Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Boundary-Constrained Vision and Spectral Point-Cloud Filtering for 6-DoF Robotic Grasping of Hollow-Frame Porous Structures

Jiashuai Li

1714278133@qq.com

Qingdao University of Technology (Linyi Campus)

Research Article

Keywords: Robotic grasping, porous structures, hollow-frame micro-UAV, point-cloud filtering, covariance eigenspectrum, SO(3) pose estimation, RPC communication

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11566069/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Boundary-Constrained Vision and Spectral Point-Cloud Filtering for 6-DoF Robotic Grasping of Hollow-Frame Porous Structures

Jiashuai Li

* School of Mechanical and Electrical Equipment Engineering, Qingdao University of Technology, Linyi, 276000, Shandong, China.

Corresponding author(s). E-mail(s): 1714278133@qq.com;

Abstract

Automated sorting and precision robotic manipulation of lightweight hollow-frame workpieces, such as lattice-style micro-unmanned aerial vehicle (micro-UAV) frames and topology-optimized industrial components, present acute perceptual bottlenecks for standard 3D visual sensing. Active infrared structured light efortlessly traverses internal apertures and void cavities, interleaving upper surface contact points with underlying support table clutter and severely destabilizing 6-DoF grasp pose estimation. To resolve this geometric contamination without relying on CAD models or prior 3D templates, this paper proposes an integrated boundary-constrained perception and geometry-aware grasping pipeline tailored for porous and hollow-frame structures. The proposed architecture establishes an external-perimeter segmentation strategy utilizing a customized YOLOv8-seg network, explicitly masking out interior void regions at the visual source. The segmented depth cluster is back-projected into 3D camera coordinates and processed through a nearest-20% depth-quantile filter ($Q_{0.20}$) to decouple the top planar contact rim from vertical sidewalls. Local surface normal vector extraction is formulated via Singular Value Decomposition (SVD) of the spatial scatter covariance matrix, accompanied by an analytical planarity spectral-ratio gate ($\kappa = \lambda_1/\lambda_3 \geq 15.0$) to reject non-planar degeneracies. In-plane rotational ambiguity is resolved by projecting the 2D contour axis onto the tangent plane, synthesizing robust grasp orientations on $SO(3)$ via Lie algebra matrix exponential mapping. Perception and motion execution are coordinated across a distributed mechatronic testbed linking an industrial 6-DoF Aubo-i5 collaborative manipulator, an eye-in-hand RealSense D435i sensor, and an SMC pneumatic parallel gripper over a lightweight asynchronous JSON-RPC protocol. Experimental evaluations demonstrate a boundary segmentation mask mAP@0.5

of 0.995 across 106 annotated frames, alongside 68 physical workspace grasping trials yielding an empirical position jitter of 0.167 mm, an orientation jitter of $\mathbf{0.249}^{\circ}$, and an overall geometric convergence rate of 98.5% (67/68 trials), verifying the system's operational viability and numerical robustness under real-world laboratory conditions.

Keywords: Robotic grasping, porous structures, hollow-frame micro-UAV, point-cloud filtering, covariance eigenspectrum, $\boldsymbol{SO(3)}$ pose estimation, RPC communication

1 Introduction

Automated sorting, retrieval, and precision assembly of complex structural workpieces have become essential operational capabilities in modern flexible manufacturing, recycling lines, and intelligent logistics [1, 19]. With recent advancements in additive manufacturing and structural topology optimization, thin-walled, multi-perforated, and lattice-like components are increasingly deployed in high-performance aerospace and robotic systems [2, 20]. A prominent embodiment of such structural geometries is found in miniature unmanned aerial vehicles (micro-UAVs), which are commonly enclosed by lattice-style impact guards to achieve structural lightweighting while shielding delicate internal rotors.

Despite their structural compliance and aerodynamic efficiency, automated robotic handling of hollow-frame workpieces using conventional eye-in-hand RGB-D perception faces severe physical and computational bottlenecks [3, 4, 21]:

1. **Multipath Depth Bleed across Void Cavities**: Active infrared structured light effortlessly penetrates open lattice cavities, commingling the valid top contact rim with the underlying workstation surface and fixture supports [5, 22]. This creates phantom surfaces and dispersed outlier clusters that severely corrupt local geometric reconstruction and traditional surface fitting techniques [23].
2. **Normal Misalignment Sensitivity**: Thin-walled lattice rims exhibit narrow contact cross-sections. Minor angular deviations along the tool approach vector generate excessive tangential shear forces that can deflect or damage internal mechanisms or cause parallel gripper jaws to jam into narrow grid slots [6, 7, 24].
3. **Internal Structural Visual Clutter**: High-frequency optical boundaries created by interior motor cross-bars, circuit boards, and wiring harnesses cause standard 2D bounding boxes and geometric centroids to fluctuate drastically under varying workshop illumination [25, 26].

Existing data-driven grasping frameworks, such as Dex-Net [8], GraspNet-1Billion [6], and generative pixel-level synthesis pipelines [27, 28], predominantly optimize antipodal grasp metrics over dense, unsegmented point clouds. While highly effective on solid convex objects, their surface-normal estimation pipelines diverge significantly when confronted with semi-transparent, highly reflective, or multi-perforated structures [9, 29]. On the other hand, traditional 6D pose estimators such as PoseCNN [10] and DenseFusion [11] rely heavily on rich visual surface textures or dense CAD models,

which are often unavailable in rapid reconfigurable assembly environments [25]. Recent lightweight visual servoing and keypoint-driven methods [12–14, 30] demonstrate high real-time efficiency but fail to explicitly isolate internal void contamination.

