Article

A Novel Multimodal Fusion Framework Based on Point Cloud Registration for Near-Field 3D SAR Perception

Tianjiao Zeng [1], Wensi Zhang [2], Xu Zhan [2], Xiaowo Xu [2], Ziyang Liu [2], Baoyou Wang [2] and Xiaoling Zhang [2,*]

1   School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China; tzeng@uestc.edu.cn
2   School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 202311012221@std.uestc.edu.cn (W.Z.); zhanxu@std.uestc.edu.cn (X.Z.); xuxiaowo@std.uestc.edu.cn (X.X.); 202322010408@std.uestc.edu.cn (Z.L.); 202021010808@std.uestc.edu.cn (B.W.)
*   Correspondence: xlzhang@uestc.edu.cn

Abstract: This study introduces a pioneering multimodal fusion framework to enhance near-field 3D Synthetic Aperture Radar (SAR) imaging, crucial for applications like radar cross-section measurement and concealed object detection. Traditional near-field 3D SAR imaging struggles with issues like target–background confusion due to clutter and multipath interference, shape distortion from high sidelobes, and lack of color and texture information, all of which impede effective target recognition and scattering diagnosis. The proposed approach presents the first known application of multimodal fusion in near-field 3D SAR imaging, integrating LiDAR and optical camera data to overcome its inherent limitations. The framework comprises data preprocessing, point cloud registration, and data fusion, where registration between multi-sensor data is the core of effective integration. Recognizing the inadequacy of traditional registration methods in handling varying data formats, noise, and resolution differences, particularly between near-field 3D SAR and other sensors, this work introduces a novel three-stage registration process to effectively address these challenges. First, the approach designs a structure–intensity-constrained centroid distance detector, enabling key point extraction that reduces heterogeneity and accelerates the process. Second, a sample consensus initial alignment algorithm with SHOT features and geometric relationship constraints is proposed for enhanced coarse registration. Finally, the fine registration phase employs adaptive thresholding in the iterative closest point algorithm for precise and efficient data alignment. Both visual and quantitative analyses of measured data demonstrate the effectiveness of our method. The experimental results show significant improvements in registration accuracy and efficiency, laying the groundwork for future multimodal fusion advancements in near-field 3D SAR imaging.

Keywords: near-field 3D SAR; multimodal fusion; point cloud registration

1. Introduction

Near-field 3D synthetic aperture radar (SAR) imaging can obtain the three-dimensional electromagnetic scattering structure of observed targets and restore their spatial position information, which has become an important trend in the development of SAR [1–5]. In recent years, near-field 3D SAR imaging has been increasingly applied in concealed object detection and radar cross-section (RCS) measurement [6]. Owing to the capability of working under all-day and all-weather conditions, the near-field 3D SAR system is not only unaffected by environmental factors such as light and smoke, but also able to reconstruct items under clothing or within boxes [7]. It is suitable for deployment in airports, high-speed railways, and other occasions for security checks. Compared to microwave anechoic chamber measurement, near-field 3D SAR systems can perform RCS measurement on the target quickly, which is beneficial for radar stealth evaluation and scattering diagnosis [8].

However, near-field 3D SAR encounters several challenges. First, the clutter, multipath interference, and noise mixed in the images obscure target–background differentiation. Second, the presence of sidelobes results in a blurry shape and structure loss of the target, which affects the scattering diagnosis of specific parts in the target. Third, near-field SAR images are limited to capturing scattering intensity and do not provide color or texture information, which complicates the accurate categorization of targets. These limitations lower the quality of perception and hinder subsequent tasks like scattering diagnosis, detection, recognition, and interpretation.

Research into scene perception based on multi-sensor fusion has recently become a hot topic [9–11]. Multi-sensor fusion can integrate complementary multimodal data to make working conditions broader and obtain more informative fusion results. Existing work has fused 2D SAR images with optical, hyperspectral and infrared images to assist in SAR image interpretation [12–14], and been applied in fields such as remote sensing surveys and disaster detection. Yinghui Quan et al. [15] developed a multi-spectral and SAR image fusion method based on weighted median filtering and Gram–Schmidt transform to improve the classification accuracy of land cover. For multi-sensor 3D SAR fusion, Xiaolan Qiu et al. [16] imaged a building using the unmanned aerial microwave vision 3D SAR (MV3DSAR) experimental system and LiDAR, and demonstrated the fusion results of LiDAR point clouds and reconstructed interferometric SAR point clouds, but did not provide relevant registration and fusion methods. It can be seen that research on the fusion of near-field 3D SAR with other heterogeneous sensors is just beginning.

Common sensors include radar, LiDAR, and cameras. LiDAR detects targets using emitted lasers, which can accurately measure distance. The captured laser point cloud can accurately describe the geometric shape, structure, and size of the target. However, its operation is greatly affected by weather, and the laser attenuates severely in environments such as heavy rain, thick smoke, and fog [17]. Optical cameras capture visible light reflected on the surface of an object for imaging, which can obtain detailed information such as the color and texture of the object. The resolution of visible light images is high, which is more in line with human cognition. However, they are greatly affected by light during operation, resulting in poor imaging results at night [18]. Due to the strong penetration of electromagnetic waves, radar can work in harsh weather, but its imaging resolution is low and lacks details [19]. To improve the capabilities of near-field 3D SAR images in scattering diagnosis and detection, this study presents the first research on multimodal fusion with near-field 3D SAR, LiDAR, and optical camera. The interference in SAR images can be suppressed by utilizing LiDAR's precision in target localization and shape description, which helps scattering diagnosis. The color and texture information of optical images can aid in categorizing objects in near-field SAR images, enhancing the perception of a scene.

Multimodal sensing uses heterogeneous sensors to capture more comprehensive scene information, and effectively addresses the afore-mentioned deficiencies by aggregating multi-sensor data through fusion [20]. The key to achieving multi-sensor data fusion is to solve the problem of coordinate system alignment. That is to say, to find the relative pose relationship of different coordinate systems. Here, the pose refers to both the position and the orientation of a subject. Two commonly used methods are calibration and registration [21]. The calibration method not only requires the manual design of the calibration object, but the object also needs to be recalibrated after the relative pose of the sensor changes, which is not flexible enough [22]. Therefore, this study adopts point cloud registration to achieve multimodal data fusion for near-field 3D SAR perception.

The existing point cloud registration research mainly focuses on the problem of homogeneous point cloud registration or LiDAR–Camera point cloud registration, while there is no published research on point cloud registration methods for near-field 3D SAR and other sensors. In 2014, Furong Peng et al. [23] first analyzed the significant differences in point cloud density, sensor noise, scale, and occlusion in multi-sensor point cloud registration, and then proposed a two-stage registration algorithm. By utilizing coarse registration based on the ensemble of shape functions (ESF) descriptor and iterative clos-

est point (ICP) [24] fine registration, the registration of LiDAR point clouds and optical structure from motion (SFM) [25] reconstruction point clouds for street buildings was completed. In 2015, Nicolas Mellado et al. [26] proposed a method for registering LiDAR point clouds and optical multi-view stereo (MVS) reconstruction point clouds. This method first achieved scale-invariant matching though the growing least squares descriptor, and then used the random sample consistency (RANSAC) method [27] for spatial transformation. In 2016, Xiaoshui Huang et al. [28] improved on the work of Furong Peng et al. [23] by using an improved generative Gaussian mixture model in the fine registration stage to achieve the high-precision fusion of street view LiDAR and SFM point clouds. In 2017, Xiaoshui Huang et al. [29] applied graphs to describe the structures extracted from multi-sensor point clouds, and used an improved graph matching method with global geometric constraints to obtain the graph matching results. After that, RANSAC and ICP were used to refine and complete the registration fusion of SFM and Kinect point clouds. In 2021, Jie Li et al. [30] utilized a unified simplified expression of geometric elements in conformal geometry algebra to construct the matching relationship between points and spheres, obtaining a more accurate alignment of LiDAR and Kinect point clouds.

From the above research, we can infer that the ICP algorithm is currently the most widely used point cloud registration method [31]. However, the ICP algorithm has strict requirements for the initial pose of the two input point clouds, and it is easy to fall into local optima when there are significant differences in the initial pose. In order to provide a good initial pose for the ICP algorithm, coarse registration algorithms such as the RANSAC method and its variants are generally used for roughly aligning the input point clouds. Currently, multi-source point cloud registration mostly uses this coarse-to-fine registration method [32].

However, the different imaging mechanisms of multiple sensors also pose some challenges to multimodal fusion. Lahat et al. [33] identified the challenges in multimodal data fusion and divided them into two parts: the challenges caused by data collection and the challenges caused by the data source. In the fusion of SAR, LiDAR, and camera, these challenges are manifested as follows: (1) Data format differences—Near-field 3D SAR images are mainly obtained by imaging radar echoes using the back projection (BP) algorithm [34], which is expressed in voxels, while the LiDAR imaging result is the point cloud and the optical camera captures the 2D image. (2) Noise difference—There are clutter and background noise in 3D SAR images. The difference is that LiDAR point clouds and optical reconstructed point clouds have outliers. (3) Resolution difference—The frequency bands of the microwave, laser, and visible light used in SAR, LiDAR, and cameras gradually increase, resulting in the highest resolution of optical images, followed by LiDAR point clouds, and the lowest resolution of SAR images. Due to these challenges, existing point cloud registration methods cannot effectively select corresponding points, making it difficult to achieve efficient and high-precision multimodal data alignment for near-field 3D SAR.

Based on the current state of research, there are no detailed public research results specifically based on the field of 3D SAR, especially near-field 3D SAR, which holds significant application value in areas such as scattering diagnosis and perception. Moreover, the fusion of 3D SAR, LiDAR and camera presents its own unique challenges that are not suitably addressed with the current methods, which are primarily aimed at homogeneous point cloud fusion. Bearing this background in mind, and following the trend of multimodal sensing for 3D SAR, we have decided to conduct a preliminary study in this work.

To address existing challenges, this study develops a novel multimodal fusion framework for near-field 3D SAR, consisting of data preprocessing, point cloud registration, and data fusion. For preprocessing, 3D SAR images are converted into point clouds and optical point clouds are reconstructed using SFM, thus standardizing the data format. This is followed by noise removal and target feature extraction from the multimodal data. For registration, LiDAR point clouds, known for their precise positioning and shape accuracy, act as an intermediate bridge for SAR–LiDAR and LiDAR–Camera pairwise registration

to achieve the spatial alignment of all three sensors. The final fusion step integrates multimodal data of varying resolutions by adding optical color textures and SAR scattering intensity to the LiDAR point clouds.

The registration process introduces a three-stage multi-sensor point cloud registration method, comprising key point extraction, coarse registration, and fine registration. Initially, a centroid distance (CED) key point extraction method with dual constraints of geometric structure and intensity is used to extract key points from the point cloud. Next, the method employs a sample consensus initial alignment (SAC-IA) coarse registration method with mixed constraints of geometric triangulation and a signature of histogram of orientation (SHOT) feature to achieve the initial pose transformation. The final step, based on the initial pose transformation, applies an adaptive-thresholding ICP fine registration algorithm for precise pose adjustment. The method enhances registration efficiency by key point extraction, and eliminates point cloud heterogeneity and uses multiple constraint terms constructed based on prior knowledge to improve registration accuracy. Through the above point cloud registration methods, the proposed multimodal data fusion framework achieves LiDAR–SAR point cloud registration and LiDAR–Camera point cloud registration, respectively, to obtain aligned SAR–LiDAR–Camera three-sensor data. After that, the nearest neighbor search algorithm is used to remove the redundancy of SAR point clouds, and the multi-sensor point cloud fusion results are obtained. The experimental data were captured by our prototype hardware system, and the processing results demonstrate the fusion of near-field 3D SAR with LiDAR and optical cameras, while verifying the effectiveness of the proposed point cloud registration method and multimodal fusion framework.

Our main contributions are as follows:

- This work presents the first attempt to enhance the perception quality of near-field 3D SAR imaging from a multi-sensor data fusion perspective, uniquely combining near-field 3D SAR with LiDAR and optical camera data to address the inherent limitations;
- This work designs a multimodal fusion framework for effectively integrating data from near-field 3D-SAR, LiDAR, and a camera, which consists of three main components—data preprocessing, point cloud registration, and data fusion;
- This work introduces a novel three-stage registration algorithm tailored to overcome the heterogeneity issue across sensors. This algorithm includes—(1) a new key point extraction method that improves the CED algorithm with structure–intensity dual constraints, (2) an enhanced coarse registration technique that integrates geometric relationship and SHOT feature constraints into SAC-IA for improved initial alignment, and (3) an adaptive-thresholding ICP fine registration algorithm for precise fine registration;
- This work validates the proposed approach using data collected from our SAR–LiDAR–Camera prototype system. The experimental results demonstrate obvious improvements in registration accuracy and efficiency over existing methods. The quantitative and qualitative results underscore the effectiveness of our multi-modal fusion approach in overcoming the inherent limitations of near-field 3D-SAR imaging.

