Autonomous Docking for Rebar Tying Robots: A Structure-Aware Perception Framework

Feng Zhang[1], Jingjing Guo[1, 2], Xiaoyi Lyu[1], Lu Deng[1, 2*], Yuning Yan[1], Bo Jin[1, 3], Jichao Hou[4, 5]

[1] College of Civil Engineering, Hunan University, Changsha, China
[2] Key Laboratory of Damage Diagnosis for Engineering Structures of Hunan Province, Hunan University, Changsha, China
[3] Hunan Huda Construction Supervision Co., Ltd, Changsha, China
[4] State Grid Jiaxing Electric Power Supply Company
[5] Jiaxing Hengguang Electric Power Construction Co., Ltd
zhangfeng86@hnu.edu.cn, guojingjing@hnu.edu.cn, lyuxiaoyi@hnu.edu.cn, denglu@hnu.edu.cn, yanyuning@hnu.edu.cn, jinbo@hnu.edu.cn, jchou@hnu.edu.cn

Abstract

Autonomous rebar tying requires precise mobile base docking relative to the rebar cage to ensure effective manipulation. However, this process is challenged by the sparse, repetitive, and texture-less nature of rebar structures, which leads to perceptual aliasing and measurement instability in conventional LiDAR and visual perception systems. To address this challenge, this paper proposes a structure-aware perception method to achieve robust and precise pose estimation. The method employs a hierarchical filtering and semantic abstraction module to extract robust structural primitives from noisy RGB-D data. Subsequently, a Structural Congruence RANSAC (SC-RANSAC) algorithm is introduced to infer the optimal spatial pose. Unlike traditional model fitting, SC-RANSAC incorporates geometric priors (coplanarity, orthogonality and coverage) to validate structural integrity. Experimental results demonstrate that SC-RANSAC achieves a 100% success rate in identifying the target plane, compared to 36–52% for state-of-the-art RANSAC variants. The proposed system maintains translation errors within 16.9 mm and orientation errors within 1.7°, exhibiting robustness against viewpoint variations, thin rebar diameters (down to 6 mm), and partial occlusions (up to 50%), demonstrates its feasibility in complex and variable construction scenarios.

Keywords:

Construction Robotics, Rebar Tying, Pose Estimation, Robot Vision, Structure-Aware Perception

1 INTRODUCTION

The construction industry, a cornerstone of the global economy, has long suffered from stagnant productivity compared to other industrial sectors [1,2]. Due to demographic trends and changing career preferences, the industry faces a structural shortage of skilled workers [3]. Concurrently, increasingly stringent standards for safety and quality traceability have intensified the urgency for automation. Consequently, introducing robotics to perform dangerous, heavy, or repetitive tasks has become a critical path for improving efficiency and quality [4].

In concrete structure construction, rebar engineering is core to structural integrity [5]. However, rebar tying remains predominantly manual, characterized by procedural repetition and physical arduousness. Workers must repeat identical tying actions thousands of times on dense rebar cages, facing significant ergonomic risks [6,7]. While this highly regular scenario makes it an ideal candidate for automation, the complexity of unstructured sites prevents existing mobile robots from operating stably without human intervention [8].

A fully autonomous mobile rebar tying system logically consists of three stages: (1) navigation; (2) fine-grained docking; and (3) manipulation. While recent research focuses heavily on stages (1) and (3), autonomously establishing an optimal base pose relative to the rebar cage remains a critical yet overlooked challenge, connecting macro-mobility with micro-manipulation. Suboptimal docking causes arm operations to become unreachable or imprecise, and disrupts the seamless coverage of adjacent work zones in segmented operations. The core difficulty stems not from recognizing disordered objects, but from the geometric regularity of the rebar cage and the limitations of sensor perception. Besides, since tasks cannot be pre-programmed due to unstructured environments and inaccurate mobile positioning [9], the robot must rely entirely on on-site perception.