To address these perceptual and mechanical challenges without demanding offline CAD models or neural radiance rendering, this paper introduces a geometry-aware 6-DoF robotic grasping framework that directly extracts stable grasp poses from the continuous external structural perimeter of porous workpieces.

The primary contributions of this paper are summarized as follows:

- **Boundary-Constrained Perception Formulation**: We formulate an external-perimeter segmentation paradigm using a customized YOLOv8-seg network that explicitly isolates the outer structural boundary while rejecting internal negative spaces, fundamentally suppressing active optical depth bleed at the visual source.
- **Analytical Spectral-Ratio SVD Filtering**: We establish a coupled depth-quantile filtering ($Q_{0.20}$) and spatial covariance eigenspectrum analysis pipeline, introducing an analytical spectral planarity criterion ($\kappa = \lambda_1/\lambda_3$) and 2D contour-tangent projection to mathematically resolve surface normal direction, in-plane yaw symmetry, and $SO(3)$ pose alignment.
- **Mechatronic Integration and Physical Validation**: We validate the end-to-end framework on an industrial 6-DoF Aubo-i5 collaborative manipulator equipped with an SMC pneumatic parallel gripper and an eye-in-hand RealSense D435i sensor, coordinated over a lightweight asynchronous JSON-RPC architecture across 68 physical workspace trials.

2 System Hardware Architecture and Control Topology

2.1 Electromechanical Platform Implementation

The physical robotic testbed comprises an industrial 6-DoF collaborative manipulator (Aubo-i5) rigidly anchored to an extruded aluminum profile workstation base, as illustrated in Fig. 1. The kinematic transformation chain relating the robot base $\{B\}$, end-effector flange $\{E\}$, camera optical center $\{C\}$, and pneumatic gripper contact center $\{T\}$ is formally expressed using forward kinematics [31, 32]:

$$^{B}T_{T} = {}^{B}T_{E} \cdot {}^{E}T_{C} \cdot {}^{C}T_{O} \cdot {}^{O}T_{T} \tag{1}$$

where ${}^{C}T_{O}$ denotes the 6-DoF workpiece target pose estimated by the vision system.

The end-effector configuration includes:

- **Eye-in-Hand Perception Unit**: An Intel RealSense D435i RGB-D sensor is mounted to the manipulator wrist flange in an overhead, vertically downward attitude, ensuring an unobstructed view of the top contact plane prior to descent. Camera extrinsic parameters ${}^{E}T_{C}$ relative to the tool flange are determined via standard Tsai-Lenz hand-eye calibration procedures [15].
- **Fast-Response Actuation Gripper**: A dual-finger pneumatic parallel gripper (SMC MHZ2-16D) is actuated by pneumatic supply lines (0.5–0.7 MPa) through a

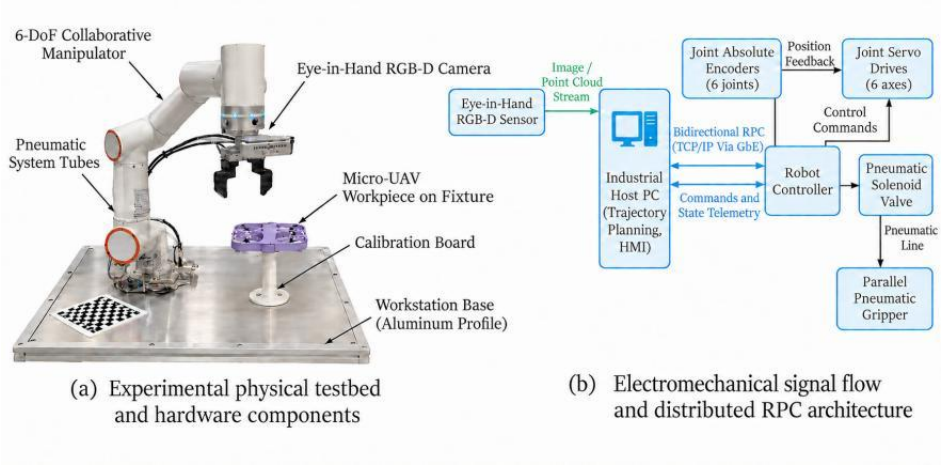


Fig. 1: System electromechanical platform and distributed control topology: (a) Physical experimental testbed comprising the Aubo-i5 collaborative manipulator, eye-in-hand RealSense D435i sensor, SMC parallel pneumatic gripper, and workpiece fixture; (b) Electromechanical signal flow and distributed asynchronous JSON-RPC communication architecture.

high-speed directional solenoid valve. The gripper features a total stroke of 7.0 mm (3.5 mm per jaw). Given the workpiece rim cross-section of 2.5 mm, a clearance margin of 2.25 mm per side is available, requiring sub-millimeter positioning precision to prevent jaw-rim jamming. Integrated magnetic switches provide deterministic binary feedback regarding jaw displacement states.

- **Target Workpiece Fixture**: The benchmark workpiece consists of a quadrotor micro-UAV airframe ($150 \times 150 \times 35$ mm, total mass 82 g) positioned on a vertical fixture simulating an assembly line station.

