The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIX-B2-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission II, 4–11 July 2026, Toronto, Canada

Large-Field binocular vision attitude determination method for rocket recovery

Yuqi Zhang [1], Xianglei Liu [1], Runjie Wang [1], Haibo Shi [1], Zhao Lu [1], Haiqian Wu [1]

[1] Key Laboratory for Urban Geomatics of National Administration of Surveying Mapping and Geoinformation, Engineering Research Centre of Representative Building and Architectural Database, Ministry of Education Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Keywords: Binocular vision, Attitude measurement, Edge detection, Central axis fitting, Rocket recovery.

Abstract

Accurate attitude estimation is essential for stable guidance and control during rocket recovery, yet it remains challenging because the target undergoes rapid pose changes, occupies only a limited image area over a large observation corridor, and often exhibits weak texture and approximate axial symmetry. To address these issues, this paper proposes a large-field binocular-vision-based attitude determination method for rocket recovery. First, a distortion-aware stereo calibration strategy based on stitched control points is developed to enable reliable geometric modeling over a large measurement field with a portable calibration target. Second, a robust contour extraction pipeline is constructed by combining bilateral filtering, gradient enhancement, and multi-threshold Canny fusion. Third, the rocket central axis is reconstructed by fitting 2D midlines in rectified stereo images and intersecting their corresponding back-projection planes, which improves stability over point-wise triangulation. Finally, pitch and yaw are derived from the recovered 3D axis direction, while roll is estimated by phase correlation on the polar-unwrapped base image under a temporal continuity constraint. Experiments on a 1:20 cylindrical scale model show RMS reprojection errors of 0.056 px and 0.066 px for the left and right cameras, respectively, and a 3D checkpoint RMSE of 33.42 mm. On a 100-frame sequence, the proposed method achieves RMSEs of 1.58°, 1.54° and 1.41° for roll, pitch, and yaw, respectively, outperforming ORB+PnP, SGBM, and Chamfer-based baselines. The results demonstrate that the proposed method provides an accurate and practical optical solution for external attitude measurement in rocket-recovery scenarios.

1. Introduction

With the rapid development of reusable launch vehicles, precise pose and attitude measurement during descent and landing has become increasingly important. Recent off-board high-speed videogrammetry studies on reusable rockets have shown that external optical systems can provide high-frequency six-degree-of-freedom measurements under uncontrolled outdoor illumination, underscoring the practical value of vision-based sensing for recovery evaluation and system optimization (Tong et al., 2025).

Conventional attitude determination still relies heavily on inertial sensing. Inertial measurement units and gyroscopes provide high-rate, fully autonomous measurements, but their outputs are affected by stochastic noise, bias instability, and drift accumulation that must be modeled and compensated carefully (El-Sheimy et al., 2008). Visual-inertial methods alleviate pure inertial drift by coupling image observations with inertial preintegration, yet their performance remains dependent on stable visual tracking, reliable initialization, and sufficient scene texture (Qin et al., 2018). In rocket recovery scenarios, however, rapid attitude variation, long observation distance, motion blur, and illumination changes make these assumptions difficult to maintain consistently (Tong et al., 2025).

Compared with monocular external vision, binocular vision can directly recover three-dimensional geometry through stereo constraints and is therefore attractive for pose measurement over a large working space. Existing vision-based pose-estimation pipelines generally follow either feature-correspondence/PnP formulations or dense stereo correspondence formulations. Feature-based approaches are effective when repeatable and well-distributed correspondences can be established between the image and the object model (Lepetit et al., 2009), whereas dense stereo methods have matured into strong geometric reconstruction tools for calibrated scenes (Scharstein et al., 2002; Hirschmüller, 2008). However, both routes become less reliable when the target is weakly textured, locally repetitive, or occupies only a limited fraction of the image.

These limitations are especially pronounced in rocket recovery. First, the measurement corridor is large, but the rocket often occupies only a small region in the image, so a single portable target is usually insufficient to constrain the full working volume during calibration. Large-field calibration studies have therefore explored onsite small-target and combined-small-target strategies to extend calibration coverage without requiring a single oversized calibration object (Wang et al., 2013; Liu et al., 2015). Second, the rocket body is typically elongated, weakly textured, and approximately axis-symmetric. For such targets, contour- or symmetry-driven methods are often more appropriate than pure sparse-feature matching, but the symmetry itself increases the difficulty of pose recovery, especially around the body axis (Fioravanti et al., 2009; Figueiredo et al., 2015).