To obtain the visual information of surrounding environment, 3D LiDARs are often favored for their precision. However, they prove inadequate in this specific scenario. The sparse scanning patterns of mechanical LiDARs frequently slip through the thin rebar rods, resulting in discontinuous point clouds and significant information loss [10]. Furthermore, LiDARs are not immune to the local measurement instability caused by the slender geometry of metal [11,12]. Industrial structured light cameras lack the frame rate required for dynamic visual servoing and are cost-prohibitive for scalable deployment [13,14].

On the other hand, consumer-grade RGB-D cameras, such as Time-of-Flight (ToF) or active stereo systems, offer a cost-effective solution with dense point clouds aligned to color data and frame rates sufficient for real-time servoing [15]. Nevertheless, practical deployment encounters two primary challenges:

- **Global Perceptual Ambiguity:** The grid structure, composed of numerous identical intersection points, induces severe perceptual aliasing. Within the robot's Field of View (FOV), these nodes are geometrically and visually indistinguishable.
- **Measurement Instability and Sensitivity:** Rebars are texture-less, reflective, and slender cylinders. The inherent spatial sparsity of the cage structure, where voids exceed solid elements, results in scarce effective depth information within the sensor field of view [18]. Furthermore, ToF sensors suffer from multipath interference on reflective surfaces and flying pixels at depth discontinuities [19]. Consequently, orientational errors could be amplified when estimating pose from these sparse and noisy 3D points, rendering attitude estimation unstable.

In order to extract stable structural poses from sparse and noisy data using cost-effective sensors, this research proposes a perception framework that leverages geometric repetition as a constraint rather than searching for unique features. Global structural regularity is utilized to validate uncertain local perception data. The specific innovations are:

1. **Hierarchical Robust Perception:** To address feature sparsity and sensor noise, a layered approach combining surface filtering, semantic keypoint extraction, and information refinement is introduced. This method utilizes point cloud redundancy to extract stable structural primitives from noisy raw data, ensuring robustness against outliers.
2. **Structure-Aware Reasoning:** To resolve perceptual aliasing in repetitive scenarios, Structural Congruence RANSAC (SC-RANSAC) is proposed. Unlike traditional methods that rely solely on inlier counts [20], SC-RANSAC employs a scoring system based on coplanarity, orthogonality, and coverage. This approach identifies the optimal physical plane by balancing data fit with geometric regularity.

2 METHODOLOGY

This research presents a perception framework to address the challenges of perceptual ambiguity and measurement instability in rebar tying. Instead of relying on local appearance, the framework utilizes intrinsic structural rules. The framework targets the rebar cage and consists of two core modules (**Figure 2**): (1) **Hierarchical Filtering and Semantic Abstraction:** Extracts robust structural primitives from complex scenes for comprehensive perception; (2) **Structural Congruence RANSAC (SC-RANSAC):** Introduces structural rules as priors and constraints to infer the optimal spatial pose from noisy sensor data.

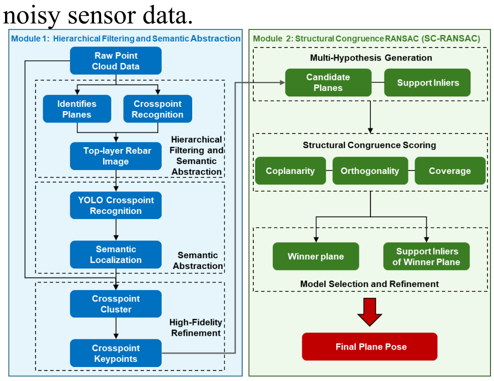


Figure 2. The overview architecture of methodology

2.1 Hierarchical Filtering and Semantic Abstraction

The goal of this stage is to extract a set of high-precision, high-confidence structural primitives from raw sensor data containing noise and background interference. This is achieved through a coarse-to-fine process divided into three sub-stages.