The rest of this paper is organized as follows: Section 2 provides a description regarding the materials adopted, including the system and the collected data. The specific framework for the fusion of data from SAR, LiDAR and the camera is presented in Section 3. Section 4 describes the experimental results and gives a discussion of the proposed framework. Finally, we summarize the paper in Section 5 and provide some prospects for future work.

2. Materials

The proposed framework is designed for near-field 3D SAR perception. The near-field 3D SAR is a type of radar imaging system that actively transmits electromagnetic waves to the observed target. These transmitted waves are often in the X band for applications like scattering diagnosis, and the W band for applications like person screening. The corresponding wavelength ranges from the level of cm to mm. Objects under these bands

present differences from human visual perception. For instance, some parts of the target might appear missing, as seen in the head of the aircraft models in Figure 1a,b. The resolution is also limited, making the grid on the surface of the satellite model appear ambiguous. Furthermore, the color of the radar image, which reflects the scattering intensity of the target, varies significantly from visual perception. These limitations make scattering diagnosis, detection, recognition, and interpretation challenging. Compared to radar, other sensors like LiDAR and cameras can supplement information. LiDAR is an active sensing method that uses a much higher frequency of electromagnetic wave and a much shorter wavelength, like 905 nm in our prototype system, achieving higher resolution as shown in Figure 2. The camera sensor is a passive sensing method that relies on the illuminated and reflected light on the object. The related electromagnetic wave lies in the spectrum of visible light, with a wavelength range between 380 and 700 nm. The resulting optical image provides color information, revealing the texture of the object in line with our visual perception, as shown in Figure 3. By fusing this additional information, the radar image (the near-field 3D SAR image) can be perceived more easily and comprehensibly. This relies on the accurate fusion framework detailed in the next section.

In the data capture system, the millimeter wave near-field array 3D SAR imaging system serves to obtain near-field 3D SAR images, the Spedal monocular camera captures multi-view optical images and Livox Avia LiDAR acquires LiDAR point clouds. As the scanning time increases, the density of the Livox LiDAR point cloud increases, and the final point cloud obtained clearly shares the shape contours of the target. The imaging resolution of the Spedal monocular camera is 1920 × 1080 pixels.

Figure 4 shows the experiment scene of the near-field array 3D SAR system. By moving the RF module on the horizontal and vertical rails, horizontal and vertical two-dimensional scanning is completed, and the virtual synthetic aperture is formed. The center frequency of the system's transmission signal is 78.8 GHz, with a maximum transmission signal bandwidth of 4 GHz. The array length of the system is 0.4 m × 0.4 m and the operating distance is 1 m. The range resolution can reach up to 3.75 cm, and the azimuth and altitude resolution can reach the millimeter level. The size of the 3D SAR image in the range, azimuth and height directions is 256 × 408 × 200.

Experiments have been conducted using multi-source data collected from four targets: aircraft model 1, aircraft model 2, pincer, and satellite model. Figure 2 shows the scene of near-field 3D SAR image acquisition, the original near-field 3D SAR imaging results, and the results obtained through the near-field SAR preprocessing process detailed in Section 3.1.1. Figure 3 exhibits the scene of LiDAR point cloud acquisition, the original LiDAR point clouds, and the results obtained through the LiDAR preprocessing process detailed in Section 3.1.2. Figure 4 depicts multi-view 2D optical image acquisition, the original 3D reconstruction results, and the results of the optical point cloud preprocessing process detailed in Section 3.1.3.
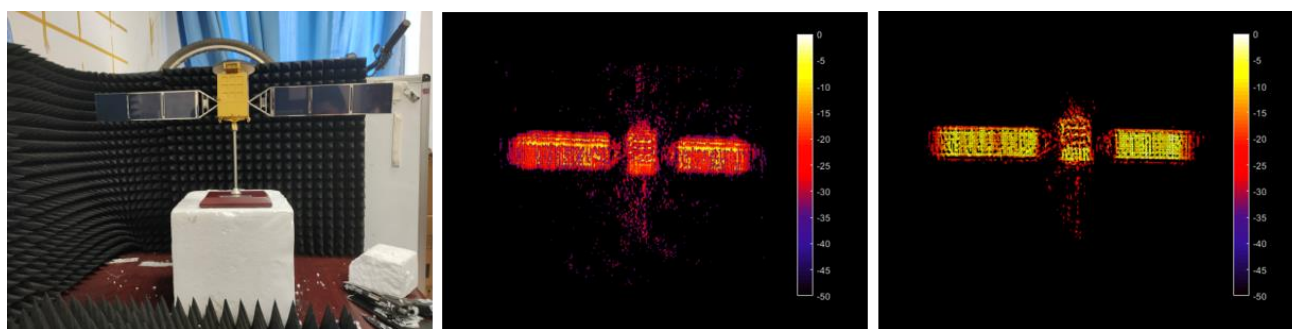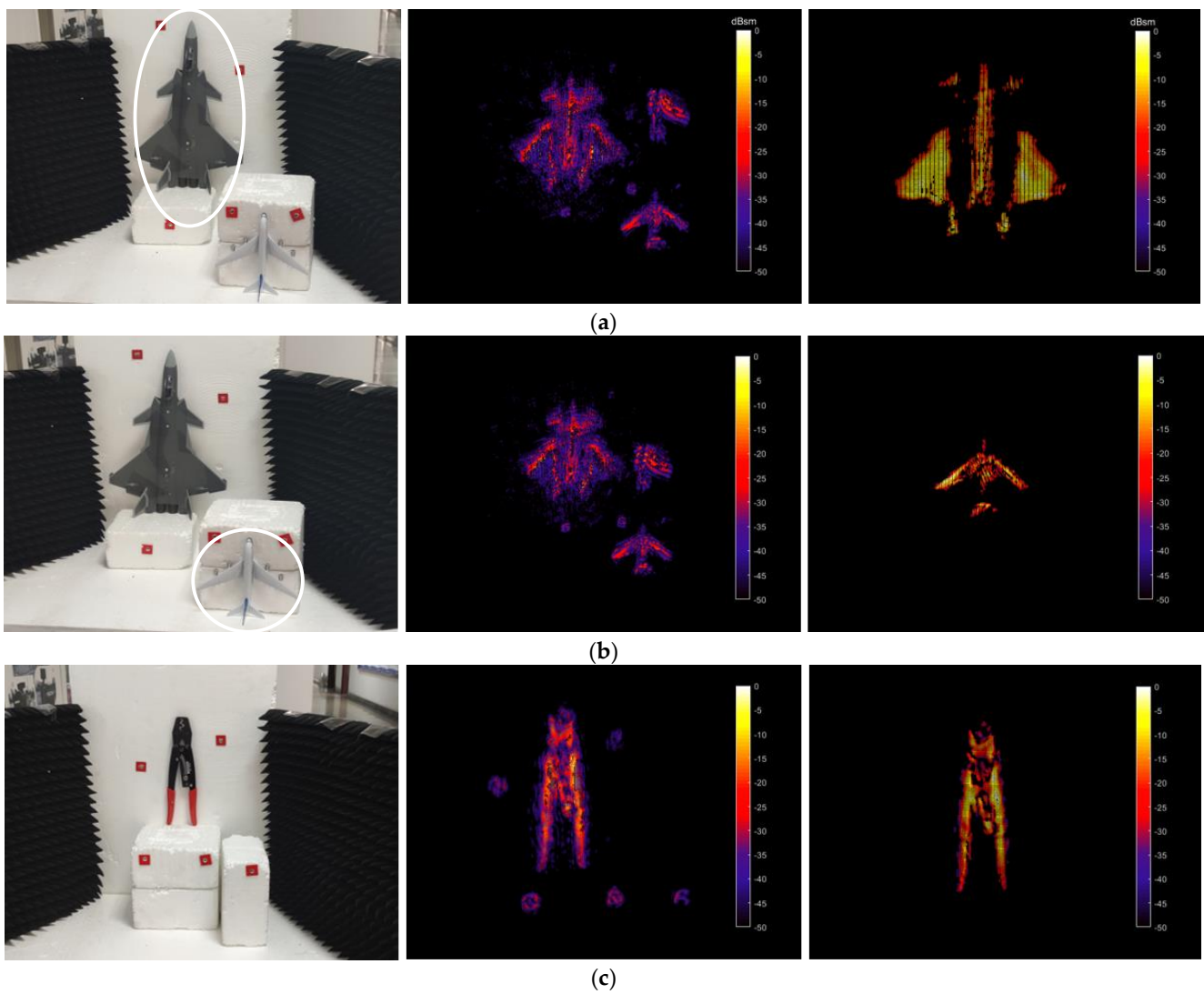

(**d**)

Figure 1. *Cont.*

Figure 1. Near-field array 3D SAR data acquisition, imaging results and preprocessing results. (**a**) Results for aircraft model 1; (**b**) results for aircraft model 2; (**c**) results for pincer; (**d**) results for satellite model.

Figure 2. *Cont.*

Figure 2. LiDAR data acquisition, imaging results and preprocessing results. (**a**) Results for aircraft model 1; (**b**) results for aircraft model 2; (**c**) results for pincer; (**d**) results for satellite model.

