

Article

Measurement-Oriented 3D Reconstruction and Attitude Estimation of Free-Tumbling Space Targets via Cooperative Multi-View Observation

Di Zhao ^{1,2}, Zhe Yue ^{1,*}, Wensong Zhang ^{3,4,*}, Jianping Yuan ², Weihua Ma ⁵, Haofei Ban ¹, Sen Li ¹ and Weiwei Lei ¹

¹ School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; zd@hpu.edu.cn (D.Z.); bhf@home.hpu.edu.cn (H.B.); 212104020069@home.hpu.edu.cn (S.L.); wwlei@hpu.edu.cn (W.L.)

² Binzhou Institute of Technology, Binzhou 256606, China; jyuan@nwpu.edu.cn

³ School of Intelligent Manufacturing, Tianmen Vocational College, Tianmen 431700, China

⁴ Aerospace Information Research Institute, Henan Academy of Sciences, Zhengzhou 450001, China

⁵ School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China; whma_npu@nwpu.edu.cn

* Correspondence: yuezhe@hpu.edu.cn (Z.Y.); zws556688@whu.edu.cn (W.Z.)

Abstract

Accurate attitude measurement of non-cooperative space targets is essential for on-orbit servicing, active debris removal, and autonomous rendezvous missions. To address the challenges associated with unknown geometry, rapid tumbling motion, and the limited observability of single-view systems, this study proposes a cooperative multi-view measurement framework for three-dimensional reconstruction and attitude estimation. Multiple spacecraft are deployed to form a stable observation configuration, and multi-view image sequences are acquired to strengthen geometric constraints. A learning-based multi-view stereo reconstruction module is used to estimate depth information and reconstruct point clouds, which are further processed through iterative closest point (ICP) registration to derive inter-frame attitude variations. An extended Kalman filter (EKF) is then introduced to improve temporal consistency and suppress measurement noise. Validation is conducted in a numerical simulation using a simplified Fengyun-1 (FY-1) satellite model under a three-spacecraft cooperative fly-around scenario. The simulation results demonstrate that the proposed method achieves high-precision attitude estimation, with attitude errors below 0.3° and positional errors within 0.05 m. Comparative experiments show that the method maintains stable measurement performance under varying observation distances and viewing configurations. The proposed framework provides a reliable and robust measurement solution for dynamic attitude determination of free-tumbling space targets.

Keywords: attitude measurement; multi-view reconstruction; cooperative observation; non-cooperative space target; measurement accuracy; point cloud registration; extended Kalman filter



Received: 19 May 2026

Revised: 16 June 2026

Accepted: 26 June 2026

Published: 27 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

With the increasing demand for on-orbit servicing and active debris removal, intelligent perception and autonomous operation of non-cooperative space targets have become key challenges in the aerospace field [1,2]. Such targets—typically defunct satellites, decommissioned spacecraft, or debris—often exhibit uncontrolled attitude states, unknown geometry,

strong surface reflectance, and complex motion patterns, which pose substantial challenges to spaceborne vision-based perception as well as autonomous rendezvous and capture [3,4].

Non-cooperative targets (e.g., failed satellites) usually retain considerable residual angular momentum in an uncontrolled state. Under the combined effects of gravity-gradient torque and structural coupling, these targets often exhibit nonlinear attitude evolution behaviors, such as periodic tumbling and nutation, with continuously varying rotation axes and angular velocities. These highly dynamic characteristics severely degrade the robustness of conventional vision-based attitude estimation methods. Feature-based or model-matching approaches often suffer from feature drift, tracking interruptions, or correspondence degeneration during continuous observations, whereas sparse 3D reconstruction strategies may fail to maintain sufficient point density and accuracy under rapid tumbling or weak surface texture [5]. Furthermore, single-spacecraft observation is limited by field-of-view occlusion, depth ambiguity in weak-texture regions, and rapid parallax decay at medium to long distances. These factors lead to a significant deterioration in 3D reconstruction accuracy, often preventing the precision required for dynamic in-orbit capture and closed-loop control from being achieved [6].

To overcome these limitations, extensive research has been conducted using traditional geometric methods, including structure-from-motion (SFM), shape-from-shading (SFS), and multi-baseline stereo reconstruction. Zhang et al. proposed a visible-sensor-based multi-view reconstruction pipeline using SFM and PMVS for dense point cloud generation of spacecraft or debris [7]. Zhang et al. explored the adaptation of mature computer-vision SFM + MVS (multi-view stereo) pipelines to satellite imagery, demonstrating the applicability of traditional geometric methods to remote sensing and space reconstruction tasks [8]. Chang et al. employed stereovision to estimate the relative pose of non-cooperative spacecraft, thereby validating the effectiveness of multi-view triangulation [9]. Uno et al. used an SFM approach to recover the motion trajectory and 3D structure of unknown space debris [10]. Although these methods can provide high-quality reconstruction under clear texture and smooth attitude motion, their strong dependence on feature continuity and strict correspondence makes them vulnerable to drift, occlusion, and matching degeneration under fast tumbling, which frequently leads to unstable registration or reconstruction failure [9,10].

With the development of deep learning, several studies have explored CNN-based end-to-end pose regression, including PoseNet-like single-image methods and CNN-PnP approaches that incorporate geometric constraints. Park et al. proposed a CNN + PnP strategy that improves the robustness of non-cooperative spacecraft pose estimation [11]. Chen et al. introduced deep landmark regression combined with nonlinear geometric refinement to enhance pose accuracy [12]. Qiao et al. fused 2D CNN features with 3D satellite point clouds through a hybrid point-cloud convolution [13]. Rondão et al. adopted a CNN + LSTM temporal modeling scheme to enable sequential attitude estimation with multimodal robustness [14]. Ye et al. integrated CNN features with transformer-based architectures and solved poses using PnP + RANSAC [15]. Although these methods improve robustness in low-texture scenarios, they remain constrained by the inherent geometric limitations of single- or dual-view depth inference. Consequently, they struggle to maintain stable accuracy when targets undergo large-amplitude, high-frequency attitude variations or severe texture degradation.

To address limited geometric constraints and insufficient depth cues, multi-spacecraft cooperative observation has emerged as a promising direction for autonomous space perception [16,17]. By leveraging cross-view and cross-orbit collaborative imaging geometries, multi-spacecraft systems can provide richer spatiotemporal information and a substantially

larger effective baseline, thereby enhancing observability, improving reconstruction quality, and stabilizing attitude estimation, particularly for highly dynamic free-tumbling targets.

Meanwhile, deep-learning-based multi-view stereo networks, represented by MVSNet, have achieved significant progress in robotics and autonomous systems [18–20]. These networks exploit deep CNN features and cross-view geometric consistency to construct 3D cost volumes and infer high-quality depth maps, thereby mitigating challenges such as weak texture and parallax degradation [21,22]. When combined with cooperative observation, MVS-based reconstruction can generate dense and consistent 3D point clouds even under highly dynamic observation conditions, providing a reliable basis for continuous attitude estimation. The integration of ICP registration and EKF-based filtering further improves the continuity, robustness, and real-time capability of attitude determination.

Motivated by these observations, this study proposes an intelligent 3D reconstruction and attitude determination framework for free-tumbling space targets by integrating multi-spacecraft cooperative observation with deep-learning-based multi-view stereo reconstruction. The proposed method exploits cross-view geometric information from the spacecraft formation to generate accurate depth maps via MVSNet, reconstructs high-fidelity point clouds, and estimates continuous attitude parameters using ICP combined with EKF fusion. The objective is to achieve stable, high-precision attitude estimation under weak-texture, limited-viewpoint, and rapid-rotation conditions, ultimately providing an effective and real-time measurement solution for autonomous on-orbit rendezvous and capture missions.

The main contributions of this study are summarized as follows. First, a cooperative multi-view measurement framework is developed for free-tumbling non-cooperative space targets by integrating formation observation, dense 3D reconstruction, point cloud registration, and filtering-based attitude fusion into a unified pipeline. Second, an MVSNet-based dense reconstruction strategy is introduced to improve the geometric completeness and robustness of target point clouds under multi-view dynamic observation, reducing the dependence on hand-crafted feature matching. Third, the proposed cooperative observation strategy is compared with single-spacecraft trinocular stereo vision and traditional SFM-based reconstruction, clarifying the roles of effective baseline, viewing-angle diversity, and learning-based depth inference in improving dynamic attitude estimation accuracy.

2. Materials and Methods

2.1. Measurement Framework

The objective of this study is to achieve accurate three-dimensional reconstruction and dynamic attitude determination of a free-tumbling non-cooperative space target by integrating multi-satellite cooperative observation with deep-learning-based multi-view stereo perception. The experimental design aims to validate the effectiveness, robustness, and real-time capability of the method under realistic orbital motion, observation geometry, and target tumbling conditions.

