Advances and Prospects in SAR Microwave Vision Three-Dimensional Imaging

<table>
<thead>
<tr>
<th>Journal</th>
<th>National Science Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>NSO20240009.R1</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Review</td>
</tr>
<tr>
<td>Date Submitted by the Author</td>
<td>06-Jun-2024</td>
</tr>
<tr>
<td>Complete List of Authors</td>
<td>Qiu, Xiaolan; Chinese Academy of Sciences Aerospace Information Research Institute</td>
</tr>
<tr>
<td></td>
<td>Jiao, Zekun; Chinese Academy of Sciences Aerospace Information Research Institute,</td>
</tr>
<tr>
<td></td>
<td>Zhang, Zhe; Chinese Academy of Sciences Aerospace Information Research Institute,</td>
</tr>
<tr>
<td></td>
<td>Yan, Qiancheng; Chinese Academy of Sciences Aerospace Information Research Institute</td>
</tr>
<tr>
<td></td>
<td>Ding, Chibiao; Chinese Academy of Sciences Aerospace Information Research Institute</td>
</tr>
<tr>
<td>Keywords</td>
<td>SAR, Microwave vision, 3D imaging, Visual semantics, Scattering mechanism</td>
</tr>
</tbody>
</table>

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Figures-1-10-JPEG.zip  
Figures-11-21-JPEG.zip
Advances and Prospects in SAR Microwave Vision Three-Dimensional Imaging

Xiaolan Qiu\textsuperscript{1,2,3,*} & Zekun Jiao\textsuperscript{1} & Zhe Zhang\textsuperscript{3} & Qiancheng Yan\textsuperscript{1,2} & Chibiao Ding\textsuperscript{1,2,*}

\textsuperscript{1}\textit{National Key Laboratory of Microwave Imaging, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China}
\textsuperscript{2}\textit{School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100094, China}
\textsuperscript{3}\textit{Key Laboratory of Intelligent Aerospace Big Data Application Technology, Suzhou 215123, China}

*Corresponding authors (emails: xlqiu@mail.ie.ac.cn (Xiaolan Qiu); cbding@mail.ie.ac.cn (Chibiao Ding))

Abstract: Synthetic Aperture Radar three-dimensional imaging enables the acquisition of more comprehensive information, making it a recent hotspot in radar imaging. Traditional three-dimensional imaging methods have evolved from two-dimensional and interferometric imaging, combining elevation aperture extension with signal processing techniques. Limitations such as long acquisition or complex system from its imaging mechanism restrict its application. In recent years, rapid development of artificial intelligence has led to a swift advancement in radar, injecting new vitality into SAR three-dimensional imaging. SAR microwave vision three-dimensional imaging theory, which is built upon advanced technologies, has emerged as a new interdisciplinary field for radar imaging. This paper reviews SAR 3D imaging’s history and present, introducing SAR microwave vision. We establish a theoretical framework covering representation models, computational models, processing paradigms, evaluation systems. Additionally, our research progress in this area is discussed, along with future prospects for SAR microwave vision 3D imaging.

Keywords: SAR, Microwave vision, 3D imaging, Visual semantics, Scattering mechanism

1. INTRODUCTION

1.1 Current Status of SAR Imaging: From 2D to 3D

Synthetic Aperture Radar (SAR), as an active microwave sensing technology, enables all-weather and all-day earth observation imaging. It plays an irreplaceable role in applications such as terrain mapping, disaster response, and smart cities. Traditional SAR imaging produces two-dimensional images, representing the projection of the three-dimensional physical reality onto a two-dimensional imaging plane. Due to the isometric projection nature of SAR imaging, issues such as perspective distortion and layover exist, significantly affecting its widespread application[1].

To address these challenges, researchers worldwide have proposed imaging systems such as Interferometric SAR (InSAR) and Stereo SAR, aiming to reconstruct stereo images of observed scenes using two images from different viewing angles[2][3]. InSAR utilizes the phase difference observed at two different elevation angles to reconstruct height information, while
Stereo SAR, also known as radar photogrammetry, employs concepts similar to optical image disparity to calculate the three-dimensional position of targets using the range-Doppler equation. However, these approaches face difficulties in resolving the layover phenomenon\cite{4}. In practical remote sensing applications, areas with steep terrain changes, such as mountains and urban environments, often experience significant layover issues. Both InSAR and Stereo SAR, inherently, provide a form of 2.5D imaging, unable to distinguish multiple scatterers affected by layover, thus limiting their application in urban and mountainous regions\cite{5}\cite{6}.

To overcome this limitation, the U.S. Naval Research Laboratory proposed the concept of SAR three-dimensional imaging using elevation synthetic aperture in 1995. That same year, European Microwave Signal Laboratory (EMSL) achieved the resolution of two overlapping metal spheres using eight-track observations, marking a leap from theory to practice for SAR three-dimensional imaging\cite{7}. Following this breakthrough, numerous research institutions and researchers worldwide have conducted extensive studies. Currently, two main approaches have emerged: the first involves repeated flights for data acquisition called Tomographic SAR, and the second utilizes a single flight with a multi-channel array for data acquisition called Arrayed Interferometric SAR\cite{8}.

In the case of Tomographic SAR, both airborne and spaceborne platforms have been utilized in experiments globally, confirming the feasibility and effectiveness of this approach. The German Aerospace Center conducted multiple TomoSAR data acquisition experiments using its airborne E-SAR and F-SAR systems, achieving three-dimensional imaging in forested areas and successfully distinguishing tree canopies, vehicles beneath the canopy, and terrain\cite{9}. In China, the Institute of Electronics, Chinese Academy of Sciences, conducted the first airborne TomoSAR data acquisition experiment in 2013\cite{8}. On the spaceborne side, the University of Naples in Italy used ERS satellite data collected between 1992 and 1998 to obtain the first results of spaceborne TomoSAR experiments in the Naples region in 2005\cite{10}\cite{11}. In 2010, the German Aerospace Center used TerraSAR-X satellite data collected over more than a year to achieve high-resolution three-dimensional imaging results in complex urban areas\cite{12}. In 2013, the University of Pisa in Italy conducted TomoSAR three-dimensional imaging experiments using 28 repeated Cosmo-Skymed satellite images\cite{13}. In China, institutions such as the Aerospace Information Research Institute, Chinese Academy of Sciences and universities like Nanjing University of Aeronautics and Astronautics used high-resolution images from domestic satellites like GF-3 to achieve three-dimensional reconstruction, demonstrating the potential of domestic satellites for three-dimensional imaging\cite{14}.

As for Array Interferometric SAR, in 2004, the French Aerospace Lab (ONERA) developed an unmanned aerial vehicle (UAV)-borne downward-looking array three-dimensional SAR system called the DRIVE system. This system employed a 12-transmitter, 12-receiver array radar for data acquisition\cite{15}\cite{16}\cite{17}. In 2005, the German FGAN also developed a UAV-borne downward-looking array three-dimensional SAR system named ARTINO, using a 44-transmitter, 32-receiver array with a total of 1408 equivalent channels for data acquisition\cite{18}\cite{19}\cite{20}. However, detailed three-dimensional imaging results from these two radar systems have not been publicly disclosed to date. Downward-looking array three-dimensional imaging, due to its use of a nadir observation configuration, faces challenges of ambiguity in the cross-track direction. In 2015, the Institute of Electronics, Chinese Academy of Sciences, used an aircraft platform carrying a 2-transmitter, 8-receiver array antenna, employing a side-looking observation mode to achieve SAR three-dimensional imaging over
large scenes[21]. This achievement marked the world's first Array Interferometric SAR three-dimensional imaging results.