Figure 3. Camera data acquisition, imaging results and preprocessing results. (**a**) Results for aircraft model 1; (**b**) results for aircraft model 2; (**c**) results for pincer; (**d**) results for satellite model.
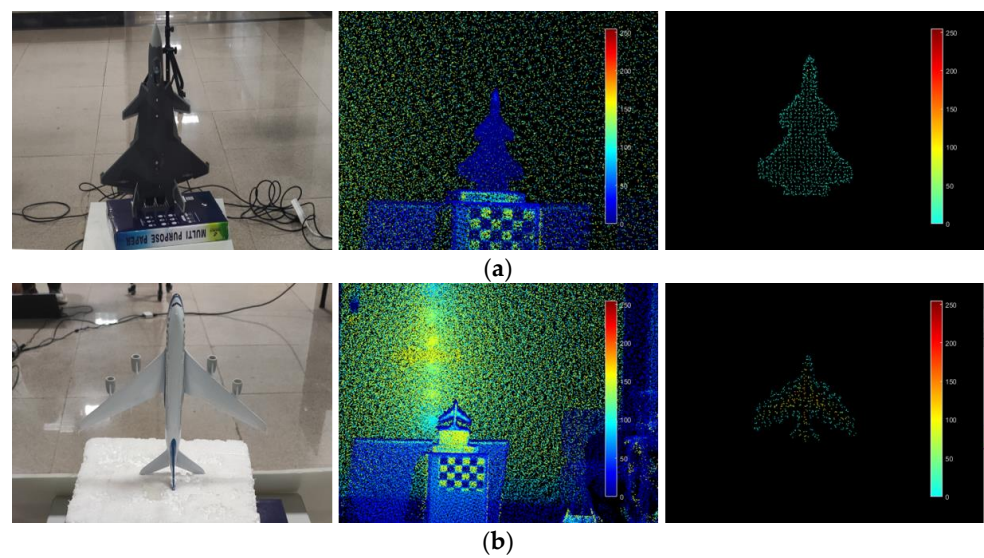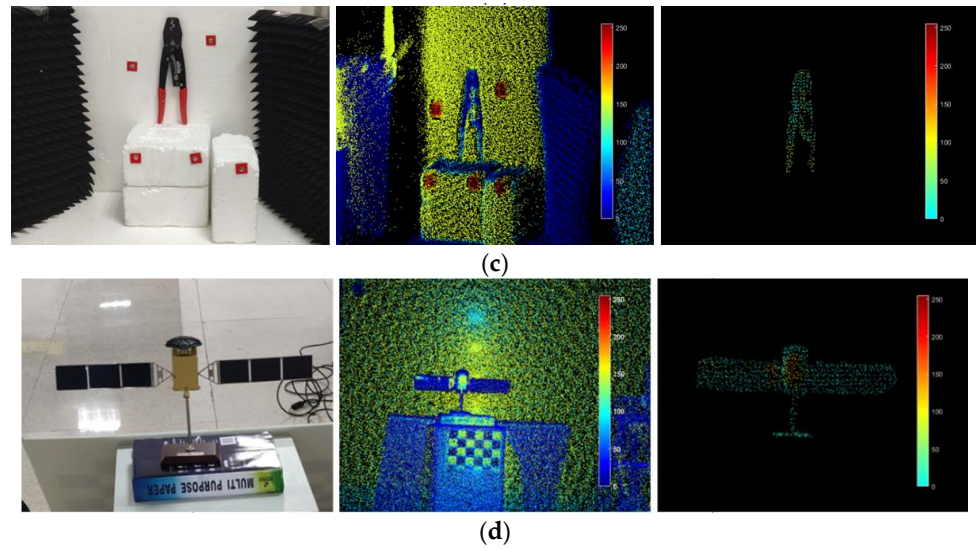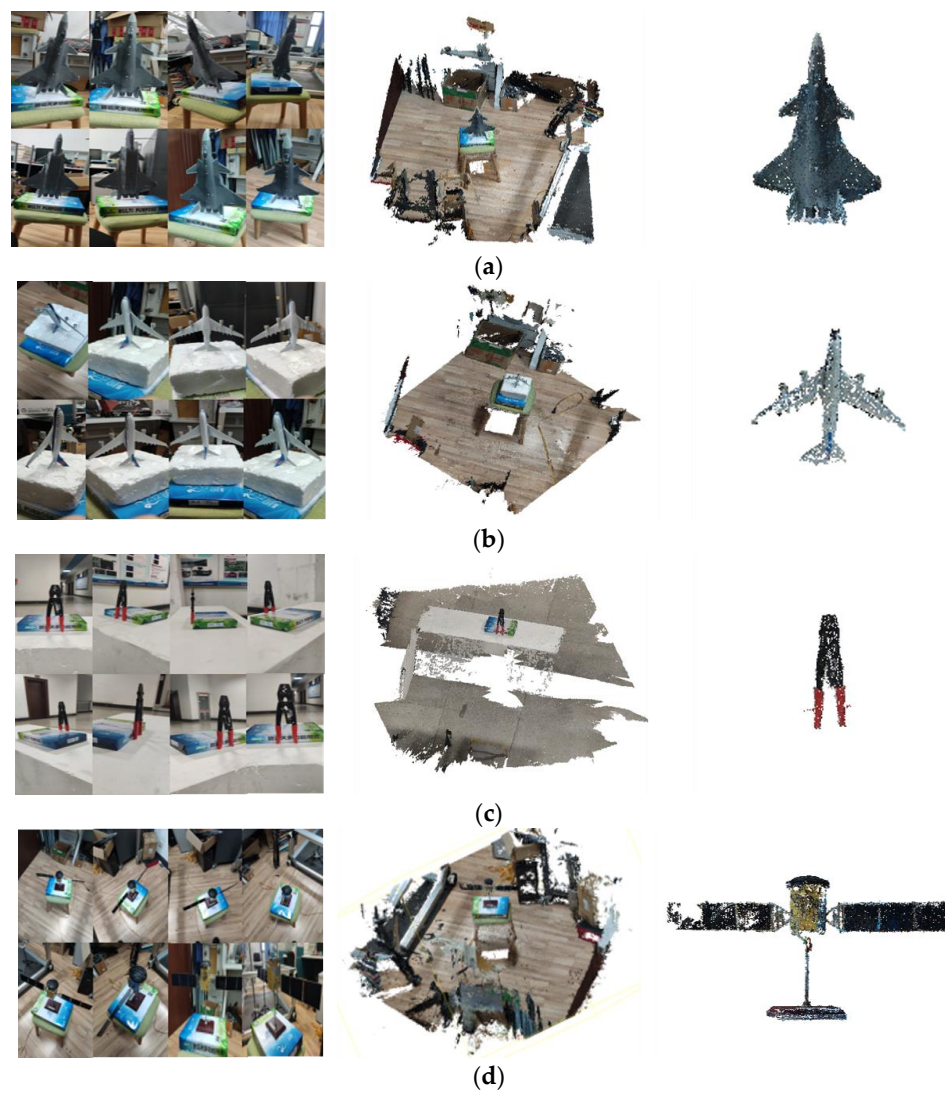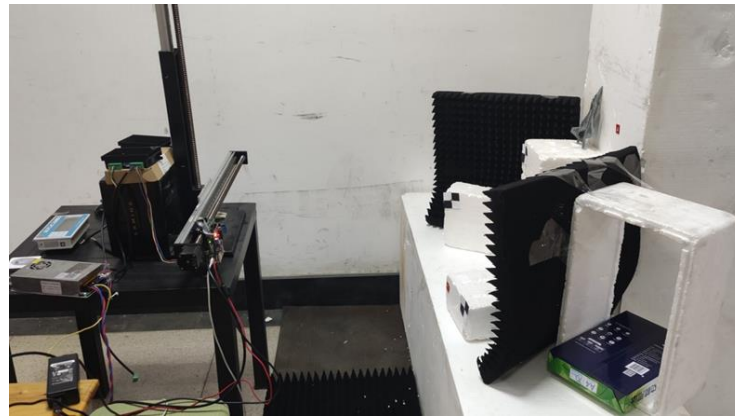
Figure 4. The experiment scene of the near-field array 3D SAR imaging system.

3. Methodology

The overall flowchart of the proposed near-field SAR multimodal fusion framework is shown in Figure 5. The framework consists of data preprocessing, point cloud registration, and data fusion. In the data preprocessing stage, near-field 3D SAR imaging, LiDAR imaging, and optical 3D reconstruction are first performed independently on measured data obtained from corresponding sensors. Then, filtering operations are used to denoise the random or spurious points. After filtering, the down-sampling reduces point density, enables uniformity and produces computational efficiency. Point clouds generated by LiDAR and optical sensors can be extremely dense. Down-sampling reduces the number of points in the cloud, making it more manageable for subsequent processing steps. It helps in achieving uniform point densities across the entire point cloud, ensuring that there are no areas with excessively high point density or gaps. Besides, processing and analyzing dense point clouds requires significant computational resources. Down-sampling reduces the computational burden by reducing the number of points while still retaining essential spatial information. Finally, segmentation operations are used to extract the target point cloud. As for the point cloud registration stage, a novel three-stage registration method including key point extraction, coarse registration, and fine registration is performed to obtain the pose transformation matrix of multi-sensor point clouds. This stage is the core of our fusion framework. It will be introduced and explained in detail in the next section. In the data fusion stage, the pose transformation matrix is used to align three point clouds, and the nearest neighbor point search algorithm is used to remove the redundancy of SAR to obtain the multi-sensor point cloud fusion result.
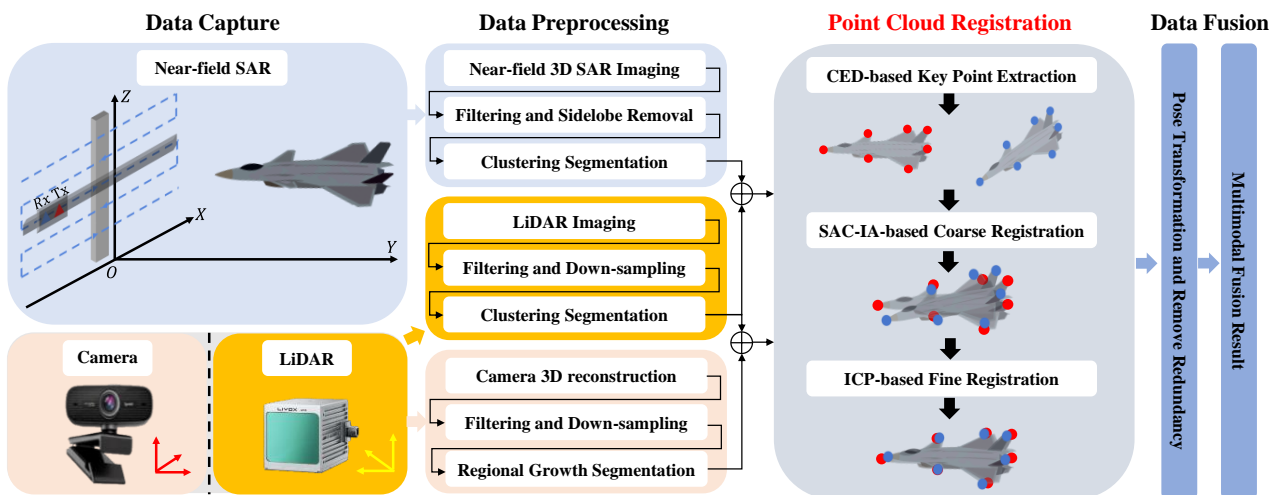

Figure 5. The overall flowchart of the proposed near-field SAR multimodal fusion framework.

3.1. Data Preprocessing

3.1.1. Near-Field SAR Preprocessing

The preprocessing of the near-field 3D SAR image to extract targets is shown in Figure 6a. The near-field 3D SAR image is generated using the BP algorithm. It is then converted into a point cloud format through global threshold filtering. Here, any pixel in the 3D image grid above the threshold is retained as a point in the point cloud. The threshold is set based on the specific dynamic range required. Based on the approximate position of the target in the observation scene, points outside the imaging area of the SAR point cloud are removed through passthrough filtering. Then, threshold extraction is performed to filter out low scattering background noise and interference by setting the absolute value of SAR scattering intensity. The sidelobes in the near-field 3D SAR point cloud are significant, and they blur the true distance and shape of the target and bring outliers to registration. Therefore, the sidelobes are removed by taking the maximum scattering intensity in the distance direction, and the main lobes are retained. Next, statistical filtering is used to process the near-field 3D SAR point cloud to remove discrete strong scattering noise, which will affect the subsequent point cloud feature calculation and registration. Finally, Euclidean distance clustering segmentation [35] is used to extract the target point cloud. After filtering out the noise, the near-field SAR point cloud is sparsely distributed in space, which is suitable for Euclidean distance segmentation.
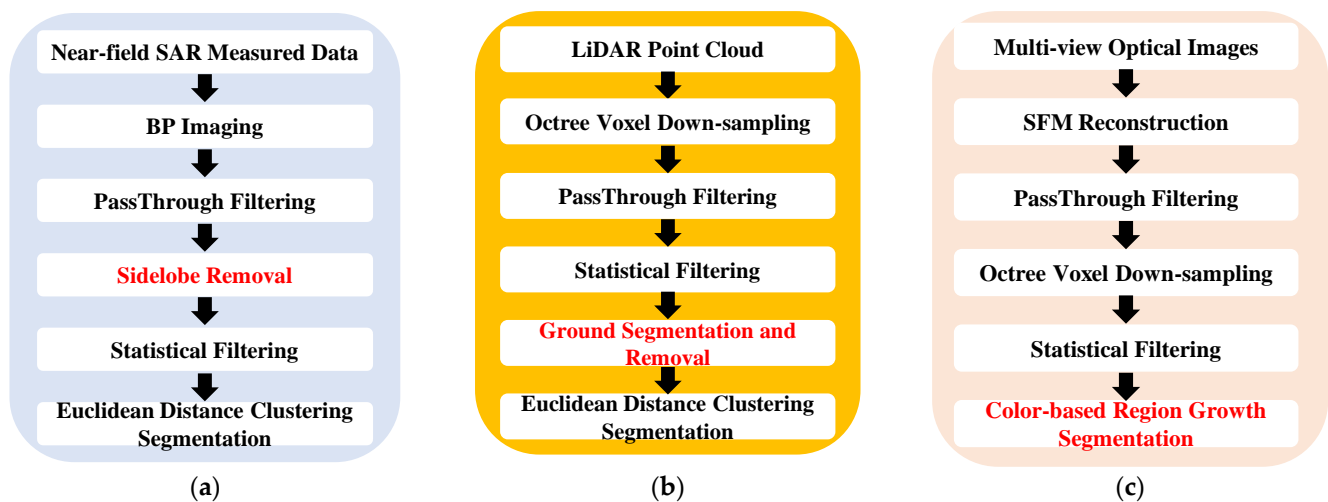

Figure 6. The data preprocessing pipeline used in our proposed multimodal fusion framework. (**a**) Specific near-field SAR data preprocessing operations; (**b**) specific LiDAR data preprocessing operations; (**c**) specific camera data preprocessing operations.