In this study, it is assumed that the relative positions and attitudes of the observing satellites are available through onboard navigation systems, and that synchronized image sequences of the target can be acquired from multiple viewpoints. These relative pose and synchronization quantities are treated as external measurement inputs for the reconstruction module. In practical missions, navigation errors, camera calibration errors, time-synchronization errors, and pointing errors would perturb the homography warping process, introduce depth-map bias, and further affect point cloud registration and EKF-based fusion. Therefore, the present validation focuses on the nominal cooperative measurement pipeline, whereas the influence of these uncertainties is discussed as a limitation and a topic for future uncertainty-propagation analysis. Based on these assumptions, the overall experimental procedure is divided into the following four stages.

Phase 1: Cooperative observation configuration: A stable multi-spacecraft formation is established around the free-tumbling target to provide complementary viewing geometries while satisfying imaging and visibility constraints.

Phase 2: Multi-view 3D reconstruction: Synchronized multi-view image sequences are processed using a CNN-based multi-view stereo framework to estimate depth maps and reconstruct the target point cloud.

Phase 3: Frame-to-frame attitude estimation: Consecutive point clouds are registered using an iterative closest point (ICP) algorithm to extract relative pose changes and obtain preliminary attitude measurements.

Phase 4: Dynamic attitude refinement: The estimated attitude measurements are fused using an extended Kalman filter (EKF) to achieve smooth and accurate attitude tracking over time.

The overall experimental workflow, including data acquisition, reconstruction, pose estimation, and filtering, is illustrated in Figure 1.

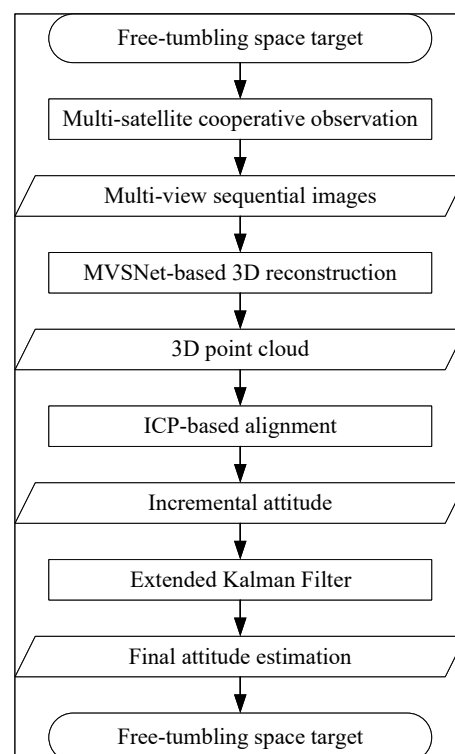


Figure 1. Process framework of the proposed method. Rounded rectangles denote the target and final output, rectangles denote processing modules, and parallelograms denote input/output data.

2.2. Characteristics of Free-Tumbling Targets and Cooperative Observation Model

2.2.1. Attitude Characteristics of Free-Tumbling Space Targets

For space targets in an uncontrolled state, the evolution of their attitude is primarily influenced by inertial torques, residual angular momentum, and external disturbances, such as Earth's gravity-gradient torque and solar radiation pressure. The attitude kinematics of the target can be described by

$$\dot{q} = \frac{1}{2} Q(q) \omega_b \quad (1)$$

where $q = [q_0, q_1, q_2, q_3]^T$ denotes the attitude quaternion and $\omega_b = [\omega_x, \omega_y, \omega_z]^T$ is the body-frame angular velocity.

The corresponding attitude dynamics are given by

$$J\omega_b + \omega_b \times J\omega_b = \tau \quad (2)$$

where J is the inertia tensor and τ represents external disturbance torques.

Because the principal inertia axes of the target spacecraft may not be aligned with its instantaneous rotation axis, nutation frequently occurs during the rotational motion. This phenomenon results in complex free-tumbling behavior, as illustrated in Figure 2. The resulting attitude variations were highly nonlinear and difficult to estimate.

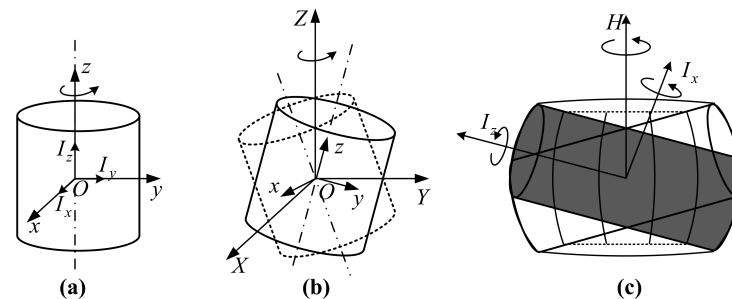


Figure 2. Motion modes of a free-tumbling target. (a) Rotation; (b) Nutation; (c) Tumbling.

Furthermore, traditional feature-based projection models are generally inapplicable, as prior information on the target's geometric and attitude parameters is unavailable and no cooperative features can be utilized. The complex and rapidly changing rotational motion introduces additional challenges, as single-spacecraft observations often suffer from field-of-view occlusion and severe depth ambiguity.

2.2.2. Multi-Spacecraft Cooperative Observation Scheme

To improve the observability and 3D reconstruction accuracy of free-tumbling targets, this study employs a cooperative observation scheme using a formation of spacecraft. As shown in Figure 3, multiple observer spacecraft are commanded to fly around the target in formation along the same periodic relative orbit, thereby maintaining long-term stability of the observation geometry.

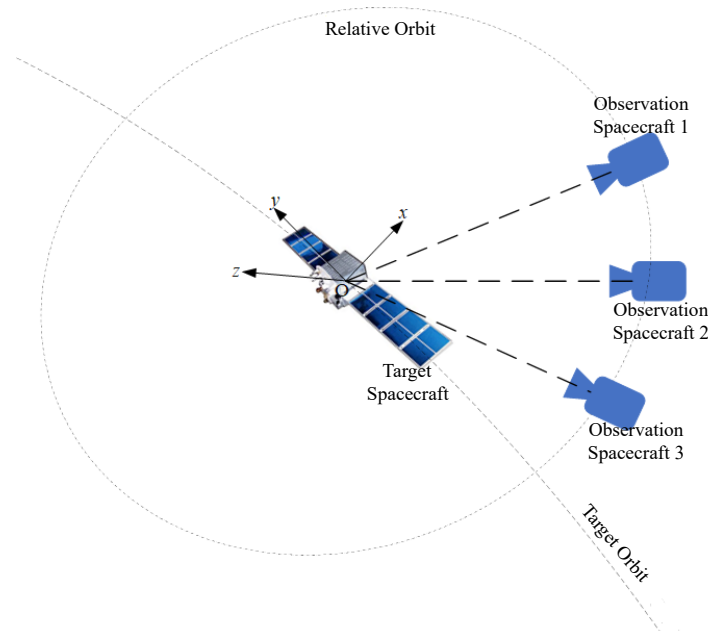


Figure 3. Multi-spacecraft cooperative observation formation. The circular curve represents the relative orbit of the observing spacecraft around the target, and the dashed lines indicate the corresponding lines of sight.

The relative motion between the target and the observer spacecraft can be modeled using the Clohessy–Wiltshire (CW) equations:

$$\begin{cases} \ddot{x} - 3n^2x - 2n\dot{y} = 0 \\ \ddot{y} + 2n\dot{x} = 0 \\ \ddot{z} + n^2z = 0 \end{cases} \quad (3)$$

which can be rearranged into the commonly used periodic solution form

$$\mathbf{X}(t) = \mathbf{\Phi}(t)\mathbf{X}(0) \quad (4)$$

and the corresponding relative orbital geometry for spacecraft i is parameterized as

$$\begin{cases} x_i = A_0 \cos(nt + \alpha + \Delta\theta_i) \\ y_i = -2A_0 \sin(nt + \alpha + \Delta\theta_i) \\ z_i = B_0 \cos(nt + \beta + \Delta\theta_i) \end{cases} \quad (5)$$

Accordingly, the relative orbital parameters of the observer spacecraft formation are designed as:

$$\mathbf{X}_p = (A_0, B_0, n, \alpha, \beta, \Delta\theta_i, y_{c0}) \quad (6)$$

where n is the mean motion of the reference orbit, y_{c0} is the cross-track bias (the y -axis offset of the fly-around orbit);

$$A_0 = \sqrt{\left(\frac{\dot{x}_0}{n}\right)^2 + x_0^2} \quad (7)$$

represents the amplitude of the periodic motion within the orbital plane of the reference spacecraft (the xy -plane), and

$$B_0 = \sqrt{\left(\frac{\dot{z}_0}{n}\right)^2 + z_0^2} \quad (8)$$

denotes the amplitude of the simple harmonic oscillation perpendicular to the orbital plane (in the z -direction). The parameters α and β are the initial phase angles of the relative motion in the xy -plane and the z -direction, respectively. Together, these parameters determine the fundamental geometric configuration of the relative orbit. $\Delta\theta_i$ is the phase offset of observer spacecraft i with respect to the initial phase angle, which specifies its relative position within the formation along the designed orbit.