(2) Challenges in 3D SAR Imaging

A comprehensive analysis of the two current 3D imaging systems, namely Tomographic SAR and Array Interferometric SAR, reveals that, at a fundamental level, both utilize multiple-angle observations in the elevation direction to construct synthetic aperture. In terms of information sources, they rely on the phase difference information provided by spatially diverse observations. In terms of processing modes, the focus is primarily on isolating processing for each pixel of the 2D SAR image. These factors collectively contribute to the significant requirement for a large number of multi-angle SAR images in the current 3D SAR imaging process.

For Tomographic SAR, this necessitates a large number of repeated flight trajectories, leading to a high volume of data acquisition, extended data acquisition cycles, and consequently, high time costs. In the case of Array Interferometric SAR, the requirement for a large number of channels in the array interferometric SAR system results in high economic costs and system complexity. These challenges have become the primary obstacles in the widespread application and promotion of 3D SAR imaging.

(3) Development Trends

To address the aforementioned challenges and reduce the time and economic costs associated with SAR 3D imaging, it is important to reevaluate the issue from the essence of 3D imaging. SAR 3D imaging primarily tackles the inverse problem of deducing the three-dimensional characteristics of objects from the information contained in the echo signals. The more comprehensive the information extracted from the echo signals, the smaller the solution space, and the higher the reconstruction accuracy. Considering the current challenges in 3D imaging, a key conclusion can be drawn. The fundamental reason for the high cost of SAR 3D imaging lies in the underutilization of information within the echo data. Analysis reveals that SAR echo signals contain rich information. From the perspective of information types, echo signals encompass multidimensional electromagnetic scattering information, while SAR images contain geometric structural information of targets.

In recent years, the advancement of computer technology and increased hardware computing power have spurred the emergence of new technologies capable of supporting the extraction of electromagnetic scattering information from echo signals and the extraction of semantic information from target images. Specifically, the rapid development of computational electromagnetics and computer vision has provided new perspectives for SAR 3D imaging.

Current SAR 3D imaging has not fully integrated microwave characteristics and image semantics, failing to efficiently utilize both microwave scattering information and visual information from the images. However, integrating these advanced technical means with traditional imaging poses significant challenges. The core difficulties lie in two aspects. The first is how to extract information favorable for 3D imaging from computational electromagnetics and computer vision. The second is how to incorporate the unstructured, qualitative information into existing frameworks based on current SAR 3D imaging geometry. SAR Microwave Vision 3D imaging theory provides answers to these challenges.

In the following sections of this paper, we will provide a detailed explanation of the conceptual connotations and theoretical framework of SAR microwave vision 3D imaging. In terms of conceptual connotations, this includes relevant definitions and core concepts, as well
as a comparison with traditional 3D imaging theories. Regarding the theoretical framework, we will present its representation model, computational model and the processing paradigm. In Section 4 and 5, recent progresses and future trends about SAR microwave vision 3D imaging are discussed.

2. Concept and Connotation

(1) Definition of SAR Microwave Vision 3D Imaging

Synthetic Aperture Radar (SAR) Microwave Vision 3D Imaging is a scientific methodology with the ultimate goal of achieving three-dimensional imaging. In this process, SAR, microwave, and vision are considered instrumental means. The scientific definition of SAR Microwave Vision 3D Imaging involves the extraction of information conducive to three-dimensional imaging from SAR echo data and images through the application of microwave vision. This extracted information is organically integrated with the mathematical models of SAR 3D imaging to formulate computationally feasible models and methods. The overarching goal is to reduce dependence on the quantity of observations and enhance the precision of three-dimensional imaging[1].

The term microwave vision represents a form of physical intelligence, wherein physical intelligence combines a robust theoretical framework of physics to construct an intelligent system better suited to adapting to the physical world than humans. In the specific context of SAR 3D imaging, microwave vision aims to develop a form of physical intelligence that integrates microwave scattering mechanisms, computer vision, and relevant radar imaging theories. This integration is envisioned to better extract information conducive to three-dimensional imaging.

The information beneficial for three-dimensional imaging is categorized as follows in the presented table, where the first and second represent universally significant information for the entire scene, while the third and fourth denote information relevant to typical large-scale and small-scale structures. As for the last kind, it mainly concerns about the microwave scattering mechanism of each pixel.
<table>
<thead>
<tr>
<th>ID</th>
<th>Contents</th>
<th>Descriptions</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Target identification, Semantic segmentation,</td>
<td>(1) Key focus targets such as buildings, bridges, oil tanks, vehicles, ships, and power facilities are identified.</td>
<td>(1) Semantic information aids in selectively choosing more suitable methods and parameters for different targets, for example, artificial targets and natural features.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2) The segmentation process involves distinguishing between the identified targets and the background of natural features, as well as segmenting the areas occupied by individual instances of similar targets.</td>
<td>(2) Target identification and instance/semantic segmentation also lay the foundation for subsequent extraction of spatial structural elements and other related tasks.</td>
</tr>
<tr>
<td>2</td>
<td>Elevation Inference</td>
<td>(1) Inferring pixel heights by learning SAR images, such as top-bottom inversion and shadows, for example.</td>
<td>(1) Providing a prior for height estimation reduces solution space in the height direction.</td>
</tr>
<tr>
<td>3</td>
<td>Large-scale spatial structures</td>
<td>(1) Extract spatial structural elements, including line elements such as straight lines, curved lines, etc. and surface elements such as horizontal planes, vertical planes, etc., from the visual content of images and the initial point cloud in 3D imaging.</td>
<td>(1) Providing prior information on three-dimensional positions helps reduce the solution space in the elevation direction, facilitating the three-dimensional imaging of overall structures.</td>
</tr>
<tr>
<td>4</td>
<td>Small-scale spatial structures</td>
<td>(1) For complex targets where it is not possible to directly extract structural elements on the aforementioned large-scale structures, numerous local surface elements are extracted to constitute the complex structure.</td>
<td>(1) Providing local position priors and constraints on correlation and continuity contributes to the three-dimensional imaging of fine structures.</td>
</tr>
<tr>
<td>5</td>
<td>Micron-scale microwave scattering elements</td>
<td>(1) Identification and parameter inversion of microwave scattering mechanisms through the analysis of spatial-temporal-frequency-polarization multidimensional data.</td>
<td>(1) Providing structural parameters of scattering mechanisms aids in the reconstruction of local fine three-dimensional structures.</td>
</tr>
</tbody>
</table>
(2) Core Concepts in SAR Microwave Vision 3D Imaging

(a) 3D SAR Observation Model

The 3D SAR observation model refers to the construction of a three-dimensional SAR imaging model using multi-channel observation data for elevation synthetic aperture. The three dimensions in 3D SAR imaging refer to range, azimuth, and elevation. The resolution in range and azimuth directions remains the same as in the original 2D SAR imaging. Range resolution is achieved through pulse compression by transmitting wide-bandwidth signals for high resolution, while azimuth resolution is obtained through platform motion forming synthetic aperture for resolution.