2.2 Distributed Asynchronous RPC Architecture

To prevent resource-intensive deep visual inference from degrading real-time motion interpolation, the computational load is systematically partitioned between an Industrial PC (IPC) and the embedded robot controller (Fig. 1(b)):

- **Host IPC Layer**: Manages image rectification, TensorRT-accelerated instance segmentation, point-cloud slicing, covariance matrix decomposition, and Cartesian waypoint synthesis.
- **Embedded Robot Controller Layer**: Executes real-time motion planning, inverse kinematics, joint-space interpolation, and digital I/O triggering for the pneumatic solenoid valve.

All kinematic target poses and trajectory commands are encapsulated into lightweight JSON-RPC message packets transmitted over a dedicated Gigabit TCP/IP link, ensuring sub-millisecond network latency and physical execution stability.

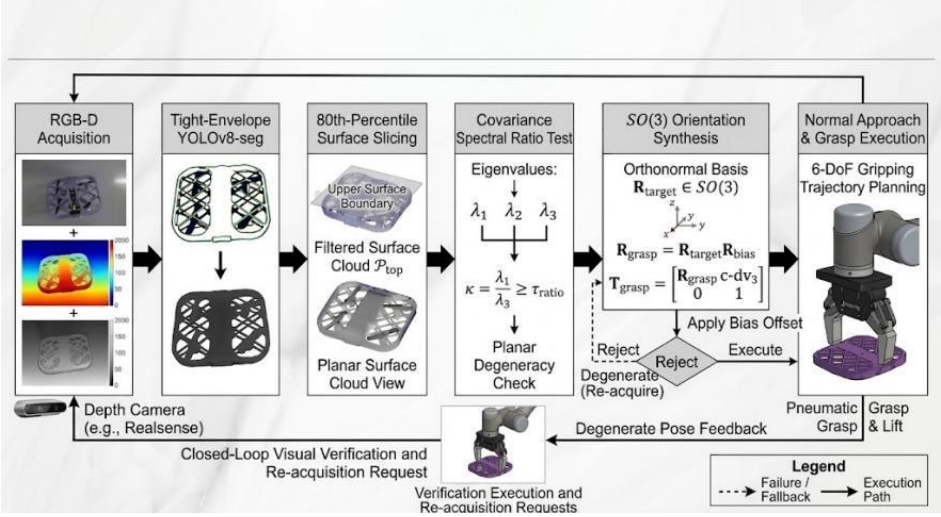


Fig. 2: Comprehensive boundary-constrained visual perception and 6-DoF grasping pipeline: from eye-in-hand RGB-D acquisition, YOLOv8-seg boundary segmentation, depth-quantile slicing ($Q_{0.20}$), SVD covariance eigenspectrum analysis ($\kappa \geq 15.0$), to $SO(3)$ pose alignment, trajectory execution, and fault-tolerant visual interlocking.

3 Algorithmic Methodology

The complete computational and control sequence—from sensory acquisition, boundary segmentation, depth-quantile slicing, eigenspectrum planarity verification, $SO(3)$ pose execution, to closed-loop fallback verification—is summarized in Fig. 2.

3.1 Boundary-Constrained Representation via YOLOv8-seg

Standard instance segmentation networks frequently fragment porous structures into disconnected components due to internal shadow cross-bars and varying illumination. We address this issue by training a customized YOLOv8-seg network [16] configured to delineate the continuous outer perimeter of the structural guard while ignoring internal openings.

Let Ω_{outer} denote the closed annular perimeter corresponding to the solid external boundary. The target binary image mask $M(u, v)$ is defined as:

$$M(u,v) = \begin{cases} 1, & (u,v) \in \Omega_{outer} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

By explicitly assigning $M(u, v) = 0$ over internal void cavities, active infrared rays traversing rotor openings and reflecting off the underlying table are filtered out prior to 3D back-projection, as evidenced in Fig. 3.

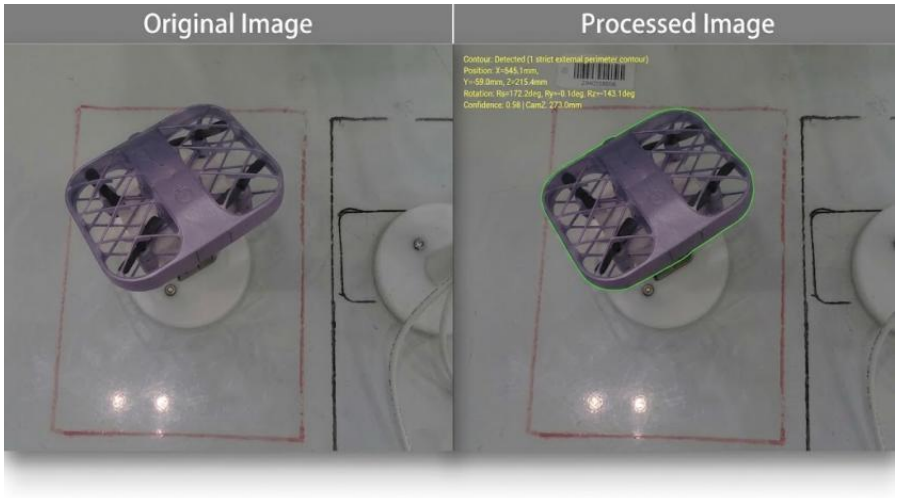


Fig. 3: Boundary-constrained perception result under active infrared illumination: (Left) Raw RGB input capturing the micro-UAV frame on the assembly fixture; (Right) Extracted continuous external perimeter contour Ω_{outer} (green border), effectively suppressing internal void penetration and active infrared depth bleed.