The cooperative fly-around observation is subject to the following constraints:

1. **Maximum observation distance constraint.** The fly-around radius must remain below the maximum effective imaging distance to ensure that the target occupies a sufficient projection area on each observer's image plane.
2. **Minimum observation distance constraint.** The fly-around radius must exceed the minimum safe distance required to capture the entire target within the camera's field of view (FOV):

$$d_{\min} > \lambda \frac{R_{\text{target}}}{\tan(\phi_{\text{FOV}}/2)}, \quad \lambda \approx 1 \quad (9)$$

where λ is a scaling factor related to the target operating region, and ϕ_{FOV} is the camera field-of-view angle.

3. **Inter-satellite viewing-angle constraint.** To ensure reliable feature correspondence between different viewing angles, the viewing-angle separation between any two spacecraft must remain below the inter-view matching threshold.
4. **Field-of-view constraint.** The target must remain inside each camera's FOV throughout the observation window; therefore, the angle between the camera optical axis and the line of sight to the target must remain less than half of the camera FOV.

5. **Long-duration feasibility.** The formation must satisfy the above constraints for a sufficiently long observation interval to ensure the availability of continuous, high-quality measurement data.

In the nominal simulation, the camera optical axes are commanded to point toward the target centroid at each imaging epoch, and the target is kept within the camera FOV before image acquisition. The optical-axis pointing process is therefore represented as an ideal target-tracking constraint rather than as a detailed attitude-control loop. Small platform jitter, pointing-rate limits, navigation errors, and calibration errors are not explicitly injected into the image-rendering process; their effects would mainly appear as relative-pose perturbations, image blur, or inter-view inconsistency, which may degrade depth estimation and point cloud registration.

Based on these constraints, a feasible relative-dynamics model and formation-configuration parameter set can be designed. Once the relative trajectories and camera attitudes of all spacecraft are computed, multi-view sequential images of the free-tumbling target are collected. These time-varying images and the associated spacecraft pose information collectively form a cooperative observation dataset, which enables robust estimation of the target's 3D reconstruction and attitude parameters.

2.3. Intelligent 3D Reconstruction Based on MVSNet

Traditional 3D reconstruction methods primarily rely on geometric matching and photometric consistency assumptions, such as normalized cross-correlation (NCC) and semi-global matching (SGM). Although such methods can achieve high accuracy under ideal imaging conditions, they are computationally expensive and highly sensitive to illumination variations, texture degradation, and occlusion, making them unsuitable for real-time on-orbit applications.

In contrast, MVSNet employs an end-to-end learning strategy that takes multi-view images as input and directly regresses the depth through 2D feature extraction, homography warping, cost-volume construction, and 3D convolution. This process yields dense or sparse 3D point cloud representations. Compared to traditional approaches, MVSNet automatically learns cross-view feature correspondences, avoids manual feature engineering and complex optimization, and demonstrates strong robustness to illumination changes and noise. In this work, MVSNet was incorporated to enable intelligent and efficient 3D reconstruction of free-tumbling space targets for attitude determination.

2.3.1. Architecture and Principles of MVSNet

As illustrated in Figure 4, the multi-view stereo reconstruction network (MVSNet) is an end-to-end architecture that takes one reference image and several source-view images as input. A 2D convolutional neural network extracts feature maps for each image. Given the relative camera poses, homography warping was applied to construct the 3D cost volume in the reference camera frustum. A 3D CNN then regularizes the cost volume to infer an initial depth map, which is further refined by a depth-refinement module to ensure sharper boundaries and reduce local noise, yielding the final high-precision depth estimation.

The reconstruction process consists of four main stages:

(1) Feature Extraction

For each input image, an 8-layer 2D convolutional neural network extracts the following feature map:

$$F_i = F(I_i; \theta) \quad (10)$$

where $F(\cdot)$ is a shared-weight CNN, θ denotes the learnable parameters.

The resulting feature maps have (1/8) of the original resolution and 32 channels. Each layer employs ReLU activation and batch normalization to improve the convergence and stability.

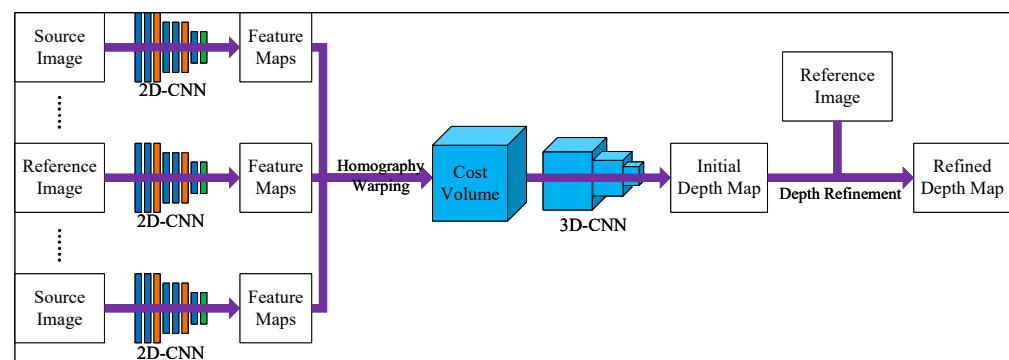


Figure 4. Architecture of the MVSNet framework. The arrows indicate the data flow among the feature extraction, cost-volume construction, 3D CNN regularization, and depth regression modules, while different colors are used to distinguish the main functional components of the network.

(2) Homography Warping and Cost-Volume Construction

Given the intrinsic and extrinsic parameters of each view, homography warping aligns the 32-channel source-view feature maps with the depth hypotheses of the reference-view frustum. The homography matrix for the depth hypothesis (d) and source view (i) is expressed as

$$H_i(d) = I \cdot K_i \cdot R_i - \frac{(t_i - t_r) \cdot n_r^T}{d} \cdot K_i \cdot R_i \cdot R_r \cdot K_r \quad (11)$$

where, K_i, K_r is the camera intrinsic matrix, R_i, t_i, R_r, t_r are the rotations and translations of the reference and source cameras, n_r is the reference camera optical axis. d denotes the depth and I denotes the identity matrix.

Warping with $H_i(d)$ aligns source-view features to the depth plane d , forming the feature cube of dimensions: $W/8 \cdot H/8 \cdot D \cdot F$, where W, H are image dimensions, D is the number of sampled depth hypotheses, and $F = 32$ is the number of feature channels.

The individual feature cubes from all source views were aggregated to form the cost volume:

$$C = \frac{1}{N} \sum_{i=1}^N (V_i - \bar{V}_i) \quad (12)$$

where $\bar{V}_i$ denotes the mean feature cube. The constructed cost volume characterizes the correspondence consistency across all the sampled depth hypotheses.

(3) 3D Convolution and Depth Refinement

The constructed cost volume is processed by a four-level encoder–decoder 3D CNN to regularize the cross-view geometric cues and construct a per-pixel depth probability distribution. The estimated depth for pixel p was calculated using the following expectation:

$$D(p) = \sum_{d=d_{\min}}^{d_{\max}} d \times P(d) \quad (13)$$

where $P(d)$ denotes the predicted probability at depth d .

The depth estimation produced by the network is highly consistent with the traditional argmax strategy. Since the depth hypotheses in the cost-volume construction are uniformly sampled within the range $[d_{\min}, d_{\max}]$, the predicted depth values form a continuous distribution. This continuity ensures differentiability, which in turn enables the end-to-end

optimization of the entire network. The initial depth map has the same spatial resolution as the 2D feature maps, i.e., $W/8 \cdot H/8$.

(4) Depth Refinement

Since the initial depth map may contain boundary blur and local noise, a residual-refinement CNN was applied. The refinement network uses the reference RGB image and the initial depth map as a four-channel input:

$$D_{\text{refined}} = G(I_r, D) \quad (14)$$

where G is a 2D residual-convolution network.

2.3.2. Training Strategy and Loss Function

To achieve a high-precision depth estimation, the loss function integrates the mean absolute error (MAE) of the initial and refined depth maps:

$$Loss = \sum_{p \in p_{\text{valid}}} (\|D(p) - D^*(p)\| + \lambda \|D_{\text{refined}}(p) - D^*(p)\|) \quad (15)$$

where D^* is the ground-truth depth, p_{valid} denotes the set of valid pixels, and λ is the weighting coefficient.

Training used the Adam optimizer with an initial learning rate of 10^{-4} and a batch size of 4. In each reconstruction step, one reference view and two source views from the three-spacecraft observation sequence were used. The depth hypothesis range was set according to the near and far bounds of the rendered target in the reference-camera frustum so that the complete target was covered by the sampled depth planes. To ensure real-time applicability in on-orbit scenarios, a compact cost volume with $D = 128$ depth hypotheses was used to balance accuracy and computational efficiency. The weighting coefficient in the refinement loss was set to $\lambda = 1$, assigning equal importance to the initial and refined depth-map errors.