(b) SAR Image Semantic Information

In the context of SAR microwave vision 3D imaging, SAR image semantic information refers to semantic details extractable from 2D SAR images using computer vision techniques to assist in 3D imaging. In computer vision, image semantics generally denote the meaning of the image content, categorized hierarchically into different layers. Despite differences in imaging mechanisms between SAR and optical images, resulting in phenomena like layover and perspective contraction, SAR images still project the physical world onto a 2D plane. Structural information of target objects in 3D space is preserved in the 2D image. Specifically, for the purpose of SAR 3D imaging, the visual layer corresponds to geometric structural information of observed objects, focusing on aspects such as shapes and edges. The object layer corresponds to category information of observed objects, such as buildings, vehicles, etc. The conceptual layer corresponds to information at the entire image level, such as identifying whether the scene represents a mountainous or urban area. Information from the visual and object layers provides strong prior information for fine-grained targets, constraining the targets to their respective geometric structures. Semantic prior information from the conceptual layer guides the selection of processing methods, as different observation scenes benefit from different processing theories, such as sparse signal processing and spectrum estimation.
(c) 3D Scattering Mechanisms

The imaging characteristics of target scattering and its correlation with geometric and physical parameters such as shape and material, waveform parameters such as frequency and polarization, and observation conditions including incident angle and mode in SAR images are multifaceted. 3D scattering mechanisms involve the exploration of pixel-level scattering mechanisms in SAR images to provide fine-grained prior information for 3D imaging.

(3) Comparison

In this subsection, a comparison will be made from different perspectives between traditional SAR 3D imaging and SAR microwave vision 3D imaging, highlighting the distinctions in their imaging modalities.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Traditional SAR 3D imaging</th>
<th>SAR Microwave vision 3D imaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaging dimensions</td>
<td>Range-azimuth-elevation</td>
<td>Range-azimuth-elevation</td>
</tr>
<tr>
<td>Resolution Mechanism</td>
<td>Range : pulse compression</td>
<td>Range : pulse compression</td>
</tr>
<tr>
<td></td>
<td>Azimuth : synthetic aperture</td>
<td>Azimuth : synthetic aperture</td>
</tr>
<tr>
<td></td>
<td>Elevation : synthetic aperture</td>
<td>Elevation : microwave vision</td>
</tr>
<tr>
<td>Information sources</td>
<td>Multi-angle observations</td>
<td>SAR image visual semantic information</td>
</tr>
<tr>
<td></td>
<td>scattering mechanism information</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 illustrates the similarities and differences between the existing SAR three-dimensional imaging system and the proposed SAR microwave vision three-dimensional imaging system in various dimensions. Both systems serve the same purpose of achieving three-dimensional imaging. Therefore, they both implement three-dimensional imaging in the range-azimuth-elevation dimensions, and the geographic coordinates of the three-dimensional images can be obtained through coordinate transformation.

In terms of the three-dimensional resolution mechanism, both systems use a pulse compression technique with a transmitted wideband signal for range resolution and employ the synthetic aperture formed by the motion of the platform for azimuth resolution. For elevation resolution in traditional methods, it relies on multi-angle observations to construct a synthetic aperture. In contrast, SAR microwave vision three-dimensional imaging achieves elevation resolution not only through multi-angle observations but also by utilizing semantic information extracted from SAR images and target scattering mechanism information. Thus, its elevation resolution mechanism is based on microwave vision theory.

Regarding the processing paradigm, traditional methods follow a single-direction open-loop approach with pixel-wise processing, isolating each pixel without considering spatial adjacency relationships between pixels. The process utilizes multi-channel two-dimensional SAR images for signal processing to obtain three-dimensional imaging results, representing a one-way processing flow. SAR microwave vision three-dimensional imaging introduces a new processing paradigm, i.e., closed-loop feedback and neighborhood collaborative processing. Neighborhood collaborative processing involves the coordinated processing of multiple pixels defined in a generalized neighborhood. This paradigm shift leads to a qualitative leap in three-dimensional imaging results. Closed-loop feedback means that the three-dimensional imaging process doesn’t end with obtaining results. Instead, the results are used to correct errors, extract semantic information, and derive scattering mechanism information in the processing flow, providing more accurate prior information constraints for the next round of three-dimensional reconstruction.

The disciplinary foundations of the two systems also differ. Traditional methods primarily involve radar signal processing, whereas the proposed method is an interdisciplinary product that incorporates advanced techniques from computational electromagnetics and computer vision theory on the basis of radar signal processing.

3. Theoretical Framework

The theoretical framework of SAR microwave vision 3D imaging is fundamentally about processing information from radar echo data to generate high-quality 3D images. Therefore, it involves addressing three challenges, namely the representation, computation, and processing of information. Thus, the theoretical framework of SAR microwave vision 3D imaging primarily encompasses the following components.

(1) Representation Model

At the current stage, the two SAR three-dimensional imaging methods, as described in the conceptual framework in previous section, are based on aperture synthesis in the elevation direction. Based on this concept, widely used 3D imaging model are as follows.

\[ y = \Phi \sigma(s) + n \]

\[ \sigma(s) \text{ s.t. } \sigma \text{ is sparse on } s \]

In above equations, \( y = [y_1, y_2, ..., y_N]^T \) is the multi-channel observation vector. \( \Phi \) is the sensing matrix,
\[ \Phi = [\phi_{n,m}]_{N \times M}, \quad \phi_{n,m} = \exp(-j2\pi r_n r_m). \] In this model, the target scattering coefficients are considered to be sparsely distributed in the elevation direction.

This paper provides the new framework for SAR microwave vision three-dimensional imaging as follows. It should be noted that SAR microwave vision three-dimensional imaging utilizes multi-angle and multi-polarization observations, and the characterization framework incorporates the concept of holographic SAR proposed by the authors. For details on holographic SAR, please refer to the reference provided. Therefore, considering the holographic characteristics of the targets, the characterization model for SAR microwave vision three-dimensional imaging is as follows:

\[ y = \Phi \sigma(p, f, k, t, r, L, \alpha, \gamma, \varphi) + n \]

\[ \sigma(p, f, k, t, r, L, \alpha, \gamma, \varphi) \text{ s.t. } \begin{cases} \sigma \text{ is sparse on } r, \\ \zeta(\sigma, r) \in R, \\ r \in F, \\ \sigma(\varphi) \in D. \end{cases} \] (2)

\( \sigma(p, f, k, t, r, L, \alpha, \gamma, \varphi) \) is used to comprehensively consider the target's holographic scattering characteristics for the sought scattering coefficients. The terms in parentheses represent dependencies on the radar system and target scattering characteristics. Specifically, \( p \) represents the unit Jones vector indicating the polarization state of the electromagnetic field. \( f \) is the carrier frequency, \( k \) is the unit vector characterizing the direction of electromagnetic wave propagation. \( t \) characterizes the moment of propagation. \( r \) is the three-dimensional position vector of the target and \( L \) is the three-dimensional size of the target. \( \alpha \) and \( \gamma \) are the frequency-dependent factor and the curvature-dependent factor, respectively. \( \varphi \) is the target's positional vector relative to the radar, including pitch and azimuth angles.

The lower part of this equation is the restrictions to be satisfied for the unknown vector \( \sigma(p, f, k, t, r, L, \alpha, \gamma, \varphi) \). Before going into detailed explanations, it is important to note that the presented expressions here are abstract representations. In practical applications, there are more specific and concrete forms. The provided analysis here offers a macroscopic perspective. The first constraint is the sparse distribution of the target's scattering coefficients in three-dimensional space. This is a constraint commonly employed in traditional SAR 3D imaging models. \( \zeta(\sigma, r) \in R \) represents semantic constraints, including the correlation between scattering points at different locations. For example, whether two scattering points at different locations belong to the same target or the same type of terrain, and whether two scattering points at different locations have continuity in three-dimensional space or not. \( r \in F \) represents constraints on the positions of scatterers and is primarily expressed in the form of spatial structure primitive functions. \( \sigma(L, \alpha, \gamma, \varphi) \in D \) is the restrictions of scattering mechanism.