The observability of the three attitude components is therefore inherently unbalanced. Pitch and yaw can often be inferred from the reconstructed body axis, whereas roll cannot be robustly recovered from the side contour alone. Studies on symmetrical and textureless objects have shown that symmetry may leave orientation partially ambiguous, while contour-based matching remains highly sensitive to the stability of global shape extraction and outlier suppression (Penman and Alwesh, 2006; Zhang et al., 2018). Moreover, any practical recovery method must remain geometrically interpretable and resistant to noise, blur, and radiometric degradation. For rotational displacement estimation, phase-correlation methods in the Fourier/polar domain provide an attractive basis because they estimate rotation efficiently and retain good robustness to random noise (Reddy and Chatterji, 1996).

Taken together, the literature suggests that a practical binocular attitude-determination method for rocket recovery should simultaneously address four requirements: accurate distortion-aware geometric modeling over a large spatial corridor, robust

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-1155-2026 |

contour extraction for weakly textured elongated targets, stable three-dimensional central-axis reconstruction for pitch/yaw estimation, and ambiguity-aware roll recovery beyond side-silhouette information. Recent rocket photogrammetry systems demonstrate that high-frequency external measurement is feasible, but they also indicate that existing solutions commonly rely on designed surface targets and multi-camera control networks; this suggests room for a more geometry-driven method tailored to elongated recovery vehicles (Tong et al., 2025).

To address these issues, this paper proposes a large-field binocular-vision-based attitude determination method for rocket recovery. In this paper, “large-field” refers to the large spatial measurement corridor and stand-off observation configuration rather than to an ultra-wide-angle imaging model. First, a distortion-aware stereo calibration strategy based on stitched control points is developed to achieve accurate geometric modeling over a large measurement corridor using a small portable target. Second, a robust contour extraction pipeline is constructed by integrating bilateral filtering, multi-threshold Canny fusion, line-support filtering, and central-axis reconstruction based on back-projection-plane intersection. Third, pitch and yaw are derived from the recovered 3D axis direction, while roll is estimated from the base-disc region in the polar domain via phase correlation under a temporal continuity constraint. In this way, the proposed framework couples large-field calibration, structure-aware reconstruction, and ambiguity-resolved roll estimation within a unified and geometrically interpretable pipeline.

The main contributions of this work are as follows:

1. A distortion-aware large-field stereo calibration strategy based on stitched control points is proposed for accurate geometric modeling over a large measurement corridor with a small portable target.
2. A robust contour extraction and central-axis reconstruction pipeline is designed by integrating bilateral filtering, multi-threshold Canny fusion, line-support filtering, and back-projection-plane intersection, making the method suitable for elongated weakly textured targets.
3. A phase-correlation-based roll recovery method is introduced in the polar coordinate domain, and the rotational ambiguity is resolved through a temporal continuity constraint.
4. Comparative experiments and ablation studies demonstrate that the proposed method achieves higher accuracy and stability than representative baseline methods.

The remainder of this paper is organized as follows. Section 2 presents the proposed methodology, including calibration, image preprocessing, axis reconstruction, and attitude estimation. Section 3 reports the experimental setup and quantitative results. Section 4 discusses applicability and limitations. Section 5 concludes the paper.

2. Methodology

As shown in Figure 1, the overall workflow contains four stages: (1) large-field stereo calibration, (2) image preprocessing and contour extraction, (3) central-axis reconstruction, and (4) roll-pitch-yaw attitude estimation.

For clarity, bold uppercase letters denote matrices, bold lowercase letters denote vectors, and italic letters denote scalars.