2.3.3. Point Cloud Reconstruction

An example of the reconstructed depth map of the Fengyun-1 (FY-1) satellite model is shown in Figure 5, where the color value represents the simulated depth in millimeters.

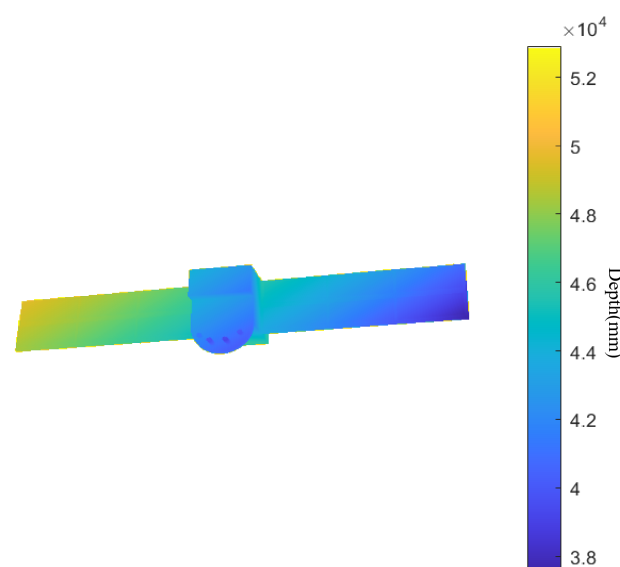


Figure 5. Depth map reconstruction of the FY-1 free-tumbling target.

Each depth pixel corresponds to a 3D point in the reference camera frame according to

$$\begin{cases} x = \frac{z_e(u - u_0)}{f_x}, \\ y = \frac{z_e(v - v_0)}{f_y}, \\ z = z_c \end{cases} \quad (16)$$

where z_e is the depth value at pixel $p(u, v)$, (u_0, v_0) is the principal point, f_x, f_y are the focal lengths in the horizontal and vertical directions, respectively.

By back-projecting all depth pixels, a full 3D point cloud of the target model was obtained, forming the basis for subsequent attitude estimation.

2.4. Attitude Determination Based on Point Cloud Registration and Filter Fusion

After reconstructing multi-view 3D point clouds, the point cloud sequences acquired at different time steps encode the target's continuous attitude evolution. Since a free-tumbling target undergoes ongoing rotational motion, its attitude angles (roll, pitch, and yaw) vary rapidly over time. To achieve high-precision, continuous attitude estimation, this study integrates point-cloud-based registration with extended Kalman filtering (EKF) for dynamic fusion and refinement.

2.4.1. Inter-Frame Point Cloud Registration Using ICP

The Iterative Closest Point (ICP) algorithm is one of the classical methods for point cloud registration. Its core idea is to iteratively establish nearest-neighbor correspondences between two point sets and solve for the optimal transformation parameters (rotation and translation) that minimize alignment error.

Consider two consecutive point clouds for the same target: $P_t = \{p_i\}_{i=1}^n$ at time t and $P_{t+\Delta t} = \{q_i\}_{i=1}^n$ at time $t + \Delta t$.

Assuming partial overlap, the goal is to find the rotation $\mathbf{R}$ and translation $\mathbf{T}$ such that

$$F(\mathbf{R}, \mathbf{T}) = \sum_{i=1}^m \|q_i - (\mathbf{R}p_i + \mathbf{T})\|^2 \quad (17)$$

is minimized.

Here, set $\{q_i\}$ consists of the nearest-neighbor points of $\{p_i\}$ determined via the Euclidean distance, $F(\mathbf{R}, \mathbf{T})$ is the registration cost function.

The ICP procedure is iterative:

1. Initialize the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{T}$.
2. Select point subsets from the frames.
3. Transform the current-frame point cloud using the current $\mathbf{R}$ and $\mathbf{T}$.
4. Establish correspondences via nearest-neighbor matching.
5. Update $\mathbf{R}$ and $\mathbf{T}$ using least-squares optimization to minimize $F(\mathbf{R}, \mathbf{T})$.
6. Check convergence; if the convergence criterion is not satisfied, repeat Step 3.
7. Output the final $\mathbf{R}$ and $\mathbf{T}$ as the registration result.

The final $\mathbf{R}, \mathbf{T}$ matrices quantify the inter-frame rotational and translational change of the target, providing raw attitude-motion measurements for subsequent filtering.

2.4.2. Extended Kalman Filter for Attitude Fusion

For a free-tumbling target with unknown inertia ratios and rapidly varying angular velocity, the motion between time t and $t + \Delta t$ can be modeled as a relative attitude transformation. The inter-frame attitude quaternion is defined as:

$$\Delta \mathbf{q}_{t,t+\Delta t} = \mathbf{q}_{t+\Delta t} \otimes \mathbf{q}_t^{-1} \quad (18)$$

The relative attitude dynamics can be expressed by

$$\Delta \dot{\mathbf{q}}_{t,t+\Delta t} = \frac{1}{2} \Omega(\Delta \boldsymbol{\omega}_{t,t+\Delta t}) \Delta \mathbf{q}_{t,t+\Delta t} \quad (19)$$

where $\boldsymbol{\omega}_{t,t+\Delta t} = \boldsymbol{\omega}_{t+\Delta t} - \boldsymbol{\omega}_t$ denotes the angular-velocity increment between t and $t + \Delta t$.

Based on the attitude dynamics and available measurements of the target, the state vector is defined as follows:

$$\mathbf{X} = \left(\Delta \mathbf{q}_{t,t+\Delta t}^T \quad \boldsymbol{\omega}_t^T \quad \boldsymbol{\omega}_{t+\Delta t}^T \quad \mathbf{p}^T \right)^T \quad (20)$$

where $\mathbf{p} = (p_x, p_y, p_z)^T$ denotes the inertia-ratio parameters, defined as

$$\begin{cases} p_x = \frac{J_{Ty} - J_{Tz}}{J_{Tx}}, \\ p_y = \frac{J_{Tz} - J_{Tx}}{J_{Ty}}, \\ p_z = \frac{J_{Tx} - J_{Ty}}{J_{Tz}}. \end{cases} \quad (21)$$

where J_{Tx}, J_{Ty}, J_{Tz} represent the principal moments of inertia of the free-tumbling target. Neglecting the external disturbance torques, the corresponding state-estimation model of the system can be written as follows:

$$\begin{cases} \Delta \dot{\mathbf{q}}_{t,t+\Delta t} = \frac{1}{2} \Omega(\Delta \boldsymbol{\omega}_{t,t+\Delta t} - \Delta \boldsymbol{\omega}_t) \Delta \mathbf{q}_{t,t+\Delta t}, \\ \dot{\boldsymbol{\omega}}_t = \begin{bmatrix} p_x \omega_{t,y} \omega_{t,z} \\ p_y \omega_{t,x} \omega_{t,z} \\ p_z \omega_{t,x} \omega_{t,y} \end{bmatrix}, \\ \dot{\boldsymbol{\omega}}_{t+\Delta t} = \begin{bmatrix} p_x \omega_{t+\Delta t,y} \omega_{t+\Delta t,z} \\ p_y \omega_{t+\Delta t,x} \omega_{t+\Delta t,z} \\ p_z \omega_{t+\Delta t,x} \omega_{t+\Delta t,y} \end{bmatrix}, \\ \dot{\mathbf{p}} = \mathbf{0} \end{cases} \quad (22)$$

The measurement is given by the target's attitude quaternion:

$$\mathbf{z} = \mathbf{h}(\mathbf{X}) = \begin{bmatrix} \Delta \mathbf{q}_{t,t+\Delta t} \\ \mathbf{0} \end{bmatrix} + \mathbf{v} \quad (23)$$

Based on the above state variables, state equations, and measurements, an Extended Kalman Filter (EKF) was constructed to estimate the target's attitude parameters. The EKF procedure is summarized as follows:

1. State prediction:

$$\hat{\mathbf{X}}_{k|k-1} = \mathbf{f}(\hat{\mathbf{X}}_{k-1|k-1}) \quad (24)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k \quad (25)$$

2. Measurement update:

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \quad (26)$$

$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + K_k (z_k - h(\hat{X}_{k|k-1})) \quad (27)$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \quad (28)$$

In the present numerical validation, no additional random navigation or camera-calibration noise is injected into the rendered image sequences. The measurement uncertainty used by the EKF is therefore mainly associated with the reconstruction and ICP-registration errors. The process-noise covariance Q_k and measurement-noise covariance R_k are tuned empirically to ensure stable filter convergence under the simulated free-tumbling motion. A more rigorous noise-identification and uncertainty-propagation model will be required for semi-physical or on-orbit validation.

Through the continuous correction of registration-derived attitude measurements, the EKF produces smooth, temporally consistent estimates of the target's attitude, maintaining high accuracy throughout the observation period.