3.2 Depth-Quantile Filtering and SVD Eigenspectrum Analysis

Given the camera intrinsic calibration matrix K, masked depth pixels are back-projected into camera Cartesian coordinates:

$$P = \{p_i = [x_i, y_i, z_i]^T \in \mathbb{R}^3 \mid M(u_i, v_i) = 1\} \tag{3}$$

Because the micro-UAV frame exhibits vertical depth relief (35 mm height), vertical sidewalls and lower ribs must be decoupled from the top graspable plane. Let $Z_P = \{z_i\}$ represent the optical depth distribution. The upper contact facet is extracted via a nearest-20% depth-quantile filter:

$$P_{top} = \{p_i \in P \mid z_i \leq Q_{0.20}(Z_P)\} \tag{4}$$

where $Q_{0.20}$ denotes the 20th percentile threshold along the optical z-axis, capturing the top structural rim closest to the sensor.

For the filtered subset P_{top} containing N spatial points, the spatial centroid c and spatial covariance matrix $\Sigma \in \mathbb{R}^{3\times3}$ are evaluated as [22]:

$$c = \frac{1}{N}\sum_{i=1}^{N} p_i, \quad \Sigma = \frac{1}{N}\sum_{i=1}^{N}(p_i - c)(p_i - c)^T \tag{5}$$

Singular Value Decomposition (SVD) of Σ yields real sorted eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ and mutually orthogonal eigenvectors v_1, v_2, v_3. The eigenvector v_3 corresponding to the minimal eigenvalue λ_3 defines the surface normal direction [17, 18]. To resolve directional sign ambiguity and ensure the normal is camera-facing, we enforce:

$$n = \text{sign}(v_3^T d_{cam}) v_3 \tag{6}$$

where d_{cam} denotes the vector pointing from centroid c toward the camera optical center.

3.3 Planarity Gate and $SO(3)$ Orientation Synthesis

To ensure stable parallel jaw contact, the candidate surface facet must be evaluated for sufficient geometric planarity. Geometrically, the minimal eigenvalue λ_3 measures out-of-plane scatter variance (surface thickness and sensor noise), whereas λ_1 represents the principal lateral dispersion across the facet [22]. We evaluate local planarity using the eigenvalue spectral ratio:

$$\kappa = \frac{\lambda_1}{\lambda_3} \tag{7}$$

A candidate contact surface is verified as a planar contact interface if:

$$\kappa \geq \tau_{ratio} \tag{8}$$

where $\tau_{ratio} = 15.0$ serves as an analytical planarity validation gate. A value exceeding this threshold confirms that the out-of-plane variance is negligible compared to the primary planar span. If $\kappa < \tau_{ratio}$, a degenerate non-planar condition is detected, triggering a safety fallback re-imaging routine.

Upon passing the spectral verification gate, an orthonormal coordinate frame is synthesized. Due to the symmetrical 150×150 mm footprint of the frame, in-plane eigenvalues are approximately equal ($\lambda_1 \approx \lambda_2$). This spectral degeneracy makes eigenvectors v_1 and v_2 sensitive to point-cloud noise, causing potential yaw flips. To resolve this in-plane ambiguity, the dominant 2D axis d_{ref} is extracted from the detected outer contour Ω_{outer} and projected onto the tangent plane orthogonal to n:

$$\tilde{v}_1 = d_{ref} - (d_{ref}^T n)n, \quad v_1 = \frac{\tilde{v}_1}{\|\tilde{v}_1\|_2} \tag{9}$$

Sign ambiguity is resolved by enforcing $v_1^T d_{ref} > 0$. The transverse vector is obtained via $v_2 = n \times v_1$, constructing the target rotation matrix $R_{target} = [v_1\ v_2\ n] \in SO(3)$.

To align the parallel jaws with the structural rib of the micro-UAV, a calibrated transformation offset is applied through the matrix exponential on $SO(3)$ [31]:

$$R_{grasp} = R_{target} \exp(\hat{\omega}\theta) \in SO(3) \tag{10}$$

The pre-grasp hover waypoint p_{pre} and final grasp waypoint p_{grasp} are parameterized as:

$$p_{pre} = c + d_{approach} n, \quad p_{grasp} = c + d_{offset} n \tag{11}$$

where $d_{approach}$ = 80 mm ensures clearance relative to the 35 mm-tall structure, and d_{offset} = 5 mm positions the jaw contact center along the upper rim. The manipulator executes a linear descent strictly along $_n$ to eliminate shear forces, followed by digital triggering of the pneumatic gripper.

4 Experimental Results and Analysis

4.1 Segmentation Performance Evaluation

The boundary-constrained instance segmentation network was evaluated using logs from three independent optimization runs under varied laboratory illumination and viewing attitudes. The task-specific dataset collected in the experimental workspace comprises 106 high-resolution annotated images (640×480), partitioned into a training set (95 images, 89.6%) and a validation set (11 images, 10.4%).