3.1.2. LiDAR Preprocessing

The preprocessing of the LiDAR point cloud to extract targets is shown in Figure 6b. The points outside the target area in the LiDAR point cloud are filtered out through passthrough filtering to reduce the size of point cloud. Then, the LiDAR point cloud is processed through octree voxel down-sampling to facilitate the further correspondence search with the voxel-transformed SAR point cloud. Due to the octree voxel down-sampling method retaining the centroid of the voxel grid as the sampling point, rather than the points in the original point cloud, the detailed features of the point cloud are destroyed. Therefore, the process selects the point closest to the centroid of the voxel grid in the original point cloud as the sampling point. Next, statistical filtering is used to filter out outliers and noise in the LiDAR point cloud. Finally, the M-estimator sample consensus (MSAC) algorithm [36] is used to obtain the platform plane where the target is located and remove it. The LiDAR point cloud of the target is segmented using the Euclidean distance clustering segmentation method mentioned in Section 3.1.1.

3.1.3. Camera Preprocessing

The preprocessing of the multi-view optical images to extract targets is shown in Figure 6c. Multi-view optical images are reconstructed through the SFM algorithm to obtain optical point clouds. And the point cloud of target area is obtained by passthrough filtering. Then, the optical point cloud is down-sampled using the octree voxel down-sampling method described in Section 3.1.2, which reduces the size of the point cloud while preserving the target structural features and overcoming the resolution differences with LiDAR and SAR point clouds. Next, the process includes statistical filtering on the optical point cloud to remove outliers generated during SFM reconstruction. Finally, a color-based region growth segmentation method [37] is used to segment the target. Optical point clouds have abundant color and texture information. The color-based region growth segmentation method utilizes color differences between points for clustering, which can effectively segment optical point clouds.

3.2. SAR–LiDAR Point Cloud Registration

3.2.1. Basic Principles of Point Cloud Registration

Point cloud registration aligns the coordinate systems of two input point clouds by solving the spatial transformation matrix between them. A point cloud is a collection of points. We assume the two input point clouds are $X = \{x_1, \cdots, x_i, \cdots, x_M\} \in \mathbb{R}^{M \times 3}$ and $Y = \{y_1, \cdots, y_i, \cdots, y_N\} \in \mathbb{R}^{N \times 3}$, where x_i and y_i are the coordinates of the ith points in the point clouds X and Y, respectively. Suppose X and Y have Z pairs of correspondences, where the corresponding point set is $D = \{(x_1, y_1), \cdots, (x_Z, y_Z)\}$. The spatial transformation matrix includes rotation, translation, and scaling transformations. The rotation transformation includes pitch angle, yaw angle, and roll angle, the translation transformation includes the translation of three coordinate axes, and the scaling transformation includes one scaling factor, which are represented as the rotation matrix $R \in \mathbb{R}^{3 \times 3}$, translation vector $t \in \mathbb{R}^3$, and scaling factor f_s, respectively. Scaling is not considered in rigid registration, so scaling factor f_s is ignored. The goal of registration is to find the rigid transformation parameters R and t that best align the point cloud X to Y, as shown below:

$$\underset{R \in \mathbb{R}^{3 \times 3}, t \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{k=1}^{Z} \|x_k - (Ry_k + t)\|_2^2 \tag{1}$$

where $\|x_k - (Ry_k + t)\|_2^2$ is the projection error of the kth corresponding point between X and transformed Y. By solving the above optimization problem to minimize the position error between the two point clouds, the optimal spatial transformation matrix (R and t) is obtained. When the corresponding points between two point clouds are obtained, singular value decomposition (SVD decomposition) is usually used to solve the transformation matrix [38].

Traditional registration methods use optimization strategies to estimate the transformation matrix. The most commonly used optimization-based registration method is the ICP algorithm, which contains two stages: correspondence searching and transformation estimation. Correspondence searching is intended to find the matched point for the input point clouds. Transformation estimation is used to estimate the transformation matrix via the correspondences. These two stages will be conducted iteratively to find the optimal transformation. If the initial pose differences of input point clouds are significant, the ICP algorithm struggles to find precise correspondences during the iterative process, and its estimated transformation matrix is also inaccurate. The two-step registration method is then adopted in homogeneous point cloud registration, which roughly aligns the point cloud pose through coarse registration. However, the different imaging mechanisms of multiple sensors pose challenges to multimodal point cloud registration in terms of data format, noise, and resolution differences. Therefore, this study proposes a multi-sensor point cloud registration method that involves three stages of key point extraction, coarse registration, and fine registration to achieve high-precision multi-source point cloud registration.

3.2.2. Key Point Extraction with Structure-Intensity Constraints

Key points are points in the point cloud that have significant features, including geometric structure, color, and intensity, which can effectively describe the original point cloud. Compared to the original point cloud, the number of key points is relatively small. In addition, as the points' relative positions remain unchanged during point cloud rotation and translation, the extracted key points have rotational and translational invariance. Therefore, key points can be utilized to replace the original point cloud for registration. Using key points for point cloud registration can preserve point cloud features, eliminate multimodal point cloud heterogeneity and improve registration efficiency.

The near-field SAR and LiDAR point cloud contain the spatial position coordinates and intensity information of points. Most existing key point detectors often focus on extracting key points from a single feature, which reduces the description ability of the extracted key points. Note that the centroid-distance (CED) detector [39] has been recently proven to be more effective. Although the CED detector is a multi-feature key point detector that can extract geometric structure and color key points from color point clouds, it does not focus on the extraction of intensity key points, and so cannot be used for SAR and LiDAR. Therefore, this study designs a novel detector based on the CED detector that can extract geometric structure and intensity key points, enabling the key point extraction of near-field SAR and LiDAR point clouds.

Specifically, the process of our key point extraction is to calculate the significance of each point in the point cloud in its sphere neighborhood, and then retain key points with higher significance compared to all neighboring points in the sphere neighborhood through non-maximum suppression, whereby the significance refers to geometric structure and intensity. Assuming there is a LiDAR point cloud set Q and $q = \left[q^G, q^S\right]^T$ is one of the points, $q^G = \{x, y, z\}$ is the geometric coordinate of point q, and q^S is the intensity of point q. We set point q as the query point and r as the radius of the spherical neighborhood, ad search for all points within the spherical neighborhood to form the set of neighboring points $N_q = \left\{q_i \middle| \left\|q^G - q_i^G\right\|_2 < r\right\}$ for point q.

The first step is to calculate the geometric significance and intensity significance of the points. The geometric centroid of the spherical neighborhood of point q can be obtained by the following equation:

$$\mu_q^G = \frac{1}{I}\sum_{i=1}^{I} q_i^G \tag{2}$$

where I is the number of neighboring points of point q. The intensity centroid of the spherical neighborhood of point q can be obtained by the following equation:

$$\mu_q^S = \frac{1}{I}\sum_{i=1}^{I} q_j^S \tag{3}$$

Intuitively, the larger the distance from the point to the geometric centroid, the more prominent the geometrical significance, such as corner points. And the greater the intensity difference between the point and the intensity centroid, the more prominent the intensity significance. Therefore, the geometric significance of point q is measured by its distance from the geometric centroid of its sphere, as follows:

$$d_G = \left\|q^G - \mu_q^G\right\|_2 \tag{4}$$

The intensity significance of point q is represented by the L1 norm of its intensity and the intensity centroid of its sphere neighborhood, as follows:

$$d_S = \left\|q^S - \mu_q^S\right\|_1 \tag{5}$$

The second step is to obtain key points with high significance. We traverse all points in the point cloud Q, and filter out the point with low significance using Equation (6).

$$\begin{cases} d_G < d_{Gt} \\ d_S < d_{St} \end{cases} \qquad (6)$$

where d_{Gt} is the geometric significance threshold and d_{St} is the intensity significance threshold. In order to select points with high geometric and intensity significance within the sphere neighborhood, the non-maximum suppression algorithm [40] is used for screening the key points that meet Equation (7), where d_{Gi} is the geometric significance threshold and d_{Si} is the intensity significance threshold of neighboring point q_i.

$$d_G \cdot d_S \geq d_{Gi} \cdot d_{Si}, \ \forall q_i \in N_q \qquad (7)$$

3.2.3. SAC-IA Coarse Registration with SHOT Feature and Geometric Relationship Constraints

After extracting the key points of the near-field SAR and LiDAR point cloud, coarse registration can be performed using these key points to give a good initial pose between the input point clouds. Enhancements to the correspondence searching process of the original SAC-IA algorithm [41] come through the signature of the histogram of orientation (SHOT) feature descriptor [42] and geometric relationship constraints for better SAR–LiDAR coarse registration. The original SAC-IA algorithm only relies on the fast point feature histogram (FPFH) feature descriptors [43] to select correspondences, without considering the geometric relationship between correspondences. When the corresponding points are incorrectly selected, a problem of ambiguous rotation angles arises. Such ambiguity caused by the three collinearity points can be overcome by triangular relationships constraints. Furthermore, compared to the SAC-IA algorithm using the FPFH feature to describe the features of points, SHOT feature descriptors are more robust to point clouds with incomplete surfaces and uneven density. Therefore, on the basis of the original SAC-IA, this study uses both the SHOT features and triangular relationships of the corresponding points to constrain the correspondence search.

Before executing our improved coarse registration algorithm, we should calculate the SHOT feature descriptor of the key points. The SHOT feature descriptor uses the adjacent points to encode the key points and obtain the corresponding feature vectors. SHOT features have rotation and translation invariance and can be used for correspondence selection in point cloud registration. The steps for constructing SHOT feature descriptors are as follows.

First, we build a unique coordinate system centered around key points. For the key point $q' \in Q$, we construct the covariance matrix E_S of point q' in a spherical neighborhood space with a search radius of r_s via the following equations, where $q' = \{x', y', z'\}$ only denotes the geometric coordinate of the key point.

$$E_S = \frac{1}{\sum_{i=1}^{|N_{q'}|} (r_s - \overline{q}'_i)} \sum_{i=1}^{|N_{q'}|} (r_s - \overline{q}'_i)(q'_i - q')(q'_i - q')^T \qquad (8)$$

$$\overline{q}'_i = \|q'_i - q'\|_2 \qquad (9)$$

where $N_{q'}$ represents the set of all points within the spherical neighborhood of point q' and $q'_i \in N_{q'}$ and $\|\cdot\|_2$ is the matrix L2 norm. Eigenvalue decomposition is performed on the covariance matrix E_S to obtain the corresponding unit eigenvector x^+, y^+, z^+ in the order of decreasing eigenvalues. The unit vector in the opposite direction is x^-, y^-, z^-. $G(k)$ represents the index set of k points in the spherical neighborhood space that are closest to the median distance $d_m = median(\|q'_i - q'\|_2), i \in \left[1, |N_{q'}|\right]$ from point q'. In order to eliminate the symbol ambiguity caused by eigenvalue decomposition in constructing the unique coordinate system, the following steps are performed. The positive direction of the X-axis for resolving ambiguity can be obtained using the following equation.

$$S^+ = \{i| \ \|q_i' - q'\|_2 \le r_s, (q_i' - q') \cdot x^+ \ge 0\} \tag{10}$$

$$S^- = \{i| \ \|q_i' - q'\|_2 \le r_s, (q_i' - q') \cdot x^- \ge 0\} \tag{11}$$

$$\tilde{S}{}^+ = \{i| \ i \in G(k), (q_i' - q') \cdot x^+ \ge 0\} \tag{12}$$

$$\tilde{S}{}^- = \{i| \ i \in G(k), (q_i' - q') \cdot x^- \ge 0\} \tag{13}$$

$$x = \begin{cases} x^+ & |S^+| > |S^-| \\ x^+ & |S^+| = |S^+| \cap \left|\tilde{S}{}^+\right| > \left|\tilde{S}{}^-\right| \\ x^- & |S^+| < |S^-| \\ x^- & |S^+| = |S^+| \cap \left|\tilde{S}{}^+\right| < \left|\tilde{S}{}^-\right| \end{cases} \tag{14}$$

The above similar Equations (10)–(14) can be used to determine the positive direction of the local coordinate system's Z-axis, and we can then obtain the positive direction of the local coordinate system's Y-axis through $y = z \times x$.