3. Results

3.1. Simulation Setup

To verify the effectiveness of the proposed method in space free-tumbling target perception, a virtual-space simulation platform was constructed based on multi-spacecraft cooperative observations. The target model was a simplified version of the FY-1 satellite, as shown in Figure 6. The FY-1 model is used as a representative non-cooperative spacecraft geometry with a central bus and deployable solar-panel-like structures; the simulation is not intended to reproduce the exact historical orbit or operational state of a specific FY-1 spacecraft.

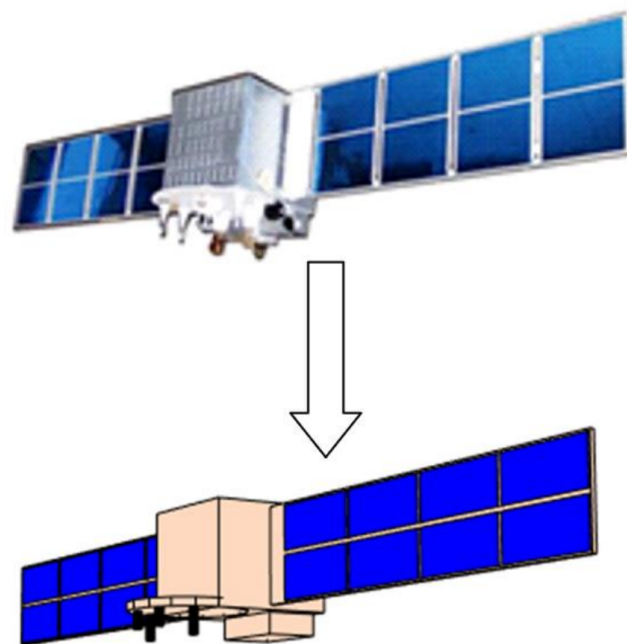


Figure 6. Simplified FY-1 target model used in simulation.

The geometric parameters of the satellite are set as follows: the main body measures $1.4 \text{ m} \times 1.4 \text{ m} \times 1.4 \text{ m}$, and two symmetrical solar panels are mounted on the sides of the bus, giving a total span of approximately 12 m. The orbital and initial attitude parameters of the target are configured as follows:

1. The spacecraft follows a representative circular low-Earth orbit at an altitude of 500 km. This altitude was selected to define a typical LEO relative-motion environment and the corresponding CW mean motion. Since the proposed visual measurement pipeline mainly depends on relative camera poses, viewing geometry, image resolution, and target imaging quality, the absolute orbital altitude affects the method primarily through relative motion and illumination conditions rather than through the reconstruction algorithm itself.
2. The initial attitude quaternion is $\mathbf{q}_{b0} = (1, 0, 0, 0)^T$.
3. The initial angular velocity is $\boldsymbol{\omega}_{b0} = (0.2, 0.2, 0.2)^T \text{ }^\circ/\text{s}$.
4. The inertia matrix is $\mathbf{J}_T = \text{diag}(75, 70, 50) \text{ kg} \cdot \text{m}^2$.

A fixed directional illumination source was used in the rendering environment to maintain a controlled sunlit observation condition. Earth eclipse, Earth albedo, specular reflection, and self-shadowing were not explicitly modeled in the present simulation. This simplification allows the influence of cooperative geometry and reconstruction strategy to be isolated, but it may underestimate the degradation caused by complex illumination and shadow boundaries in practical on-orbit imaging.

The formation of the three-observer spacecraft performs a cooperative fly-around of the free-tumbling target. The fly-around orbit is designed as a shared circular relative orbit based on the periodic solution of the Clohessy–Wiltshire (CW) equations, ensuring the long-term geometric stability of the formation. The relative orbit is described by

$$\begin{cases} x_i = -\frac{r_i}{2} \cos(nt + \theta_i), \\ y_i = r_i \sin(nt + \theta_i), \\ z_i = \pm \frac{\sqrt{3} r_i}{2} \cos(nt + \theta_i). \end{cases} \quad (29)$$

The fly-around radius for the observation formation was set to $r_i = \|\mathbf{r}_i\| = 200 \text{ m}$, and the three spacecraft were initialized at phase angles $(\theta_1, \theta_2, \theta_3) = (0^\circ, 6^\circ, 12^\circ)$. The 200 m radius was selected as a conservative stand-off distance that provides safe separation while keeping the complete target inside the camera FOV. With a 30° FOV, the transverse imaging coverage at 200 m is much larger than the approximately 12 m target span, leaving sufficient margin for tumbling-induced apparent-size variations. Meanwhile, the inter-spacecraft phase separation provides a larger effective baseline than a compact single-platform trinocular system, thereby improving depth observability under cooperative multi-view imaging.

The camera's FOV was set to 30° , with an image resolution of 800×800 pixels and focal lengths $f_x = f_y \approx 1200$. At each imaging epoch, the camera optical axes were directed toward the target centroid to satisfy the FOV constraint. The sampling duration was $t = 3000 \text{ s}$, with an imaging frequency of 0.1 Hz. The long-duration sequence was used to evaluate whether the proposed pipeline remains numerically stable over multiple tumbling cycles. In practical missions, continuous valid observation may be interrupted by illumination, occlusion, orbit geometry, communication, or operational constraints; therefore, the present 3000 s setting should be regarded as a controlled validation window rather than a strict operational requirement. Multi-view image sequences from all spacecraft were collected to form the dataset.

3.2. Validation of the Proposed Method

Based on the acquired image sequences, MVSNet is applied to reconstruct the 3D point clouds of the target. Attitude variations between sampling intervals were then derived from inter-frame point cloud registration. Outliers were removed through preliminary curve fitting. The resulting attitude angle variation estimates are shown in Figure 7, and the corresponding estimation errors are shown in Figure 8.

As shown in Figure 7, although the attitude variation curves exhibited minor fluctuations, all three attitude components clearly reflected the periodic tumbling characteristics of the target. The waveform closely matched the simulated angular velocity, demonstrating that the proposed method can effectively capture the dynamic attitude evolution of a free-tumbling target. The small oscillations mainly arise from illumination differences between adjacent frames and computational noise, however, the overall trend remains highly consistent with the reference trajectory. Figure 8 illustrates the estimation error of the attitude angles. The error distribution exhibited an approximately zero-mean characteristic, with an average error of approximately 0.3° and a maximum error below 0.5° . This indicates that the algorithm maintains a high accuracy even during rapid attitude changes.

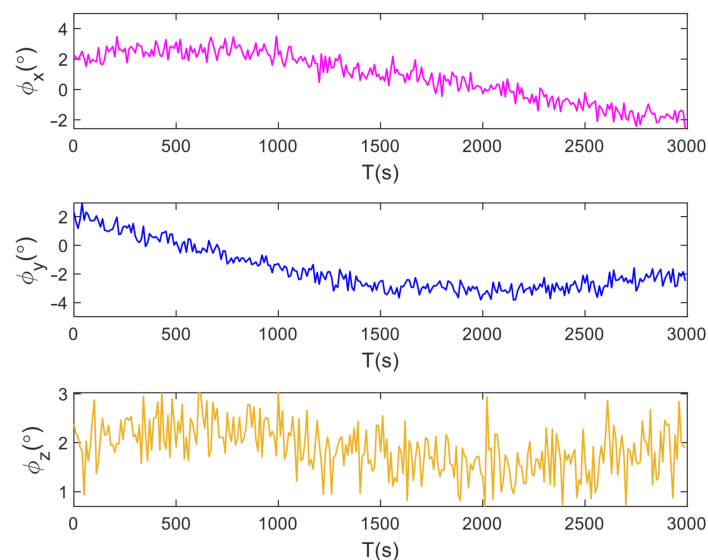


Figure 7. Estimated attitude-angle variations based on MVSNet 3D reconstruction.

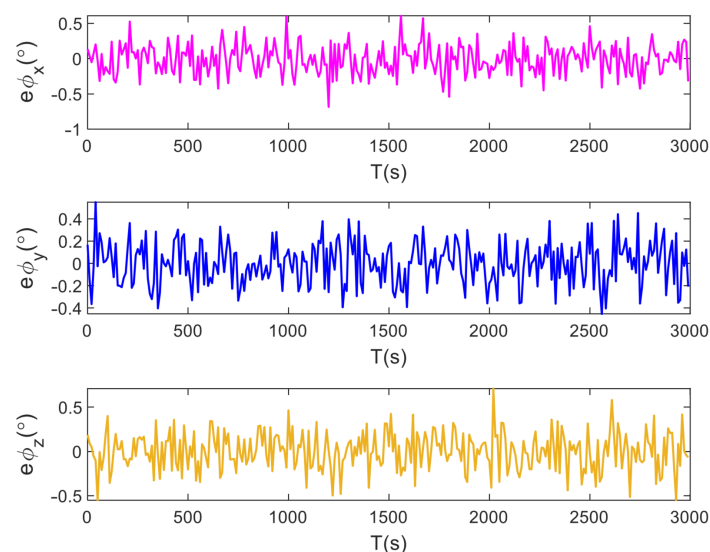


Figure 8. Attitude estimation error based on MVSNet 3D reconstruction.