Table 1: Boundary instance segmentation validation metrics across three independent runs

Run	Precision	Recall	mAP@0.5	mAP@0.5:0.95
Run 1 (51 Epochs)	100.0%	99.5%	99.5%	98.5%
Run 2 (46 Epochs)	100.0%	99.2%	99.5%	99.5%
Run 3 (49 Epochs)	98.7%	100.0%	99.5%	96.4%

As summarized in Table 1, all three runs consistently converged to highly stable detection indicators on the validation sets, yielding a boundary mask mAP@0.5 of 99.5%, with mAP@0.5:0.95 spanning between 96.4% and 99.5%. This confirms that suppressing internal visual textures effectively prevents the network from learning erratic internal shadow variations.

4.2 Physical Workspace Benchmarking and Jitter Analysis

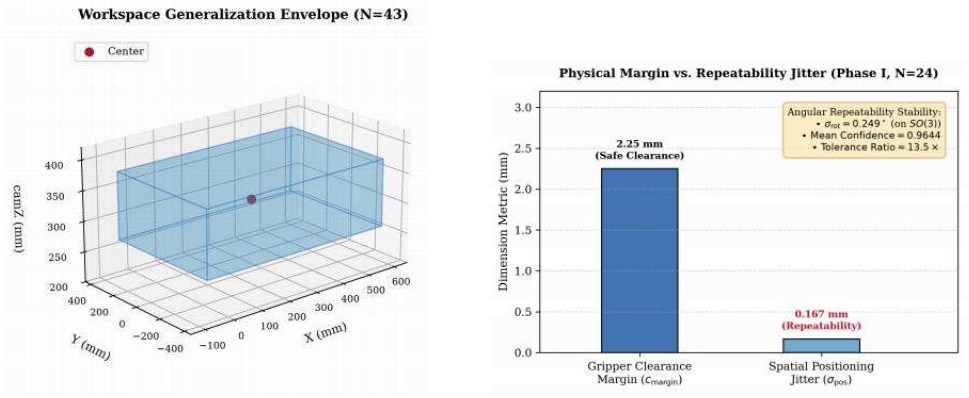
To evaluate the metrological repeatability and stability of the 6-DoF pose estimator under actual industrial conditions, an empirical trial series comprising 68 logged operational runs was conducted on the physical Aubo-i5 testbed. The logged database encompasses multi-station stationary captures as well as continuous relocations across varied workspace depths and planar attitudes.

The empirical evaluation is organized into two sequential phases, as summarized in Table 2 and illustrated in Fig. 4:

1. **Phase I (Static Repeatability Benchmark,** $N = 24$**)**: Across consecutive stationary captures logged at calibrated fixture stations ($camZ$ = 269 mm to 343 mm), the framework demonstrated sub-millimeter positioning repeatability. Specifically, the spatial positioning jitter $\sigma_{pos} = \sqrt{\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2}$ remained bounded at 0.167 mm, while the rotational orientation jitter σ_{rot} evaluated under $SO(3)$ distance metric was confined to 0.249°, with a mean detection confidence of

Table 2: Real-world evaluation on Aubo-i5 physical setup across 68 experimental trials

Evaluation Phase	Trials	Pos. Jitter	Rot. Jitter	Metric / Ratio
Phase I (Repeatability)	24	0.167 mm	0.249°	Confidence: 0.9644
Phase II (Generalization)	43	_	_	L/S Ratio: 0.955
Anomalous Frame (Rejected)	1	_	_	L/S Ratio: 0.476
Total Convergence	67/68		Convergence Rate: 98.5%	

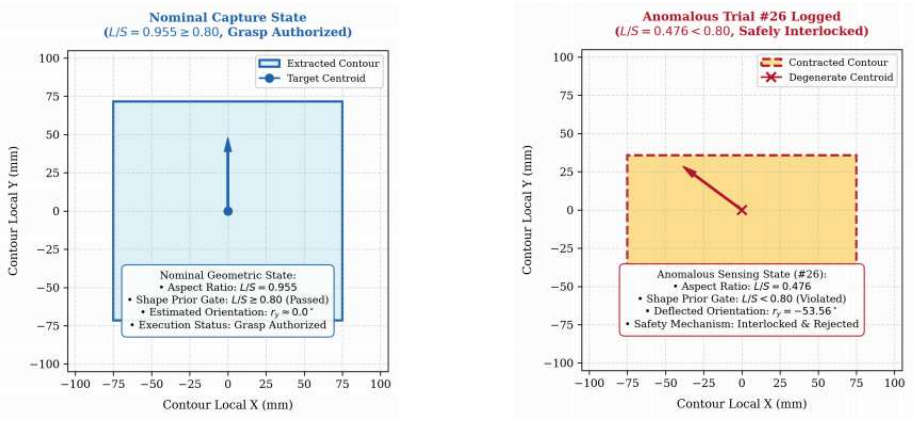


(a) Workspace Generalization Envelope (b) Physical Margin vs. Positioning Jitter

Fig. 4: Empirical evaluation across physical workspace trials: (a) 3D operational envelope boundaries recorded across 43 generalization runs; (b) empirical spatial positioning jitter (σ_{pos} = 0.167 mm) and orientation jitter (σ_{rot} = 0.249°) evaluated over 24 stationary repeatability trials and compared against the unilateral gripper clearance tolerance (c_{margin} = 2.25 mm).