2.1.1 Coarse Geometric Filtering

In rebar tying tasks, the outermost rebar layer directly in front of the robot is the primary interaction object. To reduce computational load and handle foreground clutter and multi-layer overlaps, raw point cloud data undergoes depth-based iterative layered segmentation. First, RANSAC identifies potential planes in the scene. Based on depth information, these planes are iteratively scanned from near to far using semantic crosspoint recognition (based on YOLO algorithm from **Section 2.1.2**). Planes lacking semantic features are treated as foreground clutter and discarded. The process stops when rebar crosspoints are successfully detected, and this plane is defined as the Structural Datum Surface. Finally, all 3D points belonging to this datum surface are projected to generate a 2D Top-Layer Rebar Image (**Figure 3**). This output provides an input for subsequent semantic detection with large-scale background and foreground clutter removed.

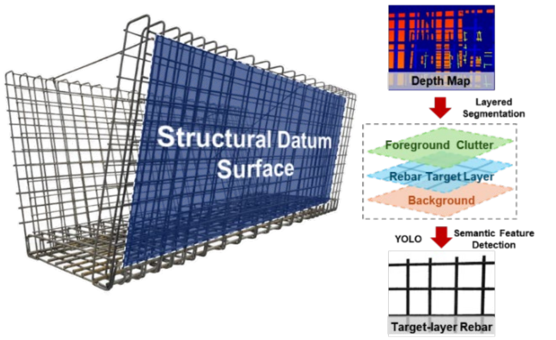


Figure 3. Generated top-layer image for semantic analysis

2.1.2 Semantic Abstraction

The rebar crosspoint is defined as the core structural primitive of the rebar cage. This is chosen because crosspoints possess distinct geometric features in point density and local curvature, and they directly relate to the target of the robot's end-effector tying task. The top-layer rebar image generated in the previous step serves as input to a deep learning-based crosspoint detector [13]. The detector outputs a series of 2D semantic detection results. Each result includes a bounding box (ROI Box) enclosing the crosspoint and the center coordinates (u_i, v_i), as shown in **Figure 4**.

2.1.3 High-Fidelity Refinement

Finally, the 2D semantic results are back-projected onto the raw depth map to locate 3D regions of interest (ROI) around each crosspoint. The raw map is used instead of the filtered point cloud from **Section 2.1.1** to avoid information loss at RANSAC segmentation boundaries. Within each ROI, the depth at the 2D detection point serves as a baseline. A threshold of twice the rebar diameter is applied to filter out background points and noise, extracting the rebar crosspoint point cloud. Each crosspoint yields a point cluster C_i, forming a group of clusters $S_{clusters} = \{C_1, C_2, \dots, C_n\}$. To abstract each crosspoint into a concise, stable mathematical representation suitable for subsequent algorithms, the Geometric Median is used to condense each cluster C_i into a 3D virtual keypoint P_i, as shown in Eq. (1). This method demonstrates superior robustness against residual noise within the cluster. The process results in a set of high-quality 3D keypoints $S_{keypoints} = \{P_1, P_2, \dots, P_n\}$. This provides the data foundation (including $S_{clusters}$ and $S_{keypoints}$) for global pose estimation in the next stage. The complete process is illustrated in **Figure 5**.

$$P_i = \arg\min_{y \in R^3} \sum_{c_j \in C_i} \left\| y - c_j \right\|_2 \quad (1)$$