Next, EKF filtering is applied to further refine the attitude parameter estimates. The EKF results are shown in Figure 9. After a brief initial convergence phase, the estimation errors rapidly stabilize and remain bounded throughout the entire observation window. The attitude error is consistently below 0.3° , and the position error remains under 0.05 m, demonstrating that the filtering performance fully meets the mission requirements.

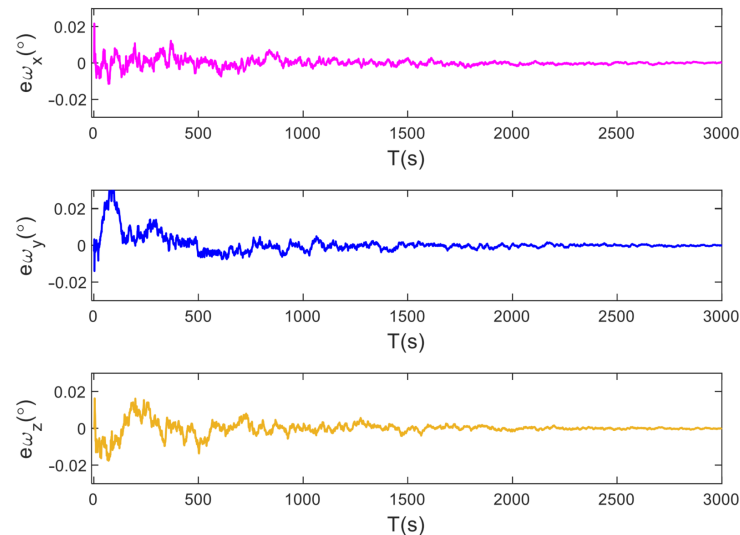


Figure 9. EKF-filtered attitude estimation error based on MVSNet reconstruction.

These results confirm that, using three-view image sequences, the proposed pipeline integrating MVSNet-based 3D reconstruction, point cloud registration, and EKF fusion achieves fast and accurate determination of the state parameters of free-tumbling space targets.

3.3. Comparative Analysis with Single-Spacecraft Trinocular Stereo Vision

To evaluate the performance advantages of the proposed cooperative-observation MVSNet approach under varying observation conditions, this subsection compares it with a single-spacecraft trinocular stereo vision system. The trinocular setup relies on depth estimation derived from fixed-baseline geometric parallax; however, its reconstruction and attitude-determination accuracy degrade significantly when parallax becomes weak.

The target orbital parameters, attitude configuration, and camera imaging model were kept identical to those used in the previous section. The three-spacecraft cooperative observation configuration was also kept unchanged. For the trinocular comparison setup, three cameras were arranged in an equilateral-triangle baseline configuration with a baseline length of 1 m to enable stereo observation and geometric calibration. The 1 m baseline was selected as a representative compact baseline that can be accommodated by a single spacecraft platform or a deployable camera boom while maintaining structural feasibility and calibration stability. This setting provides a practical baseline for evaluating the geometric limitation of single-platform trinocular vision relative to a distributed multi-spacecraft baseline.

The observation distance (i.e., the relative orbit radius of the observer formation) varied from 30 m to 100 m in increments of 5 m. The sampling duration (3000 s) and frequency (0.1 Hz) remained unchanged.

For each observation distance, both the trinocular stereo method and the proposed MVSNet-based cooperative method were applied to reconstruct the target and estimate its attitude. Based on the reconstructed 3D point clouds, attitude-angle variations were calculated using point cloud registration. The attitude-estimation error is defined as follows:

$$e_{\Delta\theta} = \|\Delta\hat{\theta} - \Delta\theta\|_2 \quad (30)$$

where $\Delta\hat{\theta}$ is the estimated attitude angle vector and $\Delta\theta$ is the ground truth value.

The mean attitude estimation error over the entire observation duration was computed for each distance, and the results are shown in Figure 10.

The fitted error curve of the trinocular method shows a clear nonlinear increase with observation distance. In the 30–55 m range, the error remained between 0.3° and 0.5° , indicating that a structured baseline still provides useful parallax for medium-to-close distances. However, when the distance exceeds approximately 60 m, the limited 1 m baseline yields insufficient parallax at a long range, causing the depth-estimation accuracy to deteriorate rapidly. In the 70–90 m interval, the error increased to 0.66 – 0.85° , and reached approximately 1.02° at 100 m.

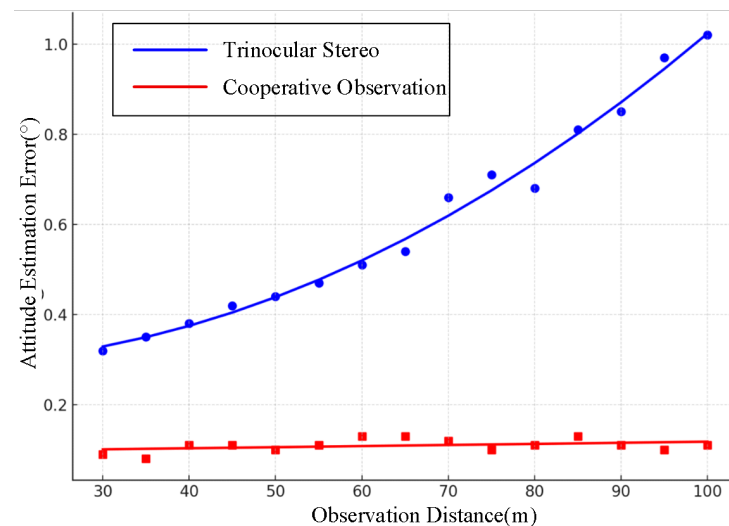


Figure 10. Attitude estimation errors obtained using different methods.

This trend clearly demonstrates that the trinocular stereo approach is fundamentally constrained by baseline length. Under long-range, weak-parallax conditions, its accuracy degrades quickly, severely limiting its applicability to space target observation tasks.

By contrast, the proposed cooperative-observation MVSNet method exhibited significantly higher stability and robustness across the full distance range. Its fitted error curve shows only mild, near-linear fluctuations, with the error consistently maintained within 0.08 – 0.13° , largely independent of distance.

3.4. Comparative Analysis with Feature-Matching-Based Methods

In the previous section, the advantages of the proposed cooperative-observation MVSNet approach over single-spacecraft trinocular vision were validated from the perspective of geometric baseline constraints. In this section, we further compare the proposed method with traditional feature-matching-based 3D reconstruction algorithms, aiming to evaluate the performance gains brought by deep-learning-based reconstruction networks under conditions involving complex pose dynamics.

The simulation environment and target parameters remained consistent with those in the previous sections, and the same three-spacecraft cooperative observation architecture was retained. To evaluate the influence of different imaging geometries on feature matching and depth reconstruction, we varied the observation distance from 30 m to 100 m and the inter-satellite viewing angle from 5° to 20° . For each configuration, the three spacecraft acquired synchronized multi-view image sequences. Subsequently, both the SFM-based feature matching method and the proposed MVSNet method were applied for 3D reconstruction, attitude-variation estimation, and attitude-error computation.

For each test configuration, the attitude estimation error was averaged over a 30 s observation window. The resulting error distributions for the SFM and MVSNet approaches are shown in Figure 11 and Figure 12, respectively. In both figures, the color bar represents the attitude estimation error in degrees.

The results revealed that the SFM-based method performed reasonably well at short distances, achieving attitude errors below 0.2° . However, as the observation distance increases, its estimation accuracy degrades steadily, with errors reaching approximately 0.7° under weaker parallax.

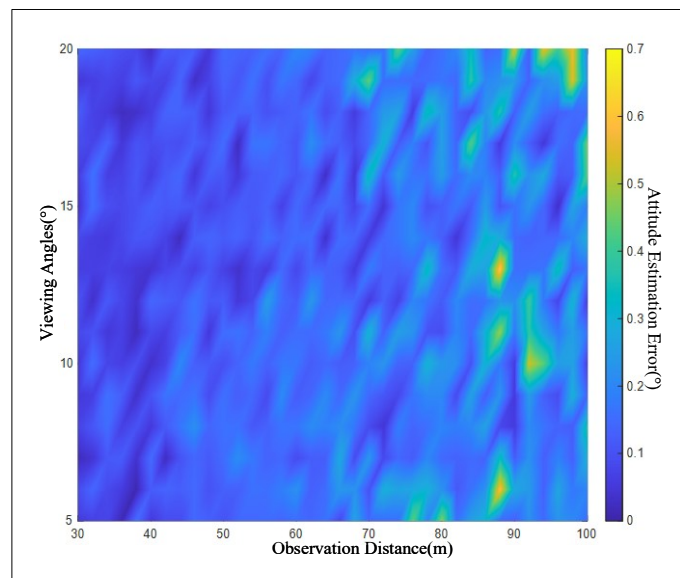


Figure 11. Attitude estimation error using SFM-based 3D reconstruction.