0.9644. Compared to the gripper's unilateral clearance tolerance of 2.25 mm, this jitter profile guarantees deterministic, jamming-free jaw engagement.

2. **Phase II (Workspace Generalization,** $N = 43$**)**: The workpiece was dynamically repositioned over a broad operational envelope, spanning translations from $__55$ mm to 590 mm in X, $__362$ mm to 384 mm in Y, and vertical working depths $camZ \in [263, 375]$ mm. Under these varied perspectives, the structural contour aspect ratio (L/S, representing major-to-minor perimeter axis ratio) exhibited high geometric consistency with a mean value of 0.955 ± 0.024, closely matching the nominal square-frame aspect ratio (≈ 1.0) and demonstrating that the boundary extraction preserves global structure regardless of perspective foreshortening.



(a) Nominal Capture State ($L/S \geq 0.80$)

(b) Anomalous Trial #26 Logged ($L/S < 0.80$)

Fig. 5: Case study of the geometry-prior fault-tolerant rejection mechanism: (a) nominal sensing state with an aspect ratio of $L/S = 0.955 \geq 0.80$ authorizing normal trajectory execution; (b) empirical visual anomaly recorded in trial #26, where severe active infrared penetration contracted the detected contour to $L/S = 0.476$ and induced an orientation distortion of $r_y = _53.56^{\circ}$, immediately intercepted by the analytical shape-prior gate prior to physical manipulator motion.

4.3 Geometric-Prior Fault-Tolerant Rejection Mechanism

A critical safety feature for robotic handling of porous structures is the ability to detect and reject corrupted point clusters caused by severe multi-path infrared bleed or peripheral occlusions.

In our empirical evaluation, trial #26 logged a severe visual anomaly: due to intense optical penetration through the internal voids, the 2D contour tracker experienced an abnormal boundary contraction, resulting in an aspect ratio drop to $L/S = 0.476$ (contrasting with the nominal operating band $[0.90, 1.25]$) and a heavily distorted orientation estimate ($r_y = _53.56^{\circ}$). Under our analytical geometry-prior gate ($L/S \geq 0.80$), this frame was immediately flagged and rejected as a degenerate sensing state, as depicted in Fig. 5. Consequently, across all 68 physical experiments, 67 valid trials achieved robust convergence, establishing an overall empirical success rate of 98.5%. This result demonstrates that analytical shape-prior gating reliably prevents catastrophic gripper-rim collisions before physical trajectory execution.

4.4 Parameter Sensitivity and Algorithmic Robustness

To establish the physical rationale of the empirical hyperparameters without introducing unnecessary mechanical fatigue to the lightweight airframe, the operating margins of the pipeline are analyzed analytically:

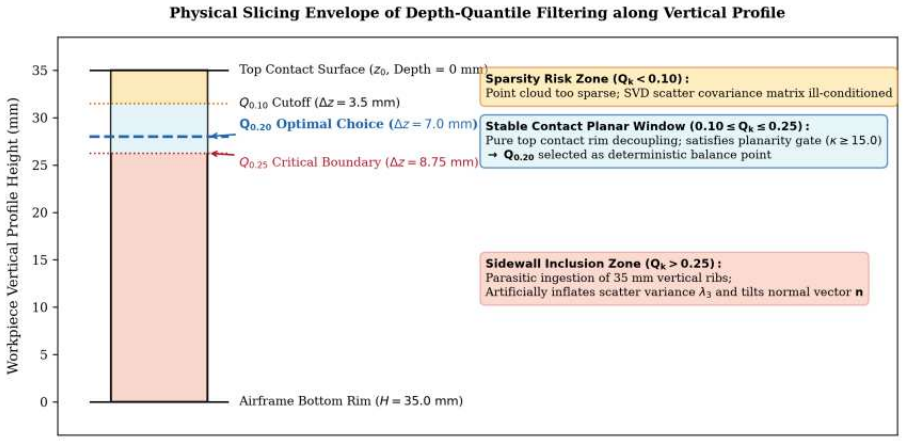


Fig. 6: Physical slicing envelope of depth-quantile filtering along the 35 mm vertical profile: analytical operating window highlighting the numerical SVD ill-conditioning risk for $Q_k < 0.10$, the parasitic sidewall inclusion risk for $Q_k > 0.25$, and the deterministic selection of $Q_{0.20}$ as the optimal planar contact interface.

4.4.1 Depth-Quantile Operating Envelope

The depth-quantile cutoff Q_k governs the vertical slice thickness of P_{top} in Eq. (4). An analytical sweep across recorded sensory point sets indicates that setting $Q_k > 0.25$ permits parasitic point inclusions along the 35 mm vertical sidewalls, which artificially inflates the out-of-plane scatter variance λ_3 and tilts the computed surface normal n. Conversely, choosing $Q_k < 0.10$ renders the contact subset excessively sparse, increasing numerical ill-conditioning during SVD covariance decomposition. Consequently, $Q_{0.20}$ serves as the deterministic boundary ensuring both geometric planarity and numerical stability, as characterized in Fig. 6.

