetic aperture radar," Brigham Young University, 2013.

[9] M. Kazhdan, M. Bolitho and H. Hoppe, "Poisson surface reconstruction," *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006.