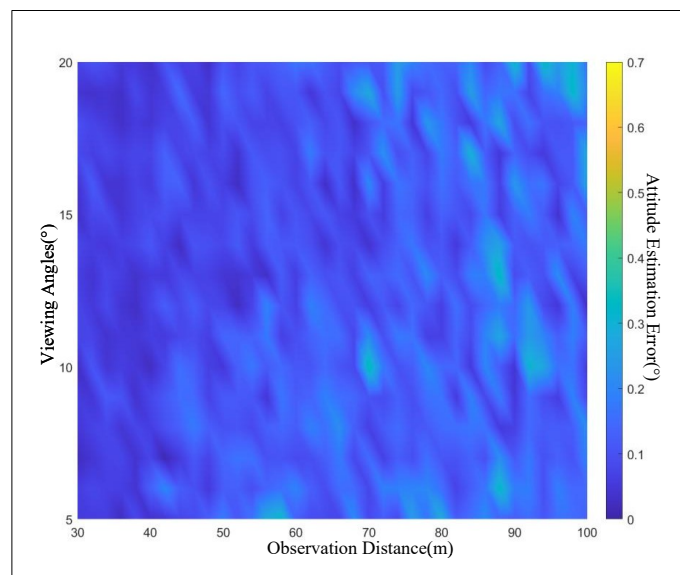


Figure 12. Attitude estimation error using MVSNet-based 3D reconstruction.

In contrast, the MVSNet-based method consistently maintained favorable performance. Although slight increases in the error can be observed as the observation distance increases, the estimation error generally remains below 0.4° , demonstrating superior distance robustness.

3.5. Computational Efficiency

To further evaluate the computational efficiency of the proposed method, Figure 13 compares the average reconstruction time of the SFM and MVSNet pipelines. The compu-

tational experiments were conducted on a Windows 11 workstation equipped with an Intel Core i7-10700 CPU, 64 GB RAM, and an NVIDIA GeForce RTX 3080 Ti GPU. The MVSNet reconstruction module was implemented in Python 3.9 with PyTorch 1.10.1 and standard scientific computing and visualization libraries. All compared methods were evaluated under the same hardware and software environment to ensure a fair comparison of their relative computational costs.

The experimental results show that the proposed MVSNet method required only 0.52 s on average for each reconstruction process over the tested 500-frame sequence, whereas the traditional SFM pipeline required 7.74 s. This indicates that the proposed MVSNet-based method achieves approximately a 15-fold speedup, demonstrating superior computational efficiency over the conventional SFM pipeline under the tested conditions. It should be noted that the absolute runtime may vary with processor, graphics hardware, memory capacity, software implementation, and batch-processing settings. Therefore, the reported values should be interpreted as simulation-based comparative runtime rather than final onboard execution time.

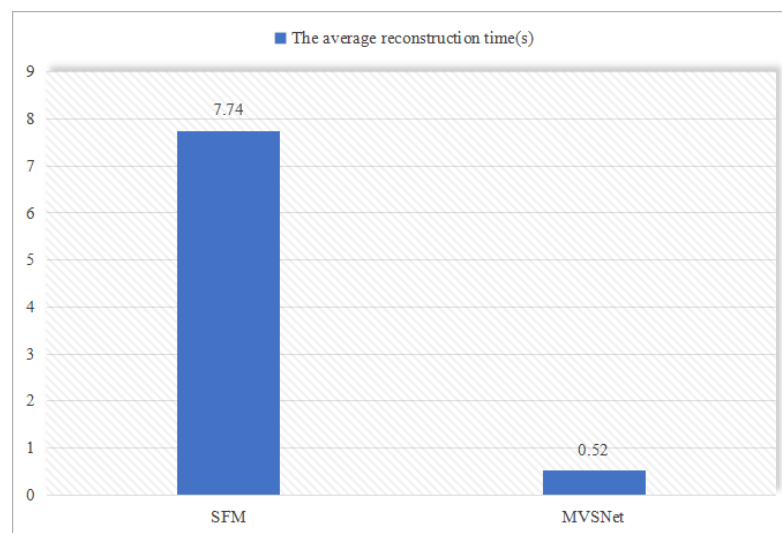


Figure 13. Average 3D reconstruction time using SFM and MVSNet.

4. Discussion

The experimental results demonstrate that the proposed cooperative multi-view reconstruction and attitude estimation framework effectively improves the measurement accuracy and robustness for free-tumbling non-cooperative space targets. Compared with single-spacecraft trinocular stereo vision, the proposed method maintains more stable attitude estimation accuracy across different observation distances. This improvement is mainly attributed to the cooperative observation geometry. In a single-spacecraft trinocular configuration, the physical baseline is limited by the size of the spacecraft platform. As the observation distance increases, the image parallax decreases rapidly, weakening depth constraints and increasing reconstruction errors. In contrast, the multi-spacecraft configuration provides a larger effective observation baseline and richer cross-view geometric constraints. Consequently, the depth uncertainty caused by weak parallax can be more effectively suppressed, and the reconstructed point clouds remain more suitable for subsequent attitude estimation.

The comparison with the SFM-based reconstruction method further confirms the advantages of the learning-based multi-view stereo strategy. Traditional SFM methods rely heavily on stable feature extraction and inter-frame feature matching. For free-tumbling space targets, the available texture is often limited, and the observed appearance may change significantly owing to rapid attitude motion, illumination variation, self-occlusion,

and changes in viewing geometry. Under such conditions, the number of reliable matched feature points may decrease, and incorrect correspondences can directly propagate into the reconstructed sparse point cloud. This explains the degradation in SFM-based attitude estimation accuracy under long-distance and weak-parallax conditions. By contrast, MVSNet uses deep image features, homography warping, cost-volume regularization, and multi-view feature aggregation to infer depth information. Even when local texture is weak or individual point estimates contain small deviations, the reconstructed point cloud provides stronger geometric redundancy for ICP registration. As a result, the MVSNet-based method exhibits better robustness and maintains lower attitude estimation errors under different observation configurations.

The overall attitude estimation accuracy is influenced by several coupled error sources. First, depth estimation uncertainty in multi-view reconstruction directly affects the geometric quality of the reconstructed point cloud. Inaccurate depth estimates may introduce local deformation or noise on the target surface, which subsequently affects point cloud registration. Second, the cooperative observation geometry, including the inter-spacecraft baseline, viewing-angle diversity, and relative distribution of the observers, plays an important role in constraining the depth solution. Insufficient angular diversity may still lead to weak geometric constraints, whereas excessively large viewing differences may increase image appearance inconsistency and matching difficulty. Third, ICP registration may be affected by point cloud sparsity, noise, partial occlusion, and local structural symmetry of the target. These factors can cause small registration deviations between adjacent frames, resulting in fluctuations in the estimated attitude parameters. The EKF module alleviates this problem by incorporating temporal consistency and filtering high-frequency measurement noise, thereby improving the continuity and stability of the final attitude estimates.

The computational efficiency results also indicate the potential of the proposed method for real-time or near-real-time on-orbit applications. In the simulation experiment, the MVSNet-based reconstruction pipeline required only 0.52 s on average for 500 frames, whereas the traditional SFM pipeline required 7.74 s. This difference suggests that the proposed method can significantly reduce reconstruction time while maintaining favorable measurement accuracy. For on-orbit servicing, autonomous rendezvous, and active debris removal missions, rapid perception of target motion is essential for subsequent trajectory planning, relative navigation, and capture control. Therefore, the combination of cooperative observation, deep multi-view reconstruction, ICP registration, and EKF fusion provides a feasible technical route for dynamic attitude measurement of non-cooperative space targets.

The performance of the proposed method is also related to the target tumbling rate and the observer formation configuration. For faster tumbling motion, the inter-frame attitude variation increases at a fixed imaging frequency, which may reduce the overlap between adjacent point clouds and increase the probability of local ICP convergence errors. A higher imaging frequency, shorter registration interval, or stronger temporal filtering would therefore be required for targets with larger angular rates. Similarly, the observer formation should provide sufficient viewing-angle diversity without causing excessive appearance differences among views. A very compact formation leads to weak parallax and poor depth observability, whereas overly large angular separation may increase self-occlusion and cross-view matching difficulty. The selected cooperative configuration represents a compromise between depth observability, FOV coverage, and cross-view consistency.

Nevertheless, several limitations remain. First, the current validation is mainly based on a numerical simulation environment, in which the target model, imaging conditions, and observation geometry can be controlled. Real on-orbit imaging may involve more complex illumination, strong shadows, specular reflection, Earth background interference,

sensor noise, and partial occlusion. In particular, self-shadowing was not explicitly modeled in the present simulation, and this assumption may lead to more continuous and higher-quality depth maps than those obtained under realistic sunlight-shadow boundaries. These factors may reduce the quality of depth estimation and point cloud registration. Second, the present study assumes that the relative poses of the observer spacecraft and camera parameters are accurately known. In practical missions, orbit determination errors, time synchronization errors, camera pointing errors, and camera calibration errors may introduce additional uncertainty into multi-view reconstruction. Third, external disturbance torques are described in the attitude dynamics but are not explicitly estimated in the current EKF implementation. Their unmodeled effects are partly absorbed by the process-noise term, but large disturbances may degrade prediction accuracy and should be incorporated into future robustness tests. Fourth, although the proposed method demonstrates favorable computational efficiency in the simulation environment, the influence of onboard computing resources, memory constraints, and communication delays among multiple spacecraft requires further evaluation.