(2) Computation Model

Based on the traditional SAR three-dimensional imaging representation model, existing computational models mainly focus on pure signal processing. As mentioned earlier, the increment of SAR microwave vision three-dimensional imaging lies in extracting more prior information from the echo signal and SAR images. To achieve this goal, this paper proposes an innovative computation model that integrates the extraction of three-dimensional information.

Specifically, based on the aforementioned representation model, this paper introduces two types of computational models. One type is based on sparse three-dimensional imaging solution models, integrating the information obtained from microwave vision into the model in the form of explicit constraints. This includes semantic regularization models, spatial deterministic constraint models, spatial statistical constraint models, joint constraints of multi-dimensional observations, etc. Another type is built upon neural network models, training the network to learn microwave vision prior information, thus achieving the implicit feedback of microwave vision constraints.
Generally, the first type of model is based on compressed sensing theory, with additional explicit constraints. Taking the semantic regularization constraint model as an example, semantic regularization constraints include but are not limited to various types of neighborhood consistency under semantic segmentation, such as variational regularization, phase gradient, etc. Neighborhood continuity constraints, such as local elevation continuity and intra-class correlation after segmentation, are also considered. These constraints mainly take into account various similarity measures between pixels of the same type after semantic segmentation and use them as constraint conditions to improve the three-dimensional imaging quality. The statistical and deterministic constraints of three-dimensional positions directly infer the spatial distribution of scatterers from the perspective of three-dimensional space distribution, using the aforementioned prior information to constrain the solution space.

The second type of model is based on neural networks to achieve the implicit feedback of microwave visual information. It combines sparse unfolding networks and deep learning networks to construct a heterogeneous neural network. On the one hand, it utilizes the network to extract microwave visual prior information. On the other hand, it trains to learn the feedback of prior information, achieving adaptive optimization of network parameters. Ultimately, it aims to combine the interpretability of signal processing with the intelligent extraction of microwave visual information.

Detailed explanations of these different computational models will be provided in the following section on research progress. Therefore, this section provides a brief overview.

(3) Processing paradigm

In Section Two, it has been mentioned that the new and old theories exhibit significant differences in their processing approaches. The figure below provides a visual comparison for better understanding. This figure illustrates a direct comparison between the two theories, highlighting their notable distinctions in processing methods.

![Figure 4: Paradigm of Processing](image)

The traditional method is a forward open-loop process, corresponding to the lower part of the diagram where the red dashed box is omitted. It mainly involves processing the 2D SAR image to obtain the final 3D result. In contrast, the proposed method in this paper adds the step of extracting 3D information by
microwave vision. As shown in the diagram, 3D information is extracted from both 2D and 3D images, including but not limited to visual semantic information and scattering mechanism information. These pieces of information, based on multi-incident-angle observations, provide prior knowledge about the spatial distribution of scattering points, anticipating the location of the scattering points. Overall, the method proposed in this paper adopts a robust estimation-based iterative processing approach, progressively improving the 3D imaging results.

(4) Evaluation criteria

To address the existing challenges of inconsistent evaluation criteria and the absence of a standardized assessment system for SAR 3D imaging methods, we have developed a comprehensive evaluation framework. This framework is designed to assess the effectiveness of 3D imaging applications from multiple perspectives, aiming at the promotion of SAR 3D imaging.

Generally, the evaluation system mainly considers the following key aspects. The first is spatial accuracy, which is used to examine the deviation of reconstructed scatterers in 3D imaging concerning the actual positions of the targets. The second is about scattering coefficient estimation. For this purpose, we mainly assess the precision of estimating target scattering coefficients, including amplitude, phase, and polarization characteristics. The third is the unwrapping capability, which aims to evaluate the resolution of 3D imaging by gauging its ability to unwrap and discriminate overlapping structures, addressing the most important challenge discussed earlier. Finally, we consider the overall performance, which mainly incorporate success rates in 3D imaging and 3D image entropy, constructing comprehensive evaluation metrics for a quantitative and integrated assessment of overall performance.

(a) Three-dimensional position accuracy

Regarding three-dimensional position accuracy, evaluations are primarily conducted from three perspectives. The accuracy of the position of scatterers focuses on the third-dimensional precision of individual scatterers, while the accuracy and completeness of the three-dimensional point cloud primarily assess the overall accuracy of the target point cloud.

The definition of accuracy of the position of scatterers is as follows. Taking two overlapped scatterers as example, the position accuracy of scatterers is given by the following formula:

$$\Delta s = \sqrt{\frac{1}{2N} \sum_{n=1}^{N} [(\hat{s}_{n1} - s_{n1})^2 + (\hat{s}_{n2} - s_{n2})^2]}$$

(3)

In which $\hat{s}_{n1}$ and $\hat{s}_{n2}$ are the position estimations of two scatterers, while $s_{n1}$ and $s_{n2}$ are the true position. It is worth mentioning that $\hat{s}_{n1}$, $\hat{s}_{n2}$, $s_{n1}$, $s_{n2}$ are all values normalized by Rayleigh resolution.

The accuracy and completeness are defined as follows[22].

$$A(\mathcal{R}) = \frac{1}{N_{\mathcal{R}}} \sum_{i=1}^{N_{\mathcal{R}}} \text{mindist} \{\mathcal{R}(i), \Gamma(j)\}$$

(4)

$$C(\mathcal{R}) = \frac{1}{N_{\Gamma}} \sum_{j=1}^{N_{\Gamma}} \text{mindist} \{\Gamma(i), \mathcal{R}(j)\}$$

(5)

In above equations, $\mathcal{R}(i)$ denotes the reconstructed 3D point cloud while $\Gamma(j)$ is the high-precision reference ground truth. $N_{\mathcal{R}}$ is the number of scatterers in reconstructed point cloud while $N_{\Gamma}$ is that for ground truth. $\text{dist}\{a,b\}$ is the distance between two scatterers. The above metrics can be quantitatively evaluated in simulations or cooperative target experiments with known positions to assess the performance of different algorithms in 3D reconstruction.

(b) Scattering coefficients accuracy

The accuracy of scattering coefficients mainly focuses on the amplitude and phase information of each
scatterer after 3D reconstruction. These details are crucial for target recognition based on 3D point clouds and for four-dimensional and five-dimensional imaging processes that consider deformations. We propose three quantitative metrics to assess the accuracy of the scattering coefficients. The first is the accuracy of scattering intensity estimation.

\[
\Delta|\sigma| (\text{dB}) = 20 \cdot \log_{10} \left( 1 + \frac{1}{2N} \sum_{n=1}^{N} \left( \left| \frac{\sigma_n}{\sigma_{n1}} - 1 \right|^2 + \left| \frac{\sigma_n}{\sigma_{n2}} - 1 \right|^2 \right) \right)
\]  

In equation (6), \(|\sigma_{n1}|, |\sigma_{n2}|\) are the reference intensity of the two overlapped scatterers. \(|\sigma_n|, |\sigma_n|\) are the estimated intensity of the two overlapped scatterers.