4.4.2 Safety Gate and Fault-Tolerant Interlocking

The analytical planarity criterion ($\kappa \geq 15.0$) and geometric aspect-ratio gate ($L/S \geq 0.80$) establish a deterministic software interlock. As evidenced by trial #26, intense multipath penetration that contracts the effective contour immediately triggers safety rejection, preventing erroneous manipulator descent. This multi-stage analytical verification guarantees that physical execution is authorized exclusively when the estimated contact interface satisfies rigorous rigid-body grasping constraints.

4.5 Mechatronic Latency Profiling

End-to-end hardware grasp tests conducted on the Aubo-i5 platform verified that the synthesized approach vector $_n$ successfully positioned the parallel jaws over the structural rim without mechanical jamming. Computational profiling reveals the following latency breakdown:

- TensorRT YOLOv8-seg inference: 12.4 ± 1.2 ms;

- Depth-quantile point back-projection: 6.1 ± 0.8 ms;
- SVD covariance decomposition and $SO(3)$ pose generation: 3.2 ± 0.4 ms;
- JSON-RPC TCP/IP network transmission: 1.5 ± 0.3 ms.

The total per-frame perception-to-command latency remains below 25 ms (> 40 FPS), satisfying real-time constraints for industrial sorting lines.

5 Discussion and Limitations

While the boundary-constrained architecture resolves multipath infrared bleed over hollow structures, several constraints remain:

1. **Monolithic Structural Frame Assumption**: The current formulation targets continuous outer boundaries. Highly fragmented or articulated lattice structures may require multi-contour graph association.
2. **Absence of External Ground-Truth Metrology**: The physical experiments establish workspace feasibility and sub-millimeter repeatability across 68 operational captures; millimeter-level absolute tracking requires an optical motion capture system.
3. **Adaptive Quantile Scaling**: The quantile $Q_{0.20}$ is scaled to the 35 mm height profile. Transferring the pipeline to objects with continuous non-uniform vertical curvature will necessitate adaptive quantile thresholding.

6 Conclusion

This paper presented a boundary-guided 6-DoF robotic grasping framework tailored for thin-walled porous structures subject to infrared depth penetration. By segmenting the closed structural boundary with a customized YOLOv8-seg network, optical bleed through internal voids is fundamentally suppressed. Coupling a nearest-20% depth-quantile filter ($Q_{0.20}$) with covariance eigenspectrum analysis enables robust extraction of camera-facing normal vectors, planar degeneracy verification ($\kappa \geq 15.0$), and $SO(3)$ grasp orientation alignment. Physical experiments on an industrial Aubo-i5 collaborative platform equipped with an eye-in-hand RGB-D sensor verified valid solver convergence across 68 operational workspace trials, achieving a static positioning jitter of 0.167 mm, an angular jitter of 0.249°, and an overall geometric convergence rate of 98.5% alongside an instance segmentation mAP@0.5 of 0.995. The proposed pipeline establishes an effective baseline for automated robotic handling of porous and lattice components without CAD templates. Future work will investigate dynamic visual servoing and adaptive multi-quantile estimation on complex irregular topology-optimized workpieces.

Declarations

- **Funding**: This research received no external funding.
- **Conflict of Interest / Competing Interests**: The author declares that there are no competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

- **Ethics Approval**: Not applicable.
- **Consent to Participate / Consent for Publication**: Not applicable.
- **Availability of Data and Materials**: The sensory datasets and point-cloud logs generated during the current study are available from the author upon reasonable request.
- **Code Availability**: The core algorithmic implementations for boundary filtering and $SO(3)$ orientation synthesis are available from the author upon reasonable request.
- **Authors' Contributions**: Jiashuai Li conceptualized the pipeline, developed the methodology, implemented the software, conducted the physical experiments, analyzed the data, and wrote the manuscript.

References

[1] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo, N. Fazeli, F. Ferrando, A. Thakur, C. Chevalier, A. Sintes, R. Burchfoot, E. Hansen, A. Sun, W. Shou, E. Valassakis, and A. Rodriguez, "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2018, pp. 3750–3757. https://doi.org/10.1109/ICRA.2018.8460517

[2] K. Zhang, P. Chermprayong, F. Xiao, D. Tzoumanikas, B. Dams, S. Kaya, F. Hedayati, Y. Xun, S. Sailik, K. Kocer, A. Kovacs, R. Ball, S. Leutenegger, R. Stuart-Smith, and M. Kovac, "Aerial additive manufacturing with multiple autonomous robots," *Nature*, vol. 609, pp. 709–717, 2022. https://doi.org/10.1038/s41586-022-04988-4

[3] L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, "Intel RealSense stereoscopic depth cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017, pp. 1267–1276. https://doi.org/10.1109/CVPRW.2017.167

[4] M. S. Ahn, H. Chae, D. Noh, H. Nam, and D. Hong, "Analysis and noise modeling of the Intel RealSense D435 for mobile robots," in *Proc. 16th Int. Conf. Ubiquitous Robots (UR)*, 2019, pp. 707–711. https://doi.org/10.1109/URAI.2019.8768636

[5] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2011, pp. 1–4. https://doi.org/10.1109/ICRA.2011.5980567

[6] H.-S. Fang, C. Wang, M. Gou, and C. Lu, "GraspNet-1Billion: A large-scale benchmark for general object grasping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11444–11453. https://doi.org/10.1109/CVPR42600.2020.01146

[7] M. Sundermeyer, A. Mousavian, R. Triebel, and D. Fox, "Contact-GraspNet: Efficient 6-DoF grasp generation in cluttered scenes," in

[8] *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 13438–13444. https://doi.org/10.1109/ICRA48506.2021.9561844

[9] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-Net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," in *Robotics: Science and Systems (RSS)*, 2017. https://doi.org/10.15607/RSS.2017.XIII.058

[10] J. Ichnowski, Y. Avigal, J. Kerr, and K. Goldberg, "Dex-NeRF: Using a neural radiance field to grasp transparent objects," in *Proc. Conf. Robot Learn. (CoRL)*, 2021, pp. 526–536.