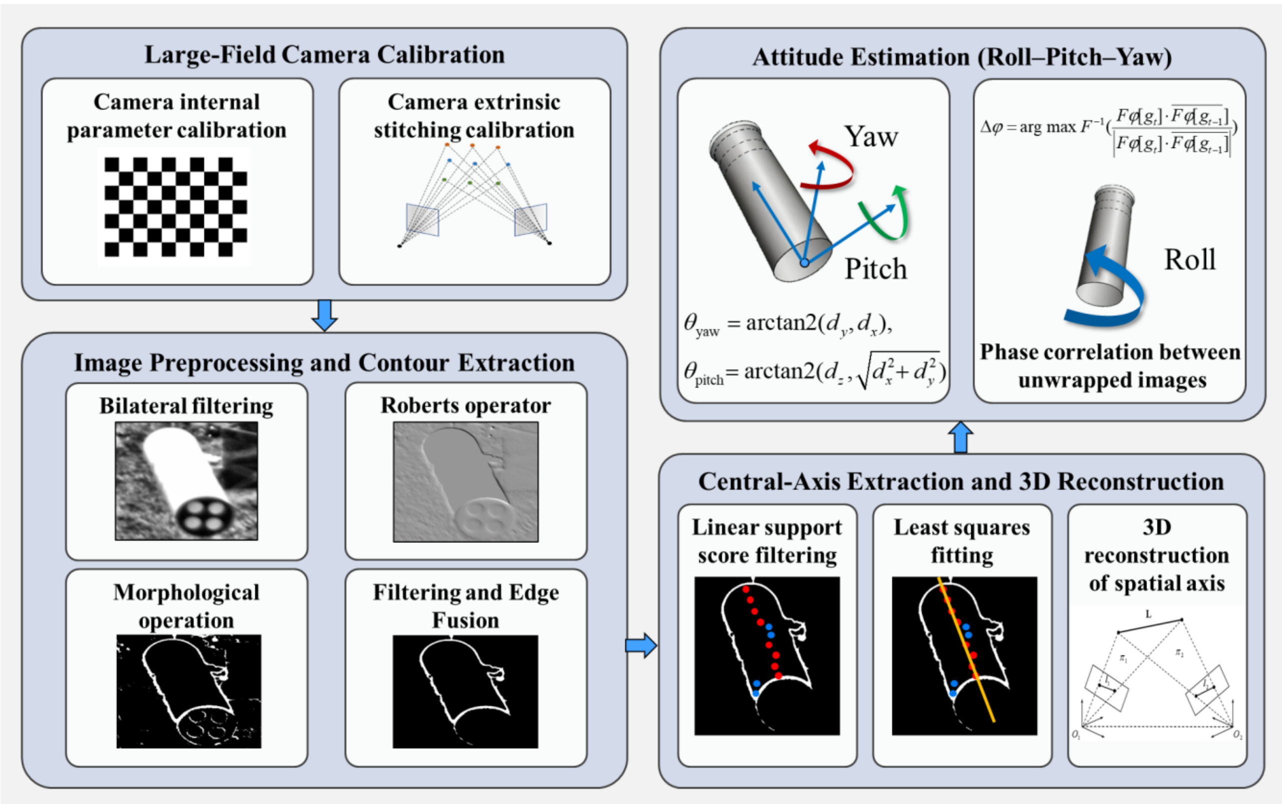

Figure 1. The flowchart of attitude determination technology for large-field binocular vision in rocket recovery.

2.1 Large-field stereo calibration with stitched control points

The image formation process is modeled by a pinhole projection with radial and tangential lens distortion. Let a 3D point in the world coordinate system be $X_w = [X_w, Y_w, Z_w]^\top$. Its coordinates in the camera system are

$$\mathbf{X}_c = \mathbf{R}\mathbf{X}_w + \mathbf{t}, \tag{1}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the extrinsic rotation matrix and translation vector, respectively. The normalized image coordinates are

$$x_n = \frac{X_c}{Z_c}, \qquad y_n = \frac{Y_c}{Z_c}. \tag{2}$$

The radial distance is

$$r^2 = x_n^2 + y_n^2. \tag{3}$$

A Brown-type distortion model is used:

$$x_d = x_n\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) + 2p_1 x_n y_n + p_2\left(r^2 + 2x_n^2\right), \tag{4}$$

$$y_d = y_n\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) + p_1\left(r^2 + 2y_n^2\right) + 2p_2 x_n y_n, \tag{5}$$

where k_1, k_2, k_3 are radial distortion coefficients and p_1, p_2 are tangential distortion coefficients. The final pixel coordinates are obtained by

$$u = f_x x_d + c_x, \qquad v = f_y y_d + c_y, \tag{6}$$

where f_x and f_y are focal lengths in pixel units, and c_x, c_y are principal point coordinates.
The intrinsic and extrinsic parameters are estimated by minimizing the total reprojection error:

$$E = \sum_{j=1}^{N_I}\sum_{i=1}^{N_j}\left\|\mathbf{x}_{ij} - \pi(\mathbf{K}, \mathbf{d}, \mathbf{R}_j, \mathbf{t}_j, \mathbf{X}_{w,i})\right\|_2^2, \tag{7}$$

where $\mathbf{x}_{ij}$ is the observed image position of the i-th control point in the j-th image, $\mathbf{K}$ is the intrinsic matrix, $\mathbf{d}$ is the vector of distortion coefficients, N_I is the number of calibration images, and N_j is the number of visible control points in image j.
Because a single target cannot cover the full large measurement field, a portable circular-target board is moved to multiple positions. All target placements are unified into one world coordinate system by total-station measurements, thereby forming a stitched control-point network. In this paper, the index i in Equation (7) denotes the global stitched control-point index, rather than the local point number on one target board.
After stereo calibration, both images are undistorted and rectified. All subsequent contour extraction and stereo processing are carried out on the undistorted and rectified images.