Second, we encode adjacent points based on the unique coordinate system above to obtain SHOT features. Point q' is the origin of the unique coordinate system, and its spherical neighborhood space is divided into two parts along the radial direction, eight parts in the vertical direction, and two parts in the horizontal direction, resulting in a total of 32 feature subspaces. Equation (15) is used to calculate the cosine value of the angle θ_i between the unit normal vector $\overrightarrow{n_i}$ of the adjacent point q_i' falling into the region and the positive direction $\overrightarrow{z}_{q'}$ of the unique coordinate system's Z-axis in each subspace.

$$\cos \theta_i = \overrightarrow{z}_{q'} \cdot \overrightarrow{n_i} \tag{15}$$

In each subspace, the cosine value is divided into 11 parts to form a local histogram, and the adjacent points are classified into different cells of the local histogram based on the cosine value. After the local histograms of all subspaces are integrated, the boundary effect is solved using a quartic linear interpolation method to obtain the SHOT descriptor of the point, totaling $32 \times 11 = 352$ dimensions.

Then, the SAC-IA algorithm with SHOT feature–geometric relationship dual constraints facilitates coarse registration and aids in determining the initial pose transformation matrix *K* between the key points of near-field 3D SAR and LiDAR.

In the correspondence searching stage, we select $s = 3$ sample points in key points of SAR point cloud *P*, where the distance between points is greater than the distance threshold so as to ensure that the SHOT features of selected sample points are different. For each sample point, the nearest neighbor search is used to find three key points in point cloud *Q* with the smallest difference in SHOT features, and a random point in the three points is selected as the corresponding point. Assuming the set of corresponding key points obtained is $\{(p_a', q_a')|p_a' \in P, q_a' \in Q, a = 1, 2, 3\}$, we calculate the edge length as follows:

$$\begin{cases} e_a^P = \|p_a' - p_b'\|_2 \\ e_a^Q = \|q_a' - q_b'\|_2 \end{cases} \tag{16}$$

where $(a, b) = \{(1, 2), (2, 3), (3, 1)\}$. A triangle judgment is performed on the calculated edge length of the corresponding points. A congruence judgment is performed on the two obtained triangles. If the triangle condition and the congruence of the triangles are not met, we reselect the sample points. In addition, due to the different resolutions of multimodal point clouds, a threshold τ should be set to adjust the edge congruence judgment conditions, as follows:

$$\frac{1}{\tau} \le \frac{e_a^P}{e_a^Q} \le \tau, a = 1, 2, 3 \tag{17}$$

In the transformation estimation stage, the obtained $s = 3$ correspondences are used to solve the rigid transformation matrix between point cloud P and Q through SVD decomposition, and the Huber penalty function is used to calculate the distance error sum $\sum_{a=1}^{s} H(e_a)$ after the rigid transformation, as follows:

$$H(e_a) = \begin{cases} \frac{1}{2}e_a^2 & \|e_a\| \leq t_d \\ \frac{1}{2}t_d(2\|e_a\| - t_d) & \|e_a\| > t_d \end{cases} \tag{18}$$

where t_d is the set distance error threshold, e_a is the distance error of the a_{th} corresponding point after transformation, $\|\cdot\|$ represents the L1 norm, and $H(e_a)$ is the distance error after imposing Huber penalty on e_a. During the iteration of the above two operations, if the current distance error sum is the smallest, the transformation matrix will be retained, and the initial pose transformation matrix $K_{SL} = \begin{bmatrix} R' & t' \\ 0 & 1 \end{bmatrix}$ of the input point cloud can be obtained until the iteration ends, where R' is the spatial rotation matrix and t' is the translation vector.

3.2.4. ICP Fine Registration with Adaptive Threshold

After the coarse registration of key points, the initial pose transformation matrix K is obtained. Then, the improved ICP algorithm is proposed to accurately align the original point clouds of near-field SAR and LiDAR, and the precise pose transformation matrix K' is obtained.

It should be noted that there is a disparity in resolution between the two types of point clouds. LiDAR point clouds typically exhibit higher intensity levels and closer proximity between adjacent points compared to SAR point clouds. The original ICP algorithm only needs to select the corresponding points with the minimum distance, which is less than the given distance threshold during the correspondence searching. Therefore, the distance threshold for the judgment is fixed. During the iterative optimization process, the ICP algorithm approaches the optimal solution, while the distance between searched corresponding points also decreases. A fixed distance threshold will introduce more corresponding points during the later iteration, resulting in an increase in registration time.

To maintain accuracy while simultaneously improving registration efficiency, the adaptive threshold is adopted. The method replaces the fixed distance threshold with an adaptive one, which increases continuously to maintain accuracy and improve registration efficiency as the iterative optimization process progresses. Initially, a smaller threshold is used to capture fine-grained correspondences and refine the alignment. As the optimization progresses and the point clouds become closer to alignment, the threshold increases, allowing for faster convergence while still ensuring accurate registration.

The steps of the improved ICP algorithm are as follows. In the parameter initialization stage, we obtain the transformed point cloud $P^{(0)}$ after coarse registration by $P^{(0)} = R'P + t'$, and set the initial distance threshold $d_t^{(0)}$, the overall distance error threshold ε and the maximum number of iterations M_{iter}. In the stage of correspondence searching of the i_{th} iteration, for $\forall p_j \in P^{(i-1)}$, we find the point $q_j \in Q$ closest to p_j through the nearest search, where p_j and q_j only denote the geometric coordinates of the point. If $\|p_j - q_j\|_2 \leq d_t^{(i)}$, p_j and q_j form the correspondence pairs, where $d_t^{(i)}$ is the distance threshold of i_{th} iteration. The final corresponding point set $\{(p_j, q_j) \mid p_j \in P, q_j \in Q, j = 1, \cdots, J\}$ is obtained. In the transformation estimation stage of the i_{th} iteration, we calculate the centroids of two point clouds in the corresponding point set using the following equations, denoted as μ_p and μ_q.

$$\mu_p = \frac{1}{J}\sum_{j=1}^{J} p_j \tag{19}$$

$$\mu_q = \frac{1}{J}\sum_{j=1}^{J} q_j \tag{20}$$

We construct the covariance matrix $E_J = \frac{1}{J}\sum_{j=1}^{J} (p_j - \mu_p)(q_j - \mu_q)^T$ and perform SVD decomposition on the covariance matrix using Equation (21).

$$E_J = U\Sigma V^T \tag{21}$$

where U and V are orthogonal matrices of 3×3, and Σ is a diagonal matrix composed of the eigenvalues of the covariance matrix E_J. Further, the rotation matrix $R^{(i)}$ and the translation vector $t^{(i)}$ can be obtained as follows:

$$R^{(i)} = VU^T \tag{22}$$

$$t^{(i)} = \mu_q - R^{(i)}\mu_p \tag{23}$$

The transformed point cloud $P^{(i)}$ can be obtained by $P^{(i)} = R^{(i)}P^{(i-1)} + t^{(i)}$. The distance error function F is calculated using Equation (24).

$$F = \frac{1}{J}\sum_{j=1}^{J} \left\| q_j - \left(R^{(i)}p_j + t^{(i)}\right) \right\|_2 \tag{24}$$

If $F \leq \varepsilon$ or has reached the maximum number of iterations, we stop the iteration. Otherwise, we update the distance threshold:

$$d_t^{(i)} = \rho \cdot d_t^{(i-1)} + (1 - \rho)d_t^{(0)} \tag{25}$$

where ρ is a constant. We use the new point cloud $P^{(i)}$ to return to the correspondence searching stage and continue with the next iteration. After the iteration ends, the final spatial rotation matrix R_{SL} and translation vector t_{SL} of the near-field SAR and LiDAR point cloud are obtained by $R_{SL} = \prod_{i=1}^{M_{end}} R^{(i)} \cdot R'$ and $t_{SL} = t' + \sum_{i=1}^{M_{end}} t^{(i)}$, where M_{end} is the number of iterations terminated. The final spatial transformation matrix is as follows:

$$K'_{SL} = \begin{bmatrix} R_{SL} & t_{SL} \\ 0 & 1 \end{bmatrix} \tag{26}$$

3.3. Camera–LiDAR Point Cloud Registration

Compared with near-field SAR and LiDAR point cloud registration, LiDAR and optical color point cloud registration have similar processes. However, optical point clouds contain target size distortion that affects the registration performance, and do not have the intensity information required by the proposed coarse registration algorithm with geometric and intensity dual constraints. Therefore, some special treatments have been developed for optical point clouds, namely, format conversion and size correction. First, we convert the color of the optical color point cloud into intensity via the following equation [44]:

$$intensity = 0.299 \cdot r + 0.587 \cdot g + 0.114 \cdot b \tag{27}$$

where r, g, b represent the red component, green component, and blue component of optical color, respectively.

After obtaining the optical intensity point cloud, the approach uses LiDAR point clouds that can truly reflect the target size as the size correction benchmark to address the target size distortion. The principal components analysis (PCA) [45] was used to correct the size of the optical point cloud. The calibration steps are as follows.

Assuming the LiDAR point cloud is $L = \{l_1, \cdots, l_N\}$ and the optical point cloud is $O = \{o_1, \cdots, o_M\}$, we calculate the centroids of point clouds L and O as follows:

$$l_c = \frac{1}{N}\sum_{i=1}^{N} l_i \tag{28}$$

$$o_c = \frac{1}{M} \sum_{i=1}^{M} o_i \tag{29}$$

We calculate covariance matrix as follows:

$$E_L = \frac{1}{N} \sum_{i=1}^{N} (l_i - l_c)(l_i - l_c)^T \tag{30}$$

$$E_O = \frac{1}{M} \sum_{i=1}^{M} (o_i - o_c)(o_i - o_c)^T \tag{31}$$

$\{v_1^L, v_2^L, v_3^L\}$ and $\{v_1^O, v_2^O, v_3^O\}$ are eigenvectors of covariance matrices E_L and E_O. Then, cross product orthogonalization is performed on a linearly independent basis to obtain the orthogonal basis and form feature space.

$$\begin{cases} w_3^L = v_1^L \times v_2^L \\ w_1^L = v_2^L \times v_3^L \\ w_2^L = v_2^L \end{cases} \tag{32}$$

$$\begin{cases} w_3^O = v_1^O \times v_2^O \\ w_1^O = v_2^O \times v_3^O \\ w_2^O = v_2^O \end{cases} \tag{33}$$

We calculate the rotation matrix R_L, R_O and translation vector t_L, t_O for converting point cloud s L and O from the original coordinate system to the feature space coordinate system, as follows:

$$\begin{cases} R_L = (w_1^L, w_2^L, w_3^L)^T \\ t_L = -R_L x_c \end{cases} \tag{34}$$

$$\begin{cases} R_O = (w_1^O, w_2^O, w_3^O)^T \\ t_O = -R_O y_c \end{cases} \tag{35}$$

$$\begin{cases} L_f = R_L L + t_L \\ O_f = R_O O + t_O \end{cases} \tag{36}$$

The point clouds L_f and O_f are converted to the feature space coordinate systems through Equation (36). We assume r_L, r_O are the coordinate ranges of point clouds L_f and O_f in the orthogonal basis direction corresponding to the maximum eigenvalue, and the scaling factor of the optical point clouds is calculated by $f_s = r_O / r_L$. The corrected optical point cloud is acquired by $O_C = O / f_s$.

After the format conversion and size correction of the optical point cloud, similar algorithms to those for key point extraction, coarse registration, and fine registration in Section 3.2 are used to obtain the spatial transformation matrix between the LiDAR and camera point cloud, as follows:

$$K'_{CL} = \begin{bmatrix} f_s R_{CL} & t_{CL} \\ 0 & 1 \end{bmatrix} \tag{37}$$

where R_{CL} and t_{CL} represent the rotation matrix and translation vector converted from camera point cloud to LiDAR point cloud.

3.4. SAR-Camera-LiDAR Data Fusion

After point cloud registration, the following steps are used for multimodal data fusion. First, the LiDAR coordinate system serves as the reference coordinate system to align the multi-sensor point cloud coordinate system. By registering near-field SAR point clouds with LiDAR point clouds, the spatial transformation matrix K'_{SL} is obtained. By registering optical color point clouds with LiDAR point clouds, the spatial transformation matrix K'_{CL} is obtained. Assuming the near-field SAR point cloud of the target is P_S, the LiDAR point

cloud is P_L, and the optical color point cloud is P_C, then multimodal point cloud coordinate alignment can be achieved through Equation (38).

$$\begin{cases} \tilde{P}_S = K'_{SL} P_s \\ \tilde{P}_C = K'_{CL} P_C \end{cases} \tag{38}$$

Then, the process mixes optical point clouds and SAR point clouds with color objects to ensure that the fused point clouds have both scattering and color information. The near-field SAR point cloud is colored based on scattering intensity to reflect the scattering information of objects. Each point in the optical point cloud is traversed and the closest point in the near-field SAR point cloud is found by the nearest neighbor search algorithm. If the distance between two points is less than the set threshold, we replace the color of the optical points with the color of the near-field SAR points. Otherwise, the optical color will still be used.