The accuracy of phase estimation is defined as follows:

\[
\Delta \varphi (\text{deg}) = \frac{1}{2N} \sum_{n=1}^{N} \left( |\phi_{n1} - \phi_{n1}|^2 + |\phi_{n2} - \phi_{n2}|^2 \right)
\]  

In equation (7), \(\phi_{n1}\) and \(\phi_{n2}\) are the estimated phase of the two overlapped scatterers. \(\varphi_{n1}\) and \(\varphi_{n2}\) are the ground truth.

The output signal to noise level of 3D imaging result comprehensively considers the accuracy of both scattering intensity and phase estimation. Taking two overlapping points as an example, the output SNR level is as follows:

\[
O(\text{dB}) = 10 \cdot \log_{10} \left( \frac{1}{\sqrt{\sum_{n=1}^{N} ||\sigma_n||^2}} \right)
\]  

\(\sigma_n = (\sigma_{n1}, \sigma_{n2})\) is a complex vector and each element stands for the complex scattering coefficient. \(\|\cdot\|\) is the Euclidean norm.

(c) 3D imaging resolution

As the two-dimensional SAR image achieves range and azimuth resolution through pulse compression and azimuth synthetic aperture, the three-dimensional imaging resolution here primarily focuses on the resolution capability in the third dimension. It is defined as the distance between overlapping points where the success rate is greater than 50%, considering a given signal-to-noise ratio. This can serve as an indicator for assessing the third-dimensional resolution.

(d) Comprehensive evaluation

The above three dimensions have been analyzed from multiple perspectives. In order to comprehensively evaluate the performance of three-dimensional imaging, this paper summarizes the comprehensive evaluation indicators for three-dimensional imaging, namely the success rate of three-dimensional imaging and the three-dimensional image entropy.

The success rate comprehensively considers the position accuracy of scatterers and the signal-to-noise ratio of the imaging output. In simulations, the success rate is calculated by statistically determining the proportion of points with position accuracy and output signal-to-noise ratio exceeding a predefined threshold among the total points.

As for the 3D entropy, it quantifies the amplitude of pixel values after three-dimensional imaging, usually 8-bit grayscale quantization, then calculate the grayscale histogram and compute the entropy. The formula is as follows[23]:

\[
P(i,j) = \frac{i(j)}{N_x \times N_y \times N_z}
\]

\[
\text{ENT} = - \sum_{i=0}^{255} \sum_{j=0}^{255} P(i,j) \cdot \ln(P(i,j))
\]  

The three-dimensional image entropy is an extension of two-dimensional image entropy into three-dimensional space, which can indirectly reflect the sharpness of three-dimensional image. The sharper the three-dimensional image, the smaller the value. It provides an evaluation metric consistent with human visual
These aspects collectively contribute to a thorough evaluation of the effectiveness and reliability of SAR 3D imaging methodologies.

4. Recent progress

The SAR Microwave Vision 3D Imaging theoretical approach is not only innovative in 3D imaging processing methods but also encompasses both hardware system design and development, as well as information extraction and processing. In summary, it involves a balanced consideration of both hardware and software aspects. Therefore, this chapter will provide an overview of the progress made by our team in the field of microwave vision SAR 3D imaging, covering theoretical framework, SAR microwave vision 3D imaging algorithms, our self-developed SAR microwave vision 3D imaging system and dataset.

(1) Theoretical framework and fundamental imaging model

In 2019, we systematically review the development of SAR 3D imaging technology, conducting an in-depth analysis of the characteristics of existing SAR 3D imaging techniques. It identifies the untapped 3D information inherent in SAR echoes and images, and is the first time that we introduce the new concept and approach of SAR microwave vision 3D imaging, which is a combination of microwave scattering mechanisms and image visual semantics. This fusion establishes the theoretical framework and methodology for SAR microwave vision 3D imaging, aiming to achieve efficient and cost-effective SAR 3D imaging. We first emphasized the concepts, objectives, and key scientific issues of SAR microwave vision 3D imaging, presenting initial technical approaches that offer novel perspectives for SAR 3D imaging.

After concept proposal, we established an accurate SAR 3D imaging model. The planar-wavefront-based TomoSAR imaging model introduces significant approximation error when applied to the low-altitude airborne platform. This will lead to significant height error of reconstructed infrastructures. In order to circumvent the aforementioned problem, we propose spherical-wave-front models with the exact form or an approximate form of the slant range formula[24]. This model is strictly built upon the spherical wave propagation model of electromagnetic waves, applicable to various application scenarios, whether in the far-field of near-field case. The proposed model has been adopted in our UAV based array InSAR experiment, demonstrating its efficacy. In contrast to the imaging model based on the conventional planar...
wave-front approximation, the reconstructed point cloud of a 69.2-meter-high building introduced a height error of approximately 10 meters due to the model’s limitations. However, utilizing our proposed precise model eliminates errors introduced by the model itself, resulting in the extracted 3D model consistent with the actual target. The following figures depict the diagrammatic sketch of the proposed model and comparison between results by classical and proposed model. It is worth mentioning that this model transforms the third dimension from elevation to off-nadir angle, which provides a new research perspective for subsequent studies.

(a)  
(b)  
(c)  

Figure 6: Spherical SAR 3D imaging model and comparison between different models. (a) Planar wavefront model. (b) Spherical wavefront model. (c) Reconstruction results by different models.

Except the approximation error induced by planar-wave-front model, in the airborne near-field case, there will exists channel migration among all channels. Existing methods for 3D imaging in TomoSAR often assume that the layover remain consistent across different channels. In low-altitude airborne scenarios, the variability of the range from the same layover to different channels becomes significant and cannot be ignored even after co-registration. Thus, we study on the phenomenon of channel migration (CM) in sparse TomoSAR under low-altitude conditions. We derive a model for the differential range, which describes the variation in range from a specific target to different antennas[25]. To address the CM issue, we propose the application of the Keystone transform (KT) along the array axis, effectively correcting for the channel migration. We can subsequently apply compressed sensing-based 3D imaging methods to obtain accurate and reliable 3D reconstruction results. To validate the effectiveness, we conduct experiments using both simulated data and real data acquired by the MV3DSAR system, a drone-borne TomoSAR system which will introduced in the following. The results demonstrate that our proposed method offers a more precise solution to the challenging problem of reconstructing low-altitude sparse TomoSAR data. Overall, our study is the first time to address channel migration and contributes to the advancement of 3D imaging in urban environments.
After imaging model construction, we conducted a series of studies regarding SAR microwave vision 3D imaging algorithms, which address high accuracy, combination with geometric semantics, combination with microwave scattering mechanisms and so on. Details are as follows.

In the classic compressed sensing algorithm process, it is necessary to divide continuous elevation directions into fixed grids and assume that the target is exactly on the divided grids. Nevertheless, this assumption is often difficult to establish, leading to off-grid effect, which is currently rarely discussed in TomoSAR. We discussed the TomoSAR observation model with off-grid targets, and proposed a solution model that uses an additive perturbation term to correct the impact of target deviation from the grid. On this basis, a combination of local optimization algorithm and L1 norm minimization is introduced to solve the proposed off-grid TomoSAR model[26]. The following figures illustrate the comparison between the proposed algorithm and the classical method based on L1 norm minimization both on simulation data and airborne array interference SAR real flight data. It turns out that for the off-grid target, the proposed method can obtain more precise position, amplitude and phase solution results, proving the superiority of the method.
Furthermore, the TomoSAR model in sparse scenes can be regarded as a single snapshot line spectrum estimation problem, while the existing methods are either limited in accuracy due to the off-grid effect or high computational complexity. We proved that its least square objective function under the condition of a single snapshot is analytical and convex within the neighborhood of the true value[27]. Gradient descent algorithm can be adopted to solve the above function iteratively and effectively. Numerical simulations for evenly distributed antenna array case show that the proposed GDLS algorithm avoids the off-grid problem with a minor amount of computation, improving the estimation accuracy at a lower cost. The figures show a comparison between the proposed method and the classical CS algorithm. We projected the estimating point clouds for real data to the range elevation plane. Our method exhibits higher estimation accuracy and stronger resolution capability, providing a point cloud with better performance without off-grid errors.