2.2 Image preprocessing and contour extraction

Reliable contour extraction is the basis for stable axis reconstruction. Let $I(\mathbf{p})$ denote the grayscale intensity at pixel $\mathbf{p}$. A bilateral filter is first applied to suppress noise while preserving edges:

$$I_b(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in\Omega(\mathbf{p})} w_s(\mathbf{p},\mathbf{q}) w_r(\mathbf{p},\mathbf{q}) I(\mathbf{q})}{\sum_{\mathbf{q}\in\Omega(\mathbf{p})} w_s(\mathbf{p},\mathbf{q}) w_r(\mathbf{p},\mathbf{q})}, \tag{8}$$

where $\Omega(\mathbf{p})$ is the local neighborhood, w_s is the spatial Gaussian kernel, and w_r is the range kernel.
To enhance local contour responses, the Roberts operator is used:

$$G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} * I_b, \qquad G_y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} * I_b, \tag{9}$$

$$G = \sqrt{G_x^2 + G_y^2}, \tag{10}$$

where $*$ denotes convolution.

A multi-threshold Canny fusion strategy is then applied to the gradient-enhanced image. After grayscale normalization to $[0,255]$, three threshold pairs are used in the experiments:

$$(T_l^{(1)}, T_h^{(1)}) = (30,90), (T_l^{(2)}, T_h^{(2)}) = (50,150), (T_l^{(3)}, T_h^{(3)}) = (70,210). \tag{11}$$

These threshold pairs generate three binary edge maps, $E^{(1)}$, $E^{(2)}$, and $E^{(3)}$. The fused edge map is obtained by majority voting:

$$E_f(\mathbf{p}) = \begin{cases} 1, & \text{if } \sum_{m=1}^{3} E^{(m)}(\mathbf{p}) \ge 2, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

This operation retains edges that are stable across multiple threshold levels while suppressing isolated spurious responses. Finally, a small morphological closing operation is applied to bridge local gaps and improve contour continuity.

2.3 Central-axis extraction and 3D reconstruction

After rectification, the rocket body contour in each image is scanned row by row. For row v , let the left and right boundary points be

$$\ell_v = (u_v^\ell, v), \qquad r_v = (u_v^r, v). \tag{13}$$

Their midpoint is defined as

$$m_v = \left(\frac{u_v^\ell + u_v^r}{2}, v\right). \tag{14}$$

For an ideal circular cylinder under orthographic projection, the midpoint of the two silhouette boundaries lies exactly on the projected central axis. Under perspective projection, the deviation is a second-order term related to the ratio between the body radius and the camera-to-target distance. In the present setup, the model diameter is approximately 0.3 m and the observation distance is on the order of several meters to tens of meters, so this midpoint approximation is sufficiently accurate for axis estimation.
To suppress outliers caused by local contour defects and noise, a line-support score is introduced. Let the candidate line be $L: au + bv + c = 0$. The distance from midpoint m_v to line L is

$$d(m_v, L) = \frac{\left| a\left(\frac{u_v^\ell + u_v^r}{2}\right) + bv + c \right|}{\sqrt{a^2 + b^2}}. \tag{15}$$

The support score is then

$$S(L) = \sum_{v\in V} \mathbf{1}\left(d(m_v, L) < \epsilon\right), \tag{16}$$

where V is the set of valid image rows, ϵ is the inlier threshold, and $\mathbf{1}(\cdot)$ is the indicator function. The line with the maximum support is selected, and total least squares is applied to the inlier midpoints to estimate the final 2D midline in each rectified image.
Let the fitted lines in the left and right images be $\mathbf{l}^L$ and $\mathbf{l}^R$, respectively, in homogeneous form. Each image line defines a 3D back-projection plane passing through the corresponding camera center. The normal of the plane is

$$\mathbf{n}^c \propto (\mathbf{K}_c^{-1})^T \mathbf{l}^c, \qquad c \in \{L, R\}. \tag{17}$$