Finally, the redundancy in SAR point clouds is removed by deleting outliers relative to optical and LiDAR point clouds, and the multimodal fusion results with near-field SAR scattering intensity, precise geometric shape size, and color texture information are obtained.

4. Experimental Results

This section will describe the multi-sensor prototype experimental hardware system from which measured data were collected, and discuss the results obtained with the multimodal data fusion framework mentioned in the previous section. This section is organized into five parts. Section 4.1 presents and evaluates SAR–LiDAR registration results for the proposed point cloud registration algorithm, and Section 4.2 presents camera–LiDAR registration results. Section 4.3 demonstrates the SAR–camera–LiDAR multimodal data fusion results based on the proposed framework. Finally, Section 4.4 discusses the applications of the current work and shows relevant experimental results. The computer system used to test method has an Intel i7-10700 CPU with RTX 2070s graphics card and 64 GB of RAM memory.

4.1. SAR–LiDAR Registration Results

Manual corresponding point selection is used to obtain the rotation matrix R_g and translation vector t_g as registration truth values. Table 1 presents the quantitative evaluation results of the proposed improved registration method and the original registration method. The evaluation indicators are registration error and registration time, whereby registration error includes rotation error E_R and translation error E_t [46].

$$\begin{cases} E_t = \left\| t_g - t_e \right\|_2 \\ E_R = \frac{180}{\pi} \left| \cos^{-1} \frac{tr\left(R_g^{-1} R_e\right) - 1}{2} \right| \end{cases} \tag{39}$$

Table 1. Comparison of registration error and time of point cloud registration algorithms before and after improvement.

Method	Aircraft Model 1			Aircraft Model 2			Pincer			Satellite		
	$E_R(°)$	E_t(m)	T(s)	$E_R(°)$	E_t(m)	T(s)	$E_R(°)$	E_t(m)	T(s)	$E_R(°)$	E_t(m)	T(s)
Original	121.4943	**0.0076**	5.724	172.1562	1.5333	1.351	10.1162	0.1991	3.433	3.4851	0.0262	11.462
Ours	**0.9885**	0.01	**2.825**	**6.2602**	**0.0739**	**0.595**	**4.8143**	**0.0916**	**0.567**	**1.521**	**0.0143**	**5.801**

Here, t_g represents the true value of the registration translation vector, t_e represents the estimated value of the registration method translation vector, R_g represents the true value of the registration rotation matrix, R_e represents the estimated value of the registration method rotation matrix, and $tr(\cdot)$ is the matrix trace operation.

By comparing the performance of the proposed improved method with that of the original two-step registration method (SAC-IA+ICP) in Table 1, it can be concluded that the improved method outperforms the original method in terms of registration error and efficiency. The original method can achieve rough alignment results in the registration of pincer and satellite models, but the rotation error is significantly large in the registration of two aircraft models, while the improved method can achieve accurate alignment on all targets. On the one hand, since the improved method extracts key points that better reflect the targets' structural characteristics compared to general points, it can provide more accurate corresponding point pairs in the correspondence searching stage of registration. On the other hand, the FPFH descriptor used in the original method is less effective in describing SAR point clouds than the SHOT descriptor, and the geometric relationship constraints also minimize the impact of rotation angle ambiguity.

In addition, we quantitatively compare the performance of the proposed method with current mainstream registration methods, including Super4PCS [47], ICP, NDT [48], and CPD [49]. Table 2 shows the registration errors of these methods, demonstrating that the proposed method achieves the lowest rotation angle errors in all experiments, with competitive performance in translation errors. The Super4PCS algorithm fails in registering the pincer, possibly due to the thin thickness of the pincer, which is similar to planar target. This resulted in the Super4PCS algorithm extracting a large set of error points. The ICP algorithm's registration requires a good initial posture. In cases of poor initial posture, it is easy to fail registration and obtain local optimal solutions. When registering the two aircraft models and the pincer, there is a large angle rotation deviation compared to other methods. Both NDT and CPD are registration methods based on hypothetical probability distribution models. When the shape difference between the two point clouds is significant, incorrect matching may occur due to the difference point distributions.

Table 2. Comparison of registration errors between the proposed method and other point cloud registration methods.

Method	Aircraft Model 1		Aircraft Model 2		Pincer		Satellite	
	$E_R(°)$	E_t(m)	$E_R(°)$	E_t(m)	$E_R(°)$	E_t(m)	$E_R(°)$	E_t(m)
Super4PCS	147.628	0.1876	17.5115	0.2197	\	\	10.5766	0.0852
ICP	75.8897	1.4288	169.8124	1.5373	173.4451	2.2754	5.0486	0.039
NDT	16.6345	0.3575	7.2386	**0.0533**	9.9256	0.1844	10.7355	0.1056
CPD	3.3796	0.0956	12.6813	0.1794	5.1054	0.0918	4.2393	0.0409
Ours	**0.9885**	**0.01**	**6.2602**	0.0739	**4.8143**	**0.0916**	**1.521**	**0.0143**

Figure 7 shows the registration results of the proposed method for near-field SAR point clouds and LiDAR point clouds. The LiDAR point cloud is displayed in white, and the near-field array 3D SAR point cloud is colored based on scattering intensity. The spatial positions of near-field SAR point clouds and LiDAR point clouds before registration are marked with ellipses, and the details of near-field SAR point clouds and LiDAR point clouds are displayed in a white box in the middle of the image.

It can be seen that LiDAR can accurately describe the geometric shapes of all targets, while there are different levels of shape missing in SAR point clouds. Therefore, relying on the precise positioning of LiDAR and the scattering information of SAR, the four targets can be accurately aligned in position, and the scattering characteristics of each target can also be clearly located in the registration results of the near-field SAR point cloud and LiDAR point cloud. However, LiDAR point clouds lack color and texture information, and so it is still necessary to fuse optical photos to assist the target category judgment.
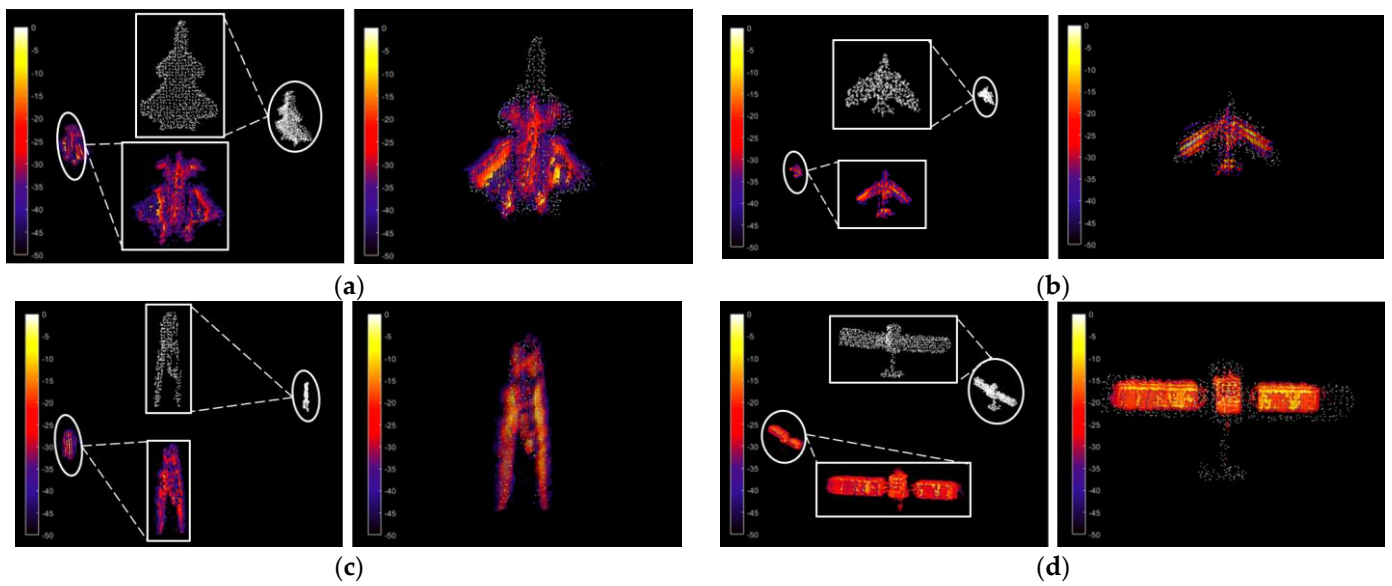
Figure 7. Near-field SAR point clouds and LiDAR point clouds before and after registration. (**a**) Aircraft model 1 before and after registration; (**b**) aircraft model 2 before and after registration; (**c**) pincer before and after registration; (**d**) satellite model before and after registration.

4.2. Camera–LiDAR Registration Results

After the format conversion and size correction of the optical point cloud, the proposed method is used to register the LiDAR point cloud with the optical point cloud. Figure 8 shows the registration results of each target's LiDAR point cloud and optical point cloud, verifying that the proposed method is also applicable in the registration of LiDAR and the optical color point cloud. LiDAR point clouds are colored in white, while optical point clouds are colored in their true colors.
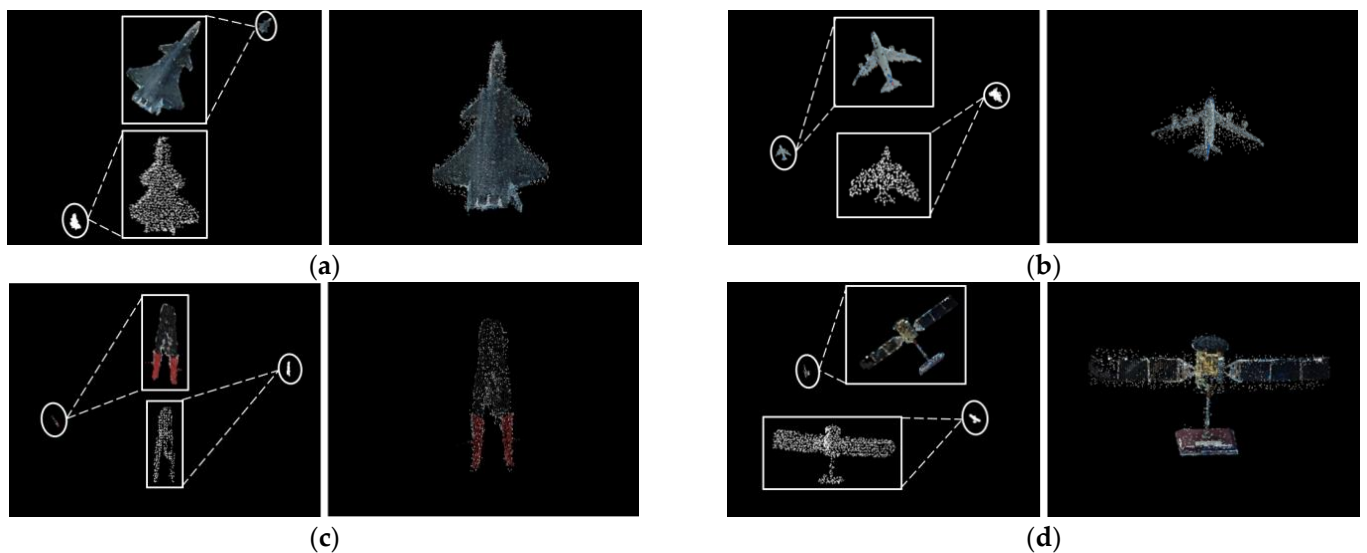

Figure 8. Optical point clouds and LiDAR point clouds before and after registration. (**a**) Aircraft model 1 before and after registration; (**b**) aircraft model 2 before and after registration; (**c**) pincer before and after registration; (**d**) satellite model before and after registration.