![Figure 8](image)  
Figure 8: Off-grid TomoSAR imaging method. (a) Magnitude estimation performance. (b) Elevation estimation performance.

Although the GDLS algorithm can provide better point cloud, treating the elevation resolution as a line spectrum estimation problem poses challenges for traditional super-resolution spectrum estimation algorithms, which necessitate multiple snapshots and uncorrelated targets. Meanwhile, compressed sensing (CS)-based methods, widely used in modern TomoSAR imaging, encounter the gridding mismatch effect, significantly compromising imaging performance. In addressing these issues, we introduced an enhanced

![Figure 9](image)  
Figure 9: Side view of reconstruction result in radar geometry by traditional method and GDLS. (a) Traditional method. (b) Proposed GDLS method.
fast atomic norm minimization (ANM) algorithm for TomoSAR elevation focusing, namely the IVDST-ANM algorithm. This approach mitigates the substantial computational complexity associated with conventional time-consuming semi-positive definite programming (SDP) by employing the iterative Vandermonde decomposition and shrinkage-thresholding (IVDST) method. Importantly, it preserves the advantages of ANM, including gridless imaging and the capability of single snapshot recovery. The effectiveness of the proposed method is demonstrated through numerical performance evaluations using simulated data and the reconstruction results of an urban area from the SARMV3D-Imaging 1.0 dataset, as illustrated in the subsequent figures[28].

![Image](https://engine.scichina.com/doi/10.1360/nso/20240009)

**Figure 10:** Performance comparison of IVDST-ANM and other state-of-art technology, and reconstruction result of urban areas. (a) RMSE versus SNR by different method. (b) Reconstructed point cloud by IVDST-ANM.

(b) SAR visual semantics extraction

Buildings are one of the typical targets in urban remote sensing images, and the extraction of buildings from remote sensing images finds extensive applications. Semantic segmentation of buildings is a crucial task and prerequisite for realizing SAR microwave vision 3D imaging. In the optical remote sensing domain, significant progress has been made in the semantic segmentation of buildings. However, in the Synthetic Aperture Radar (SAR) domain, challenges persist due to SAR imaging characteristics such as layover, foreshortening, and shadows. We addressed these challenges by proposing a novel complex-valued convolutional and multi-feature fusion network (CVCMMFF Net) for building semantic segmentation in SAR images[29]. To make better use of SAR image information, equivalent complex-valued convolution layer and equivalent complex-valued pooling layer were incorporated into the network to accommodate the input complex data. Considering the diverse shapes and sizes of buildings, an atrous spatial pyramid pooling (ASPP) method was introduced for multiscale feature fusion. The network also integrated a multichannel feature fusion module to jointly process the master SLC image, slave SLC image, and interferometric phase image. Cross entropy (CE) was employed to quantify the disparities between the actual distribution and the predicted model distribution. The loss was determined by the sum of cross entropy for all pixels. The network was trained on a simulated dataset and demonstrated favorable inference results on measured datasets.
Except extracting the building structures from the two-dimensional SAR images, we have explored the chance to obtain the 3D clues from the point cloud\cite{30}. Due to issues such as the scattering characteristics of ground objects, system noise, baseline measurement errors, etc., traditional TomoSAR inversion methods often encounter the following problems in the obtained point clouds. The first is sparsity or the absence of spatial points corresponding to partial structures, making it insufficient to represent their complete structures. The second is deviations from real structures or the presence of a significant amount of noise points. To address these challenges, W. Wang and collaborators propose a progressive building facade detection algorithm based on structural priors. The algorithm initially projects the initial array InSAR spatial points onto the ground to generate connected regions corresponding to the building facade. It then progressively detects potential line segments within each connected region guided by structural priors, subsequently generating the corresponding building facade based on the detected line segments and their associated spatial points. During this process, the detection space of line segments corresponding to the current connected region is constructed based on the detected line segments in its adjacent connected regions, effectively ensuring overall efficiency and reliability. Experimental results demonstrate that the proposed algorithm can rapidly detect numerous reliable building facades from massive, noisy array InSAR spatial points, effectively overcoming the drawbacks of traditional multi-model fitting algorithms in terms of low efficiency and reliability.

Figure 11: Network structure and segmentation result of CVCMFF Net.
(c) SAR 3D imaging algorithms with semantics constraints

In SAR image semantics, the most prominent contribution to 3D imaging is the geometric structural information of the targets. This approach relies on the premise that SAR images and 3D point clouds are mappings of the real-world physical environment. The geometric structures present in two-dimensional images exhibit similar representations in three-dimensional point clouds.

Based on above analysis, we propose a three-dimensional imaging method based on deterministic constraints using spatial structural elements. This approach primarily addresses the challenge of high clutter in current SAR three-dimensional imaging under low signal-to-noise ratio conditions. By leveraging geometric structural elements provided by semantic information in SAR images, the method constrains the positions of scatterers in the reconstructed three-dimensional point cloud. The approach relies on SAR three-dimensional imaging geometric models and extracted target geometric structures. It adaptively constrains the solution space based on information such as image signal-to-noise ratio, effectively suppressing clutter points. Experimental data validation demonstrates that this method significantly reduces clutter points, enhancing the quality of three-dimensional imaging[31].

Figure 13: Geometric constraint SAR 3D imaging. (a) Geometrical structures from SAR images. (b) Co-registered SAR and lidar point cloud.
Aforementioned is the deterministic approach for geometric constraint SAR 3D imaging. We also explored a three-dimensional imaging method based on statistical constraints using spatial structural elements, considering both large and small scales. This approach involves probabilistic modeling of the spatial distribution of scatter points in the three-dimensional point cloud based on geometric semantic information, offering greater flexibility compared to deterministic constraints. For large-scale processing targeting linear and planar structures, scatter points corresponding to these structures are modeled using a generalized Gaussian distribution, solved through maximum likelihood estimation. In small-scale processing, each pixel and its neighborhood are considered. Addressing the limitation of pixel-wise processing, which neglects spatial adjacency, this method employs a local Gaussian Markov random field to model the spatial distribution of scatter points in the neighborhood. Adaptive constraint solutions are achieved based on signal-to-noise ratio and the correlation between observation vectors. These methods produce high-quality three-dimensional imaging results with limited observations, effectively incorporating geometric semantic information from SAR images. Inspired by this approach, we further expand the usage of geometric constraints in stereo SAR 3D imaging. A novel solution for stereo three-dimensional localization combined with geometric semantic constraints is proposed and validated on spaceborne SAR data.

Figure 14: geometric constraint SAR 3D imaging based on statistical model. (a) Large-scale case. (b) Small-scale case.