The spatial central-axis direction is obtained by intersecting the two planes:

$$\mathbf{v} = \frac{\mathbf{n}^L \times \left(\mathbf{R}_{RL}^T \mathbf{n}^R\right)}{\left\|\mathbf{n}^L \times \left(\mathbf{R}_{RL}^T \mathbf{n}^R\right)\right\|}, \tag{18}$$

where $\mathbf{R}_{RL}$ is the rotation from the right camera frame to the left camera frame. Compared with point-by-point triangulation of all midpoints, this plane-intersection strategy is less sensitive to local disparity noise and therefore yields a more stable axis direction.

2.4 Attitude estimation

2.4.1 Pitch and yaw estimation

Let the normalized 3D axis direction in the left camera frame be

$$\mathbf{v} = [v_x, v_y, v_z]^T. \tag{19}$$

Pitch and yaw are computed from $\mathbf{v}$ as

$$\psi = \mathrm{atan2}(v_x, v_z), \tag{20}$$

$$\theta = \mathrm{atan2}\left(-v_y, \sqrt{v_x^2 + v_z^2}\right), \tag{21}$$

where ψ and θ denote yaw and pitch, respectively.

2.4.2 Roll estimation by polar unwrapping and phase correlation

Because the cylindrical body is approximately axis-symmetric, roll cannot be robustly recovered from the side contour alone. Therefore, the roll angle is estimated from the base-disc appearance.

Let (u_0, v_0) be the center of the base-disc region in the image. The local base image is transformed into polar coordinates:

$$U(\rho, \vartheta) = I\left(u_0 + \rho\cos\vartheta, v_0 + \rho\sin\vartheta\right), \tag{22}$$

where ρ is the radial coordinate and ϑ is the angular coordinate. In the polar image, in-plane rotation becomes a translation along the angular direction.

Given two consecutive polar-unwrapped images U_t and U_{t-1}, the phase-correlation response is

$$C_t = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(U_t)\mathcal{F}(U_{t-1})^*}{\left|\mathcal{F}(U_t)\mathcal{F}(U_{t-1})^*\right|}\right) \tag{23}$$

where $\mathcal{F}(\cdot)$ is the Fourier transform and $(\cdot)^*$ denotes complex conjugation.

The roll increment is determined by

$$\Delta\phi_t = \arg\max_{|\delta| \le \Delta\phi_{\max}} C_t(\delta), \tag{24}$$

$$\phi_t = \phi_{t-1} + \Delta\phi_t, \tag{25}$$

where ϕ_t is the roll angle at frame t, and $\Delta\phi_{\max}$ is the maximum physically admissible inter-frame roll change. In this study, $\Delta\phi_{\max}$ is set to 30°. At 20 fps, this corresponds to a maximum roll rate of 600°/s, which remains below the 90° ambiguity period of an approximately fourfold repetitive base pattern. Therefore, the temporal continuity constraint effectively resolves rotational ambiguity.

Finally, the attitude matrix is represented using the Z-Y-X Tait-Bryan convention:

$$\mathbf{R}_{bw} = \mathbf{R}_z(\psi)\mathbf{R}_y(\theta)\mathbf{R}_x(\phi). \tag{26}$$

3. Experimental results and analysis

3.1 Experimental Setup

The proposed method was evaluated on a 1:20 cylindrical scale model with an approximate length of 1.0 m and a diameter of 0.3 m. The model was rigidly mounted on a carrier UAV platform to simulate the motion of a recovering rocket. The overall experimental configuration, including the stereo-camera layout, the carrier UAV, and the flight direction, is illustrated in Figure 2. Two stationary industrial cameras formed an external stereo pair and observed the test corridor. The cameras were mounted at a height of approximately 6 m above the ground with a baseline of approximately 2 m. Hardware synchronization was achieved using TTL trigger pulses.

The hardware and acquisition settings are summarized in Table 1.

Item	Value
Camera model	Cyclone-16-300
Number of cameras	2
Image resolution	4672×3416
Pixel size	3.9μm×3.9μm
Lens focal length	35 mm
Nominal sensor size	18.22 mm×13.32 mm
Frame rate	20 fps
Stereo baseline	2 m
Evaluated sequence length	100 frames

Table 1. Hardware and acquisition parameters.

It should be noted that the term large-field in this paper refers to the large measurement field and observation corridor, rather than to an ultra-wide-angle or fisheye lens. Even so, lens distortion is explicitly modeled and removed during calibration.