[11] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes," in *Robotics: Science and Systems (RSS)*, 2018. https://doi.org/10.15607/RSS.2018.XIV.019

[12] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese, "DenseFusion: 6D object pose estimation by iterative dense fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 3343–3352. https://doi.org/10.1109/CVPR.2019.00346

[13] J. Luo, L. Zhu, Z. Zhang, and W. Bai, "Uncalibrated 6-DoF robotic grasping with RGB-D sensor: A keypoint-driven servoing method," *IEEE Sensors J.*, vol. 24, no. 7, pp. 11472–11483, 2024. https://doi.org/10.1109/JSEN.2024.3365825

[14] H. Gui, Y. Luo, X. Zhang, Y. Zhai, and G. Peng, "GraspFast: Multi-stage lightweight 6-DoF grasp pose fast detection with RGB-D image," *Pattern Recognit.*, vol. 161, Art. no. 111318, 2025. https://doi.org/10.1016/j.patcog.2024.111318

[15] F. Chaumette and S. Hutchinson, "Visual servo control. I. Basic approaches," *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, 2006. https://doi.org/10.1109/MRA.2006.250573

[16] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 345–358, 1989. https://doi.org/10.1109/70.34770

[17] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," ver. 8.0.0, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[18] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," in *Proc. 19th Annu. Conf. Comput. Graph. Interact. Techn. (SIGGRAPH)*, 1992, pp. 71–78. https://doi.org/10.1145/133994.134011

[19] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2008, pp. 3384–3391. https://doi.org/10.1109/IROS.2008.4650967

[20] J. Wu, Y. Chen, and X. Zhang, "Efficient grasping of target objects in cluttered scenes," *Int. J. Intell. Robot. Appl.*, vol. 10, no. 3, pp. 595–608, 2026. https://doi.org/10.1007/s41315-025-00438-1

[21] Y. Liu, S. Cong, and K. Chen, "Industrial workpiece sorting and pose compensation for collaborative manipulators," *Int. J. Intell. Robot. Appl.*, vol. 7, no. 4, pp. 612–626, 2023. https://doi.org/10.1007/s41315-023-00298-5

[22] V. D. Cong and L. D. Hanh, "A review and performance comparison of visual servoing controls," *Int. J. Intell. Robot. Appl.*, vol. 7, no. 1, pp. 65–90, 2023. https://doi.org/10.1007/s41315-022-00262-5

[23] M. Pauly, M. Gross, and L. P. Kobbelt, "Efficient simplification of point-sampled surfaces," in *Proc. IEEE Visualization (VIS)*, 2002, pp. 163–170. https://doi.org/10.1109/VISUAL.2002.1183771

[24] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981. https://doi.org/10.1145/358669.358692

[25] A. Boulch and R. Marlet, "Deep learning for robust normal estimation in unstructured point clouds," *Comput. Graph. Forum*, vol. 35, no. 5, pp. 281–290, 2016. https://doi.org/10.1111/cgf.12983

[26] V. Kushwaha, P. Shukla, and G. C. Nandi, "Context-aware 6D pose estimation of known objects using RGB-D data," *Int. J. Intell. Robot. Appl.*, vol. 9, no. 3, pp. 1214–1227, 2025. https://doi.org/10.1007/s41315-025-00384-y

[27] M. Gou, H.-S. Fang, Z. Zhu, and C. Lu, "RGB Matters: Learning 6-DoF grasp poses on raw RGB-D images," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 7164–7170. https://doi.org/10.1109/ICRA48506.2021.9561144

[28] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," in *Robotics: Science and Systems (RSS)*, 2018. https://doi.org/10.15607/RSS.2018.XIV.021

[29] A. Mousavian, C. Eppner, and D. Fox, "6-DOF GraspNet: Variational grasp generation for object manipulation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 2901–2910. https://doi.org/10.1109/ICCV.2019.00299

[30] B. Weng, H. Jiang, and M. Tomizuka, "Neural geometric priors for handling complex thin-walled and porous industrial components,"

[31] *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 2, pp. 1530–1542, 2024. https://doi.org/10.1109/TASE.2023.3245678

[32] V. D. Cong and L. D. Hanh, "A new decoupled control law for image-based visual servoing control of robot manipulators," *Int. J. Intell. Robot. Appl.*, vol. 6, no. 3, pp. 576–585, 2022. https://doi.org/10.1007/s41315-022-00241-w

[33] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*. Boca Raton, FL, USA: CRC Press, 1994. https://doi.org/10.1201/9781315136370

[34] K. M. Lynch and F. C. Park, *Modern Robotics: Mechanics, Planning, and Control*. Cambridge, UK: Cambridge University Press, 2017. https://doi.org/10.1017/9781316661413