In addition, we propose a novel tomographic SAR (Synthetic Aperture Radar) 3D imaging method based on sparse and low-rank structures. Traditional tomographic SAR imaging, relying on compressed sensing, only sparsely represents and reconstructs the elevation dimension for given azimuth-range cells. Our method recognizes the similar layout distribution in urban and forest areas, where targets in adjacent azimuth-range cells exhibit strong correlation in their elevation distribution. By incorporating the Karhunen-Loeve transform, our approach captures the low-rank characteristics of elevation distribution.
across neighboring azimuth-range cells. This results in a combined sparse and low-rank structured model for tomographic SAR imaging of target areas. We employ the ADMM (Alternating Direction Method of Multipliers) algorithm to solve this imaging model. The complex original optimization problem is decomposed into simpler sub-problems, which are then iteratively projected through alternating optimization variables, yielding the tomographic SAR imaging result. This method enhances reconstruction accuracy under scenarios of low overflights or channel counts, demonstrating superior imaging performance[32].

Except the urban areas, we explored the application of proposed method in mountainous areas. The predominant focus of TomoSAR in 3D reconstruction has been on low-lying targets, with limited literature addressing the reconstruction of 3D mountainous terrain. Surveying high mountain areas remains challenging due to the layover phenomenon. Therefore, conducting research on 3D mountain reconstruction using airborne array TomoSAR holds significant value. Nevertheless, the original TomoSAR mountain point cloud encounters elevation ambiguity issues. Additionally, in mountains with intricate terrain, points at different elevations may intersect, intensifying the complexity of the problem. We introduce a novel method for resolving elevation ambiguity based on the continuity of mountain structures. Initially, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Gaussian Mixture Model (GMM) are combined for point cloud segmentation[33]. DBSCAN ensures coarse segmentation based on density, while GMM enables fine segmentation of abnormal categories resulting from intersections. Subsequently, the segmentation results are reorganized along the elevation direction to reconstruct all potential point clouds. Finally, the real point cloud is automatically extracted within the constraints of boundary and elevation continuity. The efficacy of the proposed method showcases its potential for addressing challenges in 3D mountain reconstruction.

![Figure 15: Reconstructed point cloud of mountainous areas based on proposed method.](image)

(d) SAR 3D imaging algorithm with scattering mechanism

So far, 3D imaging methods have improved target localization accuracy but lack of the ability to distinguish various scattering mechanisms (SMs). In order to address this issue, we conducted a series of experiments using multi-baseline polarimetric coherence optimization and Pauli decomposition methods. They were used to establish digital elevation models (DEM) of the observed areas and preliminarily analyze SMs corresponding to real targets. Furthermore, we proposed a novel method based on maximizing polarimetric energy to enhance the 3D identification of targets. This method projects the original signal vector onto one component by rotating the polarization bases, ensuring optimal utilization for total
polarimetric power. The projected signal vectors, derived from the above methods, are then treated as the input for subsequent height inversion. Eventually, we applied these methods in two sets of array InSAR datasets from Fudan University, Shanghai, and Suzhou Aerospace Information Research Institute, Jiangsu. The results demonstrate that the 3D point clouds rendered by polarized parameters through Pauli decomposition and the new method display exceptional potential in retaining detailed structural information and distinguishing different scattering mechanisms. This improvement provides a research foundation for fully polarized array InSAR and TomoSAR.

![Flowchart of the Pol-array-InSAR](image)

**Figure 16:** Pol-array-InSAR processing model and results.

(e) Neural network-based algorithms

The multi-channel circular trajectory SAR enables omnidirectional observation of targets, addressing the shadow issues associated with conventional TomoSAR three-dimensional imaging along straight trajectories. However, conventional processing methods such as compressed sensing (CS) or back projection (BP) require a sufficiently large number of observations, leading to significant computational overhead in three-dimensional imaging. To tackle these challenges, we proposed a sparse-aspects-completion network (SACNet) based on the Generative Adversarial Network (GAN) principle[34]. The approach involves selecting a sparse subset of aspect data from the acquired omnidirectional dataset and employing traditional methods for three-dimensional imaging in each aspect, resulting in an incomplete point cloud of the target. Subsequently, these partial point clouds from selected aspects are input into SACNet, which outputs a complete three-dimensional point cloud of the target. The innovative approach involves training a GAN to learn the underlying 3D structure of vehicles and subsequently generating detailed reconstructions. The technique effectively addresses the challenge of sparse data, providing a promising solution for accurate 3D vehicle reconstruction in SAR imagery.

https://mc03.manuscriptcentral.com/nsopen
Similar to the InSAR technique, phase information is of crucial importance to SAR 3D imaging. The quality of interferometric phase significantly influences the effectiveness of the reconstruction. Therefore, interferometric phase filtering is crucial. Our team proposed an unsupervised convolutional neural network (CNN)-based method for denoising multi-channel interferometric phase data. The focus of the study is on addressing the challenges in this area. The method utilizes a CNN to automatically learn and remove noise from multi-channel interferometric phase data, achieving denoising without the need for labeled training samples. This approach demonstrates effective denoising results, enhancing the accuracy of three-dimensional reconstruction in TomoSAR[35].

CS-based algorithms have been incorporated into TomoSAR to leverage their super-resolution potential with limited samples. However, traditional CS-based methods face drawbacks, such as weak noise resistance, high computational complexity, and intricate parameter fine-tuning. In pursuit of efficient
TomoSAR imaging, we present a novel sparse unfolding network, named Adaptive Threshold ALISTA-based Sparse Imaging Network (ATASI-Net), built upon the architecture of the analytic learned iterative shrinkage thresholding algorithm (ALISTA) with adaptive threshold[36]. The weight matrix in each layer of ATASI-Net is pre-computed as the solution of an off-line optimization problem, simplifying the training stage by requiring the learning of only two scalar parameters from data. Additionally, an adaptive threshold is introduced for each azimuth-range pixel, allowing not only layer-varied but also element-wise threshold shrinkage. The following figure illustrates the framework of the proposed ATASI-Net and the data flow of each module in the $k$-th layer, experimental comparison results with conventional algorithms, and visual outcomes of semantic feedback from SAR images.

![Diagram of ATASI-Net](image)

Figure 19: Details and performance of ATASI-Net. (a) Overall architecture of ATASI-Net and the data flow of each module. (b) Point cloud comparison results based on the traditional CS algorithm and the proposed ATASI-Net model. (c) Segmentation results after threshold visualization.

### (3) System and dataset

To validate the effectiveness of the proposed theory, we overcame a series of key challenges and
developed a lightweight unmanned aerial vehicle (UAV)-borne SAR microwave vision 3D imaging system. Extensive data acquisition experiments were conducted, and a comprehensive 3D imaging dataset was established. The following provides a detailed introduction.

The theory proposed aims to reduce the requirement for the number of observations in 3D imaging to less than or equal to five, thereby lowering system complexity or data acquisition span. This enables efficient 3D imaging and promotes the widespread application of SAR 3D imaging technology. We have designed and developed a small UAV-borne interferometric SAR system, named the Microwave Vision 3D SAR (MV3D SAR). This system is utilized for data acquisition, technical validation, and the construction of a SAR microwave vision 3D imaging dataset. The system features lightweight design, full polarization, and flexible baseline configuration. During the development, we addressed challenges related to nonlinear error processing, high-precision motion compensation, and multi-channel consistency calibration. Utilizing measured data obtained from this system, we designed a tailored data processing workflow, including 2D imaging, system error calibration and compensation, and 3D imaging. The second part involves estimating and compensating for system channel delay errors, multi-channel baseline errors, and inter-channel amplitude and phase errors, providing robust support for 3D imaging processing and validating the 3D imaging capability of the system[37].

Figure 20: System configuration and multi-aspects SAR images. (a) The MV3DSAR system. (b) Optical image of the scene and multi-aspects SAR images.