4.3. SAR–Camera–LiDAR Data Fusion Results

The proposed multimodal fusion framework utilizes pairwise registration results to unify the near-field 3D SAR image and optical color point cloud into the LiDAR coordinate system, and then obtains the fusion results by attaching color and scattering intensity to the aligned point cloud. As shown in Figure 9, the near-field SAR image of each target and the corresponding SAR–camera–LiDAR data fusion results are presented.
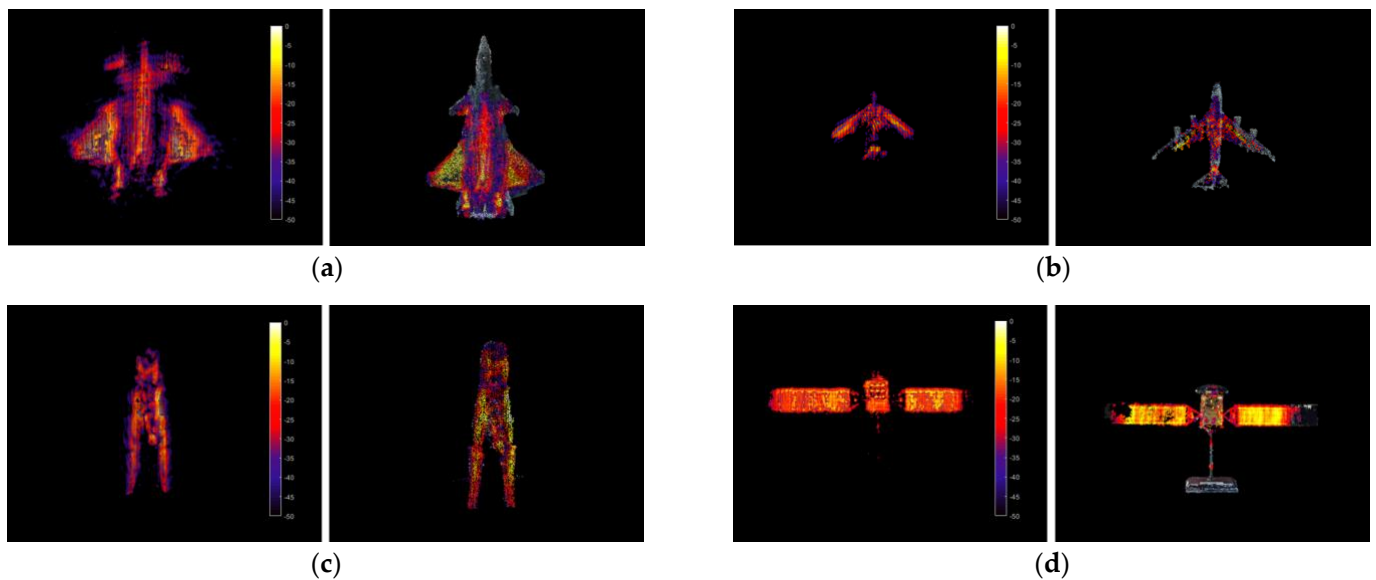

Figure 9. Near-field 3D SAR images and corresponding multimodal fusion results. (**a**) Aircraft model 1 before and after multimodal fusion; (**b**) aircraft model 2 before and after multimodal fusion; (**c**) pincer before and after multimodal fusion; (**d**) satellite model before and after multimodal fusion.

By integrating the precise geometric sizes of LiDAR point clouds and the color information of optical color point clouds, scattering characteristics can be accurately located and target categories can be intuitively determined. Multi-sensor data fusion not only reduces the difficulty of SAR scattering characteristics diagnosis, but it also improves the efficiency of SAR image interpretation. In the fusion image, it can be clearly seen that the scattering at the head of the aircraft model 1 is weak, while the scattering on both wings is strong. The scattering in the middle part of the passenger plane model is strong, while the scattering in the head and tail is weak. The pincer has high scattering characteristics in all parts and the SAR imaging contour is clear. The two wings of the satellite model have strong scattering, and there is also scattering at the vertical connecting rod. However, from the satellite model fusion image, it can also be seen that some of the satellite SAR scattering characteristics are lost due to the holes in the optical color point cloud generated during SFM reconstruction. Therefore, high-quality optical color point clouds need to be obtained in the future.

4.4. Multimodal Fusion Application Experiment

In order to demonstrate the advantages of multimodal data fusion in near-field SAR applications, application experiments have been conducted for concealed target detection and fault detection.

Figure 10 shows the experimental results of concealed target detection. Aircraft model 1 and aircraft model 2 hidden in a cardboard box are placed in the experimental scene. The millimeter wave near-field array 3D SAR imaging system can penetrate the cover (cardboard box) to image hidden targets (passenger plane model).
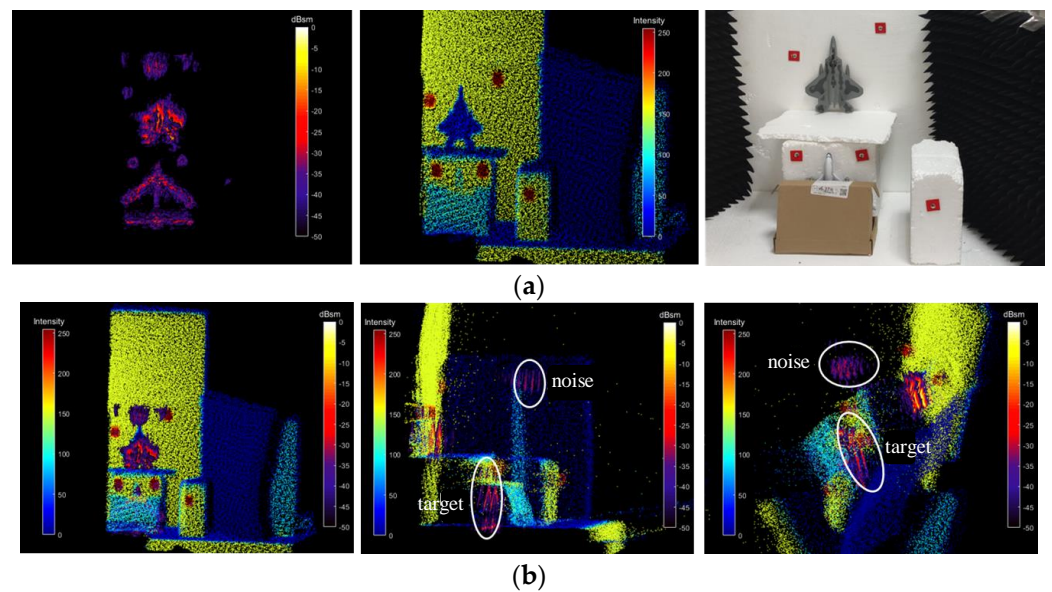
Figure 10. Application experiment of concealed target detection. (**a**) Near-field SAR image, LiDAR point cloud, and optical image of the experiment scene; (**b**) front view, left view, and top view of the fusion image of near-field SAR and LiDAR.

The results in Figure 10 show that, generally, noise often appears in 3D SAR images, which can mistakenly be considered as a target. However, by fusing the results from LiDAR, the proposed method can accurately locate the true position of the target and avoid including incorrect positions. Traditionally, this judgement relies on manual expert experience. The results also show that even a sheltered target can be detected. The LiDAR can observe the shelter wherein hidden targets are located, but cannot observe the hidden targets. The SAR–LiDAR fusion result can clearly depict the positions of hidden targets relative to the cover, which is beneficial for the application of concealed target detection.

Figure 11 shows the results of the fault detection experiment. Fault detection usually refers to the scattering enhancement of the target fault area, which is detected by the near-field 3D SAR imaging system [50]. It mainly finds possible fault areas by comparing the imaging results of the non-fault target and the fault target. The scattering from the head of aircraft model 1 is weak, indicating good stealth performance. Therefore, as shown in Figure 11a, a rivet is placed on the head of aircraft model 1 as the fault target. The scattering intensity of the rivet is relatively high, simulating targets with degraded stealth performance (with faults). In the results of radar in Figure 11b, a challenge due to the unique scattering characteristics of the target exists. Specifically, the head and nose parts of the aircraft model are missed when comparing the results from other sensors. This presents a difficulty related to identifying the part of the aircraft with a stealth performance fault. This identification is critical in determining the severity of the fault and planning appropriate solutions. However, a combination of other data helps to overcome this hurdle. Precise information on the physical structure of an aircraft can be obtained using shape and position data provided by LiDAR. This is further supplemented by the colorful texture information gathered from the camera, which provides a more detailed and visually rich representation of the aircraft's surface. This integrated approach makes the identification and location of faults considerably more straightforward and accurate, enhancing the overall effectiveness of our inspection and maintenance process. As shown in Figure 11c, the fault location can be located precisely and more intuitively by comparing the SAR–LiDAR fusion result. And Figure 11d shows the fault can be identified more intuitively by the LiDAR–camera fusion result.
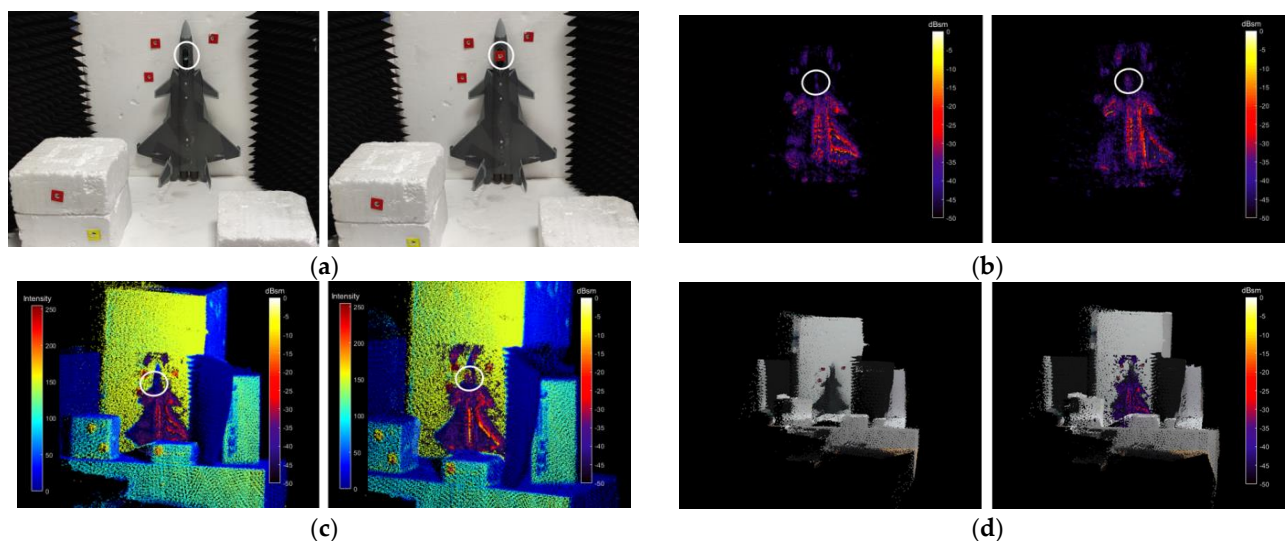
Figure 11. Application experiment of fault detection. (**a**) Optical image layout for fault detection experiment (left—without fault, right—with fault); (**b**) near-field 3D SAR imaging results (left—without fault, right—with fault); (**c**) near-field 3D SAR–LiDAR fusion results (left—without fault, right—with fault); (**d**) LiDAR—camera fusion result and multimodal fusion result without fault. The white circles in the figure indicate where the faults are set in the experiment.

5. Conclusions

This work employs multimodal data fusion for the first time to enhance the perception ability of near-field 3D SAR, which leverages the complementary strengths of multiple sensors (LiDAR's precise object localization and camera's color information). To address the difficulty in aligning the coordinate system related to data formats, noise, and resolution differences during the data fusion of SAR–camera–LiDAR, a three-step coarse-to-fine point cloud registration method is designed for our multimodal fusion framework. This method begins with a CED key point extraction algorithm with structure–intensity dual constraints, proposed to extract key points for subsequent registration. Next, the coarse registration step integrates SHOT feature–geometric relationship dual constraints into the SAC-IA algorithm to generate a rough space transformation matrix to provide a better initial pose. The subsequent fine registration leverages an ICP fine registration algorithm with adaptive thresholds, achieving the precise alignment of multi-sensor point clouds through an accurate space transformation matrix. The experimental results demonstrate that the proposed method achieves a state-of-the-art registration result in both quantitative and qualitative measurements, showing promising potential for advanced applications such as RCS measurement and concealed object detection in near-field 3D SAR scenarios.