A portable calibration target containing five circular control points was moved to eight positions within the measurement field. The calibration-field arrangement and the spatial distribution of the stitched control-point positions are shown in Figure 3. In total, 40 target observations were obtained, among which 35 stitched control points were used for calibration and 5 independent points were used as checkpoints. The 3D coordinates of these points were measured in a common world coordinate system using a total station.

For quantitative evaluation, a reference attitude sequence was obtained from the synchronized onboard flight-control IMU of the carrier UAV. Since the scale model was rigidly mounted to the carrier platform, the model attitude differed from the carrier attitude only by a constant mounting offset, which was calibrated and compensated before evaluation. The IMU output served only as a reference for error analysis and was not fused into the proposed vision-based estimator.

Three baseline methods were used for comparison:

1. ORB + PnP, representing a sparse feature-based pose estimation strategy;
2. SGBM, representing a dense stereo reconstruction strategy;
3. Chamfer matching, representing a contour-based template registration strategy.

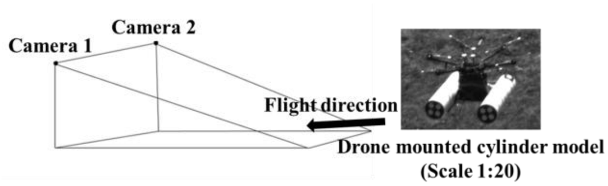

Figure 2. Schematic diagram of scale model experiment setup.

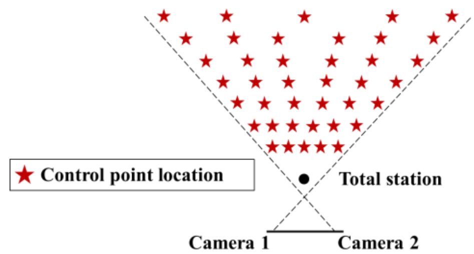

Figure 3. Calibration field setup scene.

3.2 Calibration accuracy

The intrinsic calibration yielded RMS reprojection errors of 0.056 px for the left camera and 0.066 px for the right camera. These low residuals are consistent with the distortion-aware model and the subsequent undistortion-rectification procedure.
Using the stitched control-point network, the reconstructed checkpoint coordinates were compared with the total-station measurements. The resulting 3D RMSE was 33.42 mm, with per-axis errors of 9.74 mm in X, 31.12 mm in Y, and 7.30 mm in Z. The relatively larger error in the Y direction is attributed to the anisotropic geometry of the large-field stereo setup and the spatial distribution of the calibration points.
Overall, the calibration accuracy is sufficient for the subsequent attitude estimation task and supports stable geometric reconstruction across the measurement corridor.

3.3 Attitude estimation accuracy and comparison with baseline methods

Let $\hat{\alpha}_t$ be the estimated angle at frame t, and let α_t^{ref} be the corresponding reference angle from the synchronized IMU sequence, where $\alpha \in \{\phi, \theta, \psi\}$ denotes roll, pitch, and yaw. The RMSE is defined as

$$\mathrm{RMSE}(\alpha) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{\alpha}_t - \alpha_t^{ref})^2}, \tag{27}$$

where N is the number of frames.
In addition to RMSE, the percentage of frames whose absolute errors in all three angles are simultaneously within $\pm 2°$ is reported as

$$A_{2^\circ} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{1}\left(\left|e_\phi(t)\right| \le 2^\circ \wedge \left|e_\theta(t)\right| \le 2\circ \wedge \left|e_\psi(t)\right| \le 2^\circ\right) \times 100\%. \tag{28}$$

The quantitative comparison is given in Table 2, where the last column reports Acc_{2°.

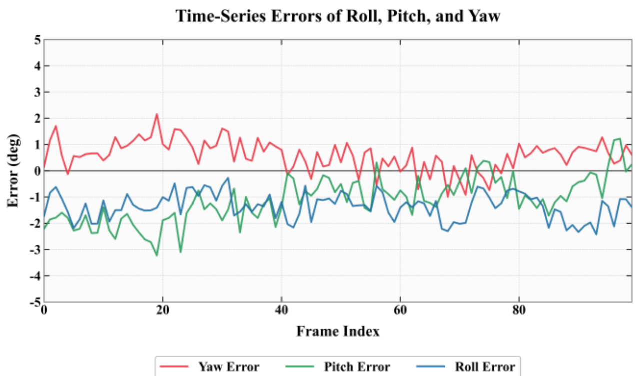

Figure 4. Time-series errors of roll, pitch, and yaw over the evaluated 100-frame sequence.