As is well known, datasets play a crucial role in the development for a certain field. The progress of SAR 3D imaging technology has long been constrained by insufficient datasets. This inadequacy manifests in two aspects. Firstly, there is a shortage of multi-channel SAR data. Secondly, there is a lack of ground truth to assess the effectiveness and accuracy of various algorithms. Therefore, leveraging the accumulated data from the previous manned platform and utilizing the data acquired by the aforementioned UAV-borne system in regions such as Tianjin and Jiangsu, we have constructed the SAR Microwave Vision 3D Imaging dataset, known as the SARMV3D dataset. This dataset addresses the scarcity of multi-channel SAR data and provides reference ground truth, enabling the evaluation of algorithmic performance and accuracy[38].
The SARMV3D dataset comprises two auxiliary datasets for 3D electromagnetic scattering mechanism research and image visual semantics perception, along with a comprehensive dataset for SAR 3D imaging research based on microwave vision. Details of these three parts are as follows. The 3D electromagnetic scattering mechanism research dataset is used for studying and validating the correctness of electromagnetic modeling, characterization, and inversion estimation of 3D feature parameters for typical scattering mechanisms and their combinations. It mainly includes simulated, anechoic chamber measurements, and outdoor field measurement data for typical scattering mechanisms and their combinations. The second is the SAR image visual semantics perception dataset. It is designed for researching and validating the effectiveness and correctness of methods for extracting and identifying 3D primitives, structures, and targets from SAR images. This subset mainly comprises high-resolution SAR images and annotated results for typical buildings and structural prior information. Last but not least is the microwave vision-based SAR 3D imaging dataset. It is the primary comprehensive dataset includes urban scenes, complex terrains, and typical targets. For each type, single-look complex images (SLC) obtained from TomoSAR or array InSAR system, corresponding imaging parameter files, 3D modeling data obtained from optical photography, and high-precision 3D point cloud data obtained from lidar, are all included.

The SARMV3D dataset has three versions, i.e., Version 1.0 released in 2021 is mainly based on data from the Aerospace Information Research Institute's array InSAR system and high-resolution SAR images from the Gaofen-3 satellite. Version 2.0 released in 2022 includes small UAV-based array InSAR data obtained in the Tianjin Lingang CBD using the MV3D SAR system, along with synchronized optical oblique photography and lidar point cloud. Version 3.0 released in 2024 includes data obtained at Suzhou aerospace information research institute using the full-polarization MV3D SAR system. Specifically, dataset of Yuncheng and Emei consists of SAR images from single aspect. Dataset of Tianjin consists of single polarization SLC images from eight aspects, while that of Suzhou consists of full polarization SLC images from eight aspects. These datasets have been made available on the website of Journal of radars, attracting hundreds of research institutions and thousands of researchers globally to download and engage in related studies. This dataset contributes significantly to alleviating the scarcity of SAR 3D imaging datasets and promoting advancements in the field. For more details on the dataset, refer to the paper available on the mentioned website.

5. Anticipated development trends
As mentioned above, we have made significant progress in SAR microwave vision 3D imaging. However, being a novel theory, there are still numerous aspects awaiting exploration. To this end, building upon the existing research achievements, we have outlined the anticipated future directions in this domain. Consistent with the previous discussion, we categorize these directions into three aspects: theoretical framework, imaging algorithm design, and hardware systems with datasets.

(1) Theoretical framework

Regarding theoretical framework, there are mainly two aspects. The first is to further refine the concept of holographic SAR, exploring more detailed theoretical models to comprehensively and accurately characterize the microwave properties of targets. Secondly, cross-disciplinary collaboration must be addressed. Increased collaboration among radar signal processing, electromagnetics, computer vision, and related fields should be conducted to leverage expertise from diverse disciplines for more holistic advancements in SAR microwave vision 3D imaging.

(2) SAR 3D imaging algorithm

As for 3D imaging algorithms, integration with advanced technologies is the of top priority. Integration with emerging technologies such as deep learning to enhance capabilities in information extraction, feature recognition, and 3D reconstruction is promising. Besides, we should further investigate unsupervised and self-supervised learning methods to reduce reliance on extensively labeled data, enhancing the model's generalization since SAR dataset is not as abundant as optical datasets. Generally, it is necessary to realize continuous refinement and advancement of algorithms and models for microwave vision-based SAR 3D imaging, focusing on improving accuracy, efficiency, and adaptability across diverse scenarios.

(3) SAR systems and dataset

First, advancements in hardware systems should be emphasized, including the development of more sophisticated SAR systems, improved sensors, and enhanced data acquisition technologies to further elevate imaging quality. Second, since the main purpose of SAR microwave vision 3D imaging is to realize high-quality low-cost reconstruction, real-time and onboard processing should be considered based on years of research accumulations. In the future, exploration and development of real-time and onboard processing capabilities should be conducted, enabling timely 3D imaging results for time sensitive applications.

As for dataset construction, it is necessary to expand existing datasets and create benchmark datasets for comprehensive evaluation and benchmarking of different microwave vision-based SAR 3D imaging approaches. In order to promote holographic SAR research, we will construct more diverse imaging datasets containing multisource heterogeneous data, covering a wider range of scenes and targets to improve the theory's robustness and generalization.

These anticipated developments collectively aim to propel microwave vision-based SAR 3D imaging into new frontiers, expanding its applications and advancing the state-of-the-art in the field.

6. Conclusion

SAR three-dimensional imaging, compared to two-dimensional images, provides richer and more diverse information, making it a recent research focus in the SAR imaging field. In response, SAR microwave vision three-dimensional imaging, as a new form emerging after two decades of three-dimensional imaging development, integrates multiple disciplines, achieving high efficiency, low cost, and has the potential to further practicalize SAR three-dimensional imaging. Traditional SAR three-dimensional imaging relies solely on signal processing techniques, while SAR microwave vision three-dimensional imaging introduces the concept of microwave vision, incorporating an intelligent theoretical framework for understanding microwave radar images. This results in an intelligent and efficient three-dimensional imaging theory.
This paper provides a systematic review of SAR microwave vision three-dimensional imaging. It starts by scientifically defining its conceptual framework and elaborating on core concepts. The theoretical framework of SAR microwave vision three-dimensional imaging includes representation models, computational models, processing paradigms, and evaluation systems. Building on the concept of holographic SAR imaging, the paper proposes a representation model for microwave vision three-dimensional imaging. It introduces an integrated model for intelligent processing and information extraction, evolving from traditional unidirectional processing models to new closed-loop feedback processing models. From the perspective of practical application promotion, a comprehensive evaluation system for three-dimensional imaging performance is constructed.

The author's team conducted extensive research on the above content, accumulating rich achievements over several years in imaging theory, algorithms, system development, and dataset construction. Notably, the constructed dataset is the most comprehensive in the field of SAR three-dimensional imaging in China. The team's theoretical algorithm research results were validated on this dataset, confirming the effectiveness of SAR microwave vision three-dimensional imaging methods. However, as an emerging development direction, SAR microwave vision three-dimensional imaging still faces foundational and key technological challenges, requiring further breakthroughs in basic theory, imaging algorithms, and system development. The key is to integrate practical application needs, further improve, and promote the development of relevant technologies.

Funding
This work was supported by the National Natural Science Foundation of China (No. 61991420, No. 61991421 and No. 61991424).

Author contributions

Conflict of interest
The authors declare that they have no conflict of interest.

Data availability
The original data are available from corresponding authors upon reasonable request.

References