Regarding the limitations of our current work, it is noted that near-field 3D SAR and LiDAR point clouds can be obtained via single perspective measurements. This restricts their ability to comprehensively perceive and interpret scenes. Hence, future work will explore reconstruction of multi-view near-field 3D SAR point clouds and corresponding multi-sensor data fusion methods to improve modeling and perception. Moreover, current learning-based methods have demonstrated impressive performance in handling point cloud data, such as feature description and matching [51]. The next phase of our fusion framework will adopt these learning-based processing methods, replacing the existing ones.

Author Contributions: Conceptualization, W.Z.; methodology, W.Z.; software, B.W.; validation, Z.L. and B.W.; formal analysis, X.X.; investigation, X.X. and Z.L.; resources, X.X.; data curation, W.Z. and Z.L.; writing—original draft preparation, W.Z., X.Z. (Xu Zhan) and T.Z.; writing—review and editing, T.Z.; visualization, T.Z.; supervision, T.Z.; project administration, X.Z. (Xiaoling Zhang); funding acquisition, X.Z. (Xiaoling Zhang). All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Dataset available on request from the authors on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, Z.; Wang, J.; Wu, J.; Liu, Q.H. A Fast Radial Scanned Near-Field 3-D SAR Imaging System and the Reconstruction Method. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1355–1363. [CrossRef]
2. Xu, X.; Zhang, X.; Zhang, T. Lite-YOLOv5: A Lightweight Deep Learning Detector for On-Board Ship Detection in Large-Scene Sentinel-1 SAR Images. *Remote Sens.* **2022**, *14*, 1018. [CrossRef]
3. Xu, X.; Zhang, X.; Shao, Z.; Shi, J.; Wei, S.; Zhang, T.; Zeng, T. A Group-Wise Feature Enhancement-and-Fusion Network with Dual-Polarization Feature Enrichment for SAR Ship Detection. *Remote Sens.* **2022**, *14*, 5276. [CrossRef]
4. Xu, X.; Zhang, X.; Zhang, T.; Yang, Z.; Shi, J.; Zhan, X. Shadow-Background-Noise 3D Spatial Decomposition Using Sparse Low-Rank Gaussian Properties for Video-SAR Moving Target Shadow Enhancement. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
5. Xu, Y.; Zhang, X.; Wei, S.; Shi, J.; Zeng, T.; Zhang, T. A Target-Oriented Bayesian Compressive Sensing Imaging Method with Region-Adaptive Extractor for mmW Automotive Radar. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–16. [CrossRef]
6. Wang, Y.; Zhang, X.; Zhan, X.; Zhang, T.; Zhou, L.; Shi, J.; Wei, S. An RCS Measurement Method Using Sparse Imaging Based 3-D SAR Complex Image. *IEEE Antennas Wirel. Propag. Lett.* **2022**, *21*, 24–28. [CrossRef]
7. Chen, X.; Luo, C.; Yang, Q.; Yang, L.; Wang, H. Efficient MMW Image Reconstruction Algorithm Based on ADMM Framework for Near-Field MIMO-SAR. *IEEE Trans. Microw. Theory Tech.* **2023**, *72*, 1326–1338. [CrossRef]
8. Pu, L.; Zhang, X.; Shi, J.; Wei, S.; Zhang, T.; Zhan, X. Precise RCS Extrapolation via Nearfield 3-D Imaging with Adaptive Parameter Optimization Bayesian Learning. *IEEE Trans. Antennas Propag.* **2022**, *70*, 3656–3671. [CrossRef]
9. Shahian Jahromi, B.; Tulabandhula, T.; Cetin, S. Real-Time Hybrid Multi-Sensor Fusion Framework for Perception in Autonomous Vehicles. *Sensors* **2019**, *19*, 4357. [CrossRef]
10. Wang, Z.; Wu, Y.; Niu, Q. Multi-Sensor Fusion in Automated Driving: A Survey. *IEEE Access* **2020**, *8*, 2847–2868. [CrossRef]
11. Yeong, J.; Velasco-Hernandez, G.; Barry, J.; Walsh, J. Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review. *Sensors* **2021**, *21*, 2140. [CrossRef] [PubMed]
12. Chen, Y.; Bruzzone, L. Self-Supervised SAR-Optical Data Fusion of Sentinel-1/-2 Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]
13. Kulkarni, S.C.; Rege, P.P. Pixel level fusion techniques for SAR and optical images: A review. *Inf. Fusion* **2020**, *59*, 13–29. [CrossRef]
14. Li, W.; Gao, Y.; Zhang, M.; Tao, R.; Du, Q. Asymmetric Feature Fusion Network for Hyperspectral and SAR Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 8057–8070. [CrossRef] [PubMed]
15. Quan, Y.; Tong, Y.; Feng, W.; Dauphin, G.; Huang, W.; Xing, M. A Novel Image Fusion Method of Multi-Spectral and SAR Images for Land Cover Classification. *Remote Sens.* **2020**, *12*, 3801. [CrossRef]
16. Jiao, Z.; Qiu, X.; Dong, S.; Yan, Q.; Zhou, L.; Ding, C. Preliminary exploration of geometrical regularized SAR tomography. *ISPRS J. Photogramm. Remote Sens.* **2023**, *201*, 174–192. [CrossRef]
17. Li, Y.; Ibanez-Guzman, J. Lidar for Autonomous Driving: The Principles, Challenges, and Trends for Automotive Lidar and Perception Systems. *IEEE Signal Process. Mag.* **2020**, *37*, 50–61. [CrossRef]
18. Bijelic, M.; Gruber, T.; Mannan, F.; Kraus, F.; Ritter, W.; Dietmayer, K.; Heide, F. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11682–11692.
19. Qian, K.; Zhu, S.; Zhang, X.; Li, L.E. Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 444–453.
20. Wei, Z.; Zhang, F.; Chang, S.; Liu, Y.; Wu, H.; Feng, Z. MmWave Radar and Vision Fusion for Object Detection in Autonomous Driving: A Review. *Sensors* **2022**, *22*, 2542. [CrossRef]
21. Zhen, W.; Hu, Y.; Liu, J.; Scherer, S. A Joint Optimization Approach of LiDAR-Camera Fusion for Accurate Dense 3-D Reconstructions. *IEEE Robot. Autom. Lett.* **2019**, *4*, 3585–3592. [CrossRef]
22. Bai, Z.; Jiang, G.; Xu, A. LiDAR-Camera Calibration Using Line Correspondences. *Sensors* **2020**, *20*, 6319. [CrossRef]
23. Peng, F.; Wu, Q.; Fan, L.; Zhang, J.; You, Y.; Lu, J.; Yang, J.-Y. Street view cross-sourced point cloud matching and registration. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 2026–2030.

24. Li, P.; Wang, R.; Wang, Y.; Tao, W. Evaluation of the ICP Algorithm in 3D Point Cloud Registration. *IEEE Access* **2020**, *8*, 68030–68048. [CrossRef]

25. Jiang, S.; Jiang, C.; Jiang, W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 230–251. [CrossRef]

26. Mellado, N.; Dellepiane, M.; Scopigno, R. Relative Scale Estimation and 3D Registration of Multi-Modal Geometry Using Growing Least Squares. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 2160–2173. [CrossRef]

27. Shen, X.; Darmon, F.; Efros, A.A.; Aubry, M. RANSAC-Flow: Generic Two-Stage Image Alignment. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Lecture Notes in Computer Science; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 618–637.

28. Huang, X.; Zhang, J.; Wu, Q.; Fan, L.; Yuan, C. A coarse-to-fine algorithm for registration in 3D street-view cross-source point clouds. In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 30 November–2 December 2016; pp. 1–6.

29. Huang, X.; Zhang, J.; Fan, L.; Wu, Q.; Yuan, C. A Systematic Approach for Cross-Source Point Cloud Registration by Preserving Macro and Micro Structures. *IEEE Trans. Image Process.* **2017**, *26*, 3261–3276. [CrossRef]

30. Li, J.; Zhuang, Y.; Peng, Q.; Zhao, L. Pose Estimation of Non-Cooperative Space Targets Based on Cross-Source Point Cloud Fusion. *Remote Sens.* **2021**, *13*, 4239. [CrossRef]

31. Cheng, L.; Chen, S.; Liu, X.; Xu, H.; Wu, Y.; Li, M.; Chen, Y. Registration of Laser Scanning Point Clouds: A Review. *Sensors* **2018**, *18*, 1641. [CrossRef] [PubMed]

32. Ma, W.; Zhang, J.; Wu, Y.; Jiao, L.; Zhu, H.; Zhao, W. A Novel Two-Step Registration Method for Remote Sensing Images Based on Deep and Local Features. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4834–4843. [CrossRef]

33. Lahat, D.; Adali, T.; Jutten, C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [CrossRef]

34. Zhou, Z.; Wei, S.; Wang, M.; Liu, X.; Wei, J.; Shi, J.; Zhang, X. Comparison of MF and CS Algorithm in 3-D Near-Field SAR Imaging. In Proceedings of the 2021 7th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Bali, Indonesia, 1–3 November 2021; pp. 1–5.

35. Sun, Z.; Li, Z.; Liu, Y. An Improved Lidar Data Segmentation Algorithm Based on Euclidean Clustering. In Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019), Tianjing, China, 13–15 July 2019; Lecture Notes in Electrical Engineering. Springer: Singapore, 2020; pp. 1119–1130.

36. Pleansamai, K. M-Estimator Sample Consensus Planar Extraction from Image-Based 3d Point Cloud for Building Information Modelling. *Int. J. Geomate* **2019**, *17*, 69–76. [CrossRef]

37. Zeng, J.; Wang, D.; Chen, P. Improved color region growing point cloud segmentation algorithm based on octree. In Proceedings of the 2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China, 22–24 July 2022; pp. 424–429.

38. Arun, K.S.; Huang, T.S.; Blostein, S.D. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *PAMI-9*, 698–700. [CrossRef]

39. Teng, H.; Chatziparaschis, D.; Kan, X.; Roy-Chowdhury, A.K.; Karydis, K. Centroid Distance Keypoint Detector for Colored Point Clouds. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 1196–1205.

40. Chu, J.; Zhang, Y.; Li, S.; Leng, L.; Miao, J. Syncretic-NMS: A Merging Non-Maximum Suppression Algorithm for Instance Segmentation. *IEEE Access* **2020**, *8*, 114705–114714. [CrossRef]

41. Salti, S.; Tombari, F.; Di Stefano, L. SHOT: Unique signatures of histograms for surface and texture description. *Comput. Vis. Image Underst.* **2014**, *125*, 251–264. [CrossRef]

42. Li, W.; Cheng, H.; Zhang, X. Efficient 3D Object Recognition from Cluttered Point Cloud. *Sensors* **2021**, *21*, 5850. [CrossRef] [PubMed]

43. Rusu, R.B.; Blodow, N.; Beetz, M. Fast point feature histograms (FPFH) for 3D registration. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 3212–3217.

44. Song, M.; Tao, D.; Chen, C.; Li, X.; Chen, C.W. Color to gray: Visual cue preservation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1537–1552. [CrossRef] [PubMed]

45. Xue, S.; Zhang, Z.; Lv, Q.; Meng, X.; Tu, X. Point Cloud Registration Method for Pipeline Workpieces Based on PCA and Improved ICP Algorithms. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2019; p. 032188.

46. Yang, R.; Meng, X.; Yao, Y.; Chen, B.Y.; You, Y.; Xiang, Z. An analytical approach to evaluate point cloud registration error utilizing targets. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 48–56. [CrossRef]

47. Liu, W.; Sun, W.; Wang, S.; Liu, Y. Coarse registration of point clouds with low overlap rate on feature regions. *Signal Process. Image Commun.* **2021**, *98*, 116428. [CrossRef]

48. Yang, J.; Wang, C.; Luo, W.; Zhang, Y.; Chang, B.; Wu, M. Research on Point Cloud Registering Method of Tunneling Roadway Based on 3D NDT-ICP Algorithm. *Sensors* **2021**, *21*, 4448. [CrossRef]

49. Liu, W.; Wu, H.; Chirikjian, G.S. LSG-CPD: Coherent point drift with local surface geometry for point cloud registration. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 15293–15302.

50. Pham, T.H.; Kim, K.H.; Hong, I.P. A Study on Millimeter Wave SAR Imaging for Non-Destructive Testing of Rebar in Reinforced Concrete. *Sensors* **2022**, *22*, 8030. [CrossRef]
51. Bai, X.; Luo, Z.; Zhou, L.; Fu, H.; Quan, L.; Tai, C.L. D3Feat: Joint Learning of Dense Detection and Description of 3D Local Features. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; 2020.