Methods	Roll RMSE	Pitch RMSE	Yaw RMSE	Acc_{2° $\leqslant \pm 2°$
Proposed	1.58°	1.54°	1.41°	93%
ORB + PnP	2.11°	2.06°	2.18°	82%
SGBM	2.50°	2.45°	2.67°	76%
Chamfer	3.00°	2.80°	3.10°	58%

Table 2. Comparison of different attitude determination methods.

The proposed method achieves the best performance in all three attitude components. Compared with ORB + PnP, it reduces the RMSE by 0.53°, 0.52°, and 0.77° roll, pitch, and yaw, respectively. The advantage over SGBM is larger, indicating that the proposed line- and plane-based geometric strategy is more robust than dense stereo matching in this weak-texture scenario. The performance gain over Chamfer matching further shows that direct silhouette-template registration is less stable under the present imaging conditions.
The time-series errors of roll, pitch, and yaw over the evaluated 100-frame sequence are shown in Figure 4. The three error curves remain centered near zero and exhibit no obvious drift or abrupt jumps, which confirms the temporal stability of the proposed method.

3.4 Ablation Study

To evaluate the contribution of each component, an ablation study was conducted. Four variants were tested:
1. removing line-support scoring,
2. replacing multi-threshold Canny fusion with a single-threshold Canny detector,
3. replacing plane-intersection reconstruction with point-by-point triangulation,
4. replacing phase correlation with standard cross-correlation for roll recovery.
The results are shown in Table 3.

Variant	Roll RMSE	Pitch RMSE	Yaw RMSE	Acc_{2° $\leqslant \pm 2°$
Baseline (Full Method)	1.58°	1.54°	1.41°	93%
No Line-Support Scoring	1.85°	2.05°	1.75°	85%
Single Threshold Canny Fusion	1.72°	1.80°	1.65°	88%
Point-by-Point Triangulation	1.75°	1.82°	1.68°	87%
Cross-Correlation Recovery	1.61°	1.58°	1.43°	91%

Table 3. Ablation study results.

Removing line-support scoring causes the most notable degradation in pitch estimation, showing that outlier suppression is essential for robust midline fitting. Replacing multi-threshold Canny fusion with a single-threshold detector leads to worse accuracy in all three angles, confirming that the proposed contour extraction strategy improves stability under varying edge strength. Point-by-point triangulation also performs worse than the proposed plane-intersection method, which supports the claim that row-wise noise accumulation can be mitigated by reconstructing the axis through plane intersection. Finally, phase correlation performs slightly better than cross-correlation in roll recovery, especially in the presence of illumination variation.

4. Discussion

The proposed method is motivated by a rocket-recovery application, but its geometric principle is not restricted to one specific rocket design. For pitch and yaw estimation, the method requires only that the observed body contain a sufficiently long, approximately straight, axis-symmetric segment. This condition is satisfied by many cylindrical or quasi-cylindrical aerial bodies, such as reusable stages, sounding rockets, and other elongated vehicles. Therefore, the method is more general than a design-specific contour template.
The midpoint-based central-axis approximation deserves further discussion. For a perfect circular cylinder under orthographic projection, the midpoint of the silhouette boundaries lies exactly on the image of the central axis. Under perspective projection, the approximation error depends on the ratio between body

radius and observation distance. In the present setup, this ratio is small, and the experimental results confirm that the approximation is sufficiently accurate. This is also indirectly supported by the fact that the proposed plane-intersection method outperforms point-by-point triangulation in the ablation study.

The roll-estimation ambiguity also requires clarification. The base-disc texture may exhibit approximate rotational repetition. In this paper, the ambiguity is resolved by restricting the phase-correlation search to a physically reasonable inter-frame angular window. Because the image sequence is acquired at 20 fps and the actual rotational change between adjacent frames is limited, the correct phase-correlation peak can be selected using temporal continuity. This mechanism avoids large discrete jumps in roll estimation and explains why the time-series result is smooth even when the base-disc appearance is not fully unique.

The current method also has limitations. First, it depends on sufficiently clear contour visibility and a usable base-disc appearance; strong occlusion, severe blur, or plume interference may reduce robustness. Second, the current quantitative evaluation is based on a representative 100-frame scale-model sequence rather than a full-scale outdoor recovery campaign. Third, the evaluation reference is an IMU-based attitude sequence rather than a laboratory-grade optical motion-capture system. These issues will be addressed in future work through larger-scale experiments, more diverse trajectories, and possible visual-inertial integration.