Future work will focus on improving the realism and applicability of the proposed framework. More complex target models, non-uniform surface materials, realistic illumination conditions, and background interference should be introduced to further evaluate the robustness of the reconstruction and attitude estimation algorithms. In addition, uncertainty propagation from spacecraft relative navigation, camera calibration, and time synchronization should be incorporated into the measurement model. Hardware-in-the-loop experiments or semi-physical simulation platforms will also be considered to verify the feasibility of the method under more realistic sensing and computation conditions. Furthermore, the reconstructed attitude information can be integrated with autonomous guidance and control algorithms to support closed-loop observation planning and robotic capture of free-tumbling non-cooperative space targets.

5. Conclusions

This paper proposed a cooperative multi-view framework for 3D reconstruction and attitude estimation of free-tumbling non-cooperative space targets. The framework integrates multi-spacecraft cooperative observation, MVSNet-based dense reconstruction, ICP-based point cloud registration, and EKF-based attitude fusion to improve the robustness of dynamic attitude measurement.

The simulation results show that the proposed method can generate denser and more stable point clouds than traditional SFM-based reconstruction methods. Compared with single-spacecraft trinocular stereo vision, the proposed cooperative observation scheme maintains better attitude estimation accuracy as the observation distance increases. The computational comparison also shows that the MVSNet-based reconstruction pipeline achieves higher efficiency than the traditional SFM pipeline in the tested scenario.

Future work will focus on more realistic on-orbit imaging conditions, including complex illumination, self-shadowing, sensor noise, background interference, partial occlusion, and target surface reflectance. The effects of relative navigation errors, camera calibration errors, camera pointing errors, time synchronization errors, external disturbances, faster tumbling motion, and different formation configurations will also be further investigated through semi-physical simulation and hardware-in-the-loop experiments.

Author Contributions: Conceptualization, D.Z., Z.Y. and W.Z.; methodology, D.Z.; software, D.Z.; validation, D.Z., H.B. and S.L.; formal analysis, D.Z. and W.L.; investigation, D.Z.; resources, Z.Y., W.Z., J.Y. and W.M.; data curation, D.Z.; writing—original draft preparation, D.Z.; writing—review and editing, Z.Y., W.Z., J.Y., W.M., H.B., S.L. and W.L.; visualization, D.Z.; supervision, Z.Y. and W.Z.;

project administration, Z.Y., W.Z. and J.Y.; funding acquisition, Z.Y. and W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, Grant Nos. 42574047 and 42508023; the Henan Province Science and Technology R&D Program, Grant No. 262102210234; the Henan Provincial Natural Science Foundation, Grant No. 262300422738; and the High-level Talent Research Start-up Project Funding of Henan Academy of Sciences, Project No. 20262325005.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study were generated from numerical simulation experiments. The simulation parameters, experimental settings, and main results are described in the manuscript.

Acknowledgments: The authors would like to thank the anonymous reviewers and editors for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional neural network
CW	Clohesy–Wiltshire
EKF	Extended Kalman filter
FOV	Field of view
FY-1	Fengyun-1
ICP	Iterative closest point
LSTM	Long short-term memory
MAE	Mean absolute error
MVS	Multi-view stereo
MVSNet	Multi-view stereo network
NCC	Normalized cross-correlation
PMVS	Patch-based multi-view stereo
PnP	Perspective-n-Point
RANSAC	Random sample consensus
RGB	Red–green–blue
SFM	Structure from motion
SFS	Shape from shading
SGM	Semi-global matching

References

1. Opromolla, R.; Grishko, D.; Auburn, J.; Bevilacqua, R.; Buinhas, L.; Cassady, J.; Jäger, M.; Jankovic, M.; Rodriguez, J.; Perino, M.A.; et al. Future in-orbit servicing operations in the space traffic management context. *Acta Astronaut.* **2024**, *220*, 469–477.
2. Flores-Abad, A.; Ma, O.; Pham, K.; Ulrich, S. A review of space robotics technologies for on-orbit servicing. *Prog. Aerosp. Sci.* **2014**, *68*, 1–26.
3. Amaya-Mejía, L.M.; Ghita, M.; Dentler, J.; Olivares-Mendez, M.A.; Martinez, C. Visual servoing for robotic on-orbit servicing: A survey. In Proceedings of the 2024 International Conference on Space Robotics (iSpaRo), Luxembourg, 24–27 June 2024; pp. 178–185.
4. Li, S.; Li, J.; Chen, Q.; Li, X.; Wang, J. Review of pose estimation methods for spacecraft targets. *Aerosp. China* **2025**, *26*, 53–58.
5. Pauly, L.; Rharaoui, W.; Shneider, C.; Rathinam, A.; Gaudilliere, V.; Aouada, D. A survey on deep learning-based monocular spacecraft pose estimation: Current state, limitations and prospects. *Acta Astronaut.* **2023**, *212*, 339–360.

6. Wang, Z.; Liu, Y.; Zhang, E. Pose estimation for cross-domain non-cooperative spacecraft based on spatial-aware keypoints regression. *Aerospace* **2024**, *11*, 948.
7. Zhang, H.; Wei, Q.; Jiang, Z. 3D reconstruction of space objects from multi-views by a visible sensor. *Sensors* **2017**, *17*, 1689.
8. Zhang, K.; Snavely, N.; Sun, J. Leveraging vision reconstruction pipelines for satellite imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 2139–2148.
9. Chang, L.; Liu, J.; Chen, Z.; Bai, J.; Shu, L. Stereo vision-based relative position and attitude estimation of non-cooperative spacecraft. *Aerospace* **2021**, *8*, 230.
10. Uno, K.; Matsuoka, T.; Uchida, A.; Yoshida, K. Structure-from-motion-based motion estimation and 3D reconstruction of unknown-shaped space debris. In Proceedings of the IEEE International Conference on Automation Science and Engineering, Bari, Italy, 28 August–1 September 2024; pp. 1283–1288.
11. Park, T.H.; Sharma, S.; D’Amico, S. Towards robust learning-based pose estimation of non-cooperative spacecraft. *arXiv* **2019**, arXiv:1909.00392.
12. Braun, C.; Caverlee, J.; Pagani, A.; Johnston, J.C. Satellite pose estimation with deep landmark regression and nonlinear pose refinement. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019; pp. 2816–2824.
13. Qiao, S.; Zhang, H.; Meng, G.; An, M.; Xie, F.; Jiang, Z. Deep-learning-based satellite relative pose estimation using monocular optical images and 3D structural information. *Aerospace* **2022**, *9*, 768.
14. Rondao, D.; Aouf, N.; Richardson, M.A. ChiNet: Deep recurrent convolutional learning for multimodal spacecraft pose estimation. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 937–949.
15. Ye, R.; Wang, L.; Ren, Y.; Wang, Y.; Chen, X.; Liu, Y. FilterformerPose: Satellite pose estimation using Filterformer. *Sensors* **2023**, *23*, 8633.
16. Zhao, D.; Sun, C.; Zhu, Z.; Wan, W.; Zheng, Z.; Yuan, J. Multi-spacecraft collaborative attitude determination of space tumbling target with experimental verification. *Acta Astronaut.* **2021**, *185*, 1–13.
17. Matsuka, K.; Santamaria-Navarro, A.; Capuano, V.; Harvard, A.; Rahmani, A.; Chung, S. Collaborative pose estimation of an unknown target using multiple spacecraft. In Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA, 6–13 March 2021; pp. 1–11.
18. Yao, Y.; Luo, Z.; Li, S.; Fang, T.; Quan, L. MVSNet: Depth inference for unstructured multi-view stereo. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 767–783.
19. Gu, X.; Fan, Z.; Zhu, S.; Dai, Z.; Tan, F.; Tan, P. Cascade cost volume for high-resolution multi-view stereo and stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2495–2504.
20. Sormann, C.; Knöbelreiter, P.; Kuhn, A.; Rossi, M.; Pock, T.; Fraundorfer, F. BP-MVSNet: Belief-propagation layers for multi-view stereo. In Proceedings of the International Conference on 3D Vision, Fukuoka, Japan, 25–28 November 2020; pp. 394–403.
21. Wang, F.; Galliani, S.; Vogel, C.; Speciale, P.; Pollefeys, M. PatchMatchNet: Learned multi-view patchmatch stereo. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14194–14203.
22. Yu, A.; Guo, W.; Liu, B.; Chen, X.; Wang, X.; Cao, X.; Jiang, B. Attention-aware cost volume pyramid based multi-view stereo network for 3D reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 448–460.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.