5. Conclusions

This paper proposed a large-field binocular-vision-based attitude determination method for rocket recovery. The method combines distortion-aware stereo calibration, robust contour extraction, midpoint-based central-axis fitting, plane-intersection reconstruction, and phase-correlation-based roll recovery under temporal continuity constraints. Experiments on a 1:20 cylindrical scale model using two Cyclone-16-300 cameras demonstrated RMS reprojection errors of 0.056 px and 0.066 px, a 3D checkpoint RMSE of 33.42 mm, and attitude RMSEs of 1.58°, 1.54° and 1.41° for roll, pitch, and yaw, respectively. Comparative experiments showed that the proposed method outperforms ORB + PnP, SGBM, and Chamfer-based baselines, while the ablation study verified the effectiveness of its main components.

Overall, the proposed method provides a practical and interpretable optical solution for external attitude measurement in rocket-recovery scenarios with a large observation field. Future work will focus on full-scale experiments, improved robustness under challenging outdoor conditions, and integration with inertial measurements.

Acknowledgements

The authors would like to sincerely thank the organizers of the ISPRS Congress for accepting this work for presentation and inclusion in the proceedings. The constructive comments and valuable suggestions provided by the anonymous reviewers are also gratefully acknowledged, which have significantly contributed to improving the quality of this manuscript.

References

El-Sheimy, N., Hou, H., Niu, X., 2008: Analysis and modeling of inertial sensors using Allan variance. *IEEE Transactions on Instrumentation and Measurement*, 57(1), 140–149. https://doi.org/10.1109/TIM.2007.908635.

Figueiredo, R.P., Moreno, P., Bernardino, A., 2015: Efficient pose estimation of rotationally symmetric objects. *Neurocomputing*, 150(Part A), 126–135. https://doi.org/10.1016/j.neucom.2014.07.070.

Fioravanti, D., Colombo, C., Allotta, B., 2009: Self-calibrated visual servoing with respect to axial-symmetric 3D objects. *Robotics and Autonomous Systems*, 57(4), 451–459. https://doi.org/10.1016/j.robot.2008.05.006.

Hirschmüller, H., 2008: Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 328–341. https://doi.org/10.1109/TPAMI.2007.1166.

Lepetit, V., Moreno-Noguer, F., Fua, P., 2009: EPnP: An accurate O(n) solution to the PnP problem. *International Journal of Computer Vision*, 81(2), 155–166. https://doi.org/10.1007/s11263-008-0152-6.

Liu, Z., Li, F., Li, X., Zhang, G., 2015: A novel and accurate calibration method for cameras with large field of view using combined small targets. *Measurement*, 64, 1–16. https://doi.org/10.1016/j.measurement.2014.11.027.

Penman, D.W., Alwesh, N.S., 2006: 3D pose estimation of symmetrical objects of unknown shape. *Image and Vision Computing*, 24(5), 447–454. https://doi.org/10.1016/j.imavis.2006.01.019.

Qin, T., Li, P., Shen, S., 2018: VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020.

Reddy, B.S., Chatterji, B.N., 1996: An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8), 1266–1271. https://doi.org/10.1109/83.506761.

Scharstein, D., Szeliski, R., Zabih, R., 2002: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1–3), 7–42.

Tong, X., Shi, H., Liu, X., Chen, P., Gan, Q., Gao, Y., Liu, Z., Hong, Z., Ye, Z., Xie, H., 2025: Precise pose measurement during reusable rocket launch and landing using high-speed videogrammetry under uncontrolled outdoor lighting conditions. *Measurement*, 247, 116786. https://doi.org/10.1016/j.measurement.2025.116786.

Wang, Z., Wu, Z., Zhen, X., Yang, R., Xi, J., 2013: An onsite structure parameters calibration of large FOV binocular stereovision based on small-size 2D target. *Optik*, 124(21), 5164–5169. https://doi.org/10.1016/j.ijleo.2013.03.077.

Zhang, X., Jiang, Z., Zhang, H., Wei, Q., 2018: Vision-based pose estimation for textureless space objects by contour points matching. *IEEE Transactions on Aerospace and Electronic Systems*, 54(5), 2342–2355. https://doi.org/10.1109/TAES.2018.2815879.