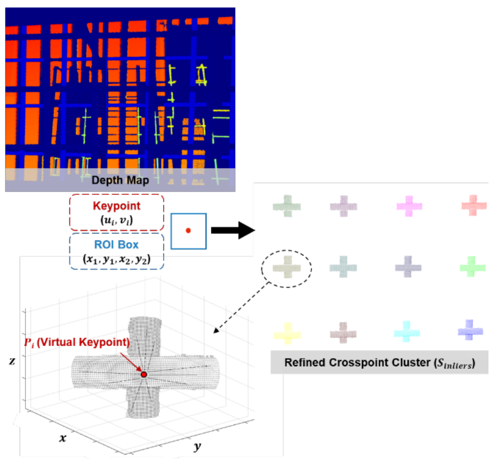


Figure 5. Illustration of the high-fidelity keypoint refinement

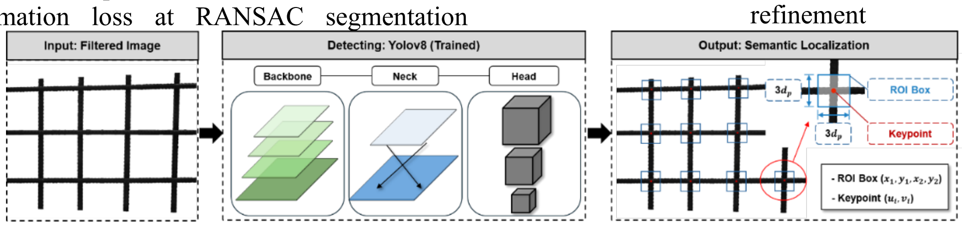


Figure 4. 2D crosspoint detections on the filtered image

2.2 Structural Congruence RANSAC (SC-RANSAC): Global Optimal Interpretation

Traditional model fitting methods (e.g., RANSAC) maximize the number of inliers as the sole optimization objective. When resolving Perceptual Ambiguity inherent in structured repetitive scenes, this approach is insufficient and risky for sparse, repetitive, and noisy data. Algorithms easily converge to physically unreasonable or merely locally optimal models. Therefore, structural priors of the target are integrated into the model selection process, proposing the Structural Congruence RANSAC (SC-RANSAC).

SC-RANSAC establishes an evaluation paradigm for structural perception. It does not rely solely on finding the mathematical solution that best fits the model equation but constructs a multi-dimensional physical constraint system. In the context of rebar cage perception, this is materialized into three indicators: (1) **Data Fidelity (Coplanarity):** Verifies whether the solution represents the majority at the data level; (2) **Geometric Regularity (Orthogonality):** Introduces the geometric regularity prior of the target. It verifies structural orthogonality by reconstructing the local topological connectivity between keypoints, ensuring the physical rationality of the solution; (3) **Global Completeness (Coverage):** Quantifies the global representativeness of the solution using the convex hull area to prevent the final solution from falling into local optima.

This mechanism aims to enable the robot to robustly infer a globally optimal spatial pose from imperfect and partial sensor data.

2.2.1 Multi-Hypothesis Generation

A candidate plane pool is established by iteratively sampling from the 3D discrete semantic keypoint set $S_{keypoints}$. In each iteration k, three non-collinear points are randomly selected from $S_{keypoints}$ to calculate the candidate plane model M_k:

$$M_k \triangleq (\boldsymbol{n_k}, d_k), \boldsymbol{n_k} \cdot \boldsymbol{x} + d_k = 0, \|\boldsymbol{n_k}\| = 1 \qquad (2)$$

Where n_k and $x \in \mathbb{R}^3, d_k \in \mathbb{R}$.

Subsequently, the orthogonal distance from the remaining keypoints in $S_{keypoints}$ to M_k is calculated. A keypoint P_i is classified as a consistent inlier if its perpendicular distance to the plane falls within $\delta = 10$ mm (the value of δ considers sensor precision and rebar diameter). The support inliers set $S_{inliers}$ is defined as:

$$S_k^{inliers} = \{\boldsymbol{P_i} \in S_{keypoints} \,|\, |\boldsymbol{n_k} \cdot \boldsymbol{P_i} + d_k| \leq \delta\} \qquad (3)$$

After exhaustively evaluating all C(n, 3) three-point combinations, a collection of tuples $\{(M_k, S_k^{inliers})\}_{k=1}^{N}$ is output as data for subsequent structural evaluation.

2.2.2 The Structural Congruence Score

(1) Coplanarity (S_{cop}): Data fidelity

As shown in **Figure 6a**, the coplanarity score is defined as the ratio of the number of consistent inliers $S_k^{inliers}$ that support plane M_k to the total number of detected semantic keypoints $S_{keypoints}$:

$$S_{cop\_k} = \frac{|S_k^{inliers}|}{|S_{keypoints}|} \qquad (4)$$

The higher the S_{cop}, the more M_k aligns with the perceived semantic features, indicating greater data fidelity.

(2) Orthogonality (S_{ort}): Geometric regularity

To eliminate perspective distortion, the set of support inliers $S_k^{inliers}$ is first projected onto the 2D local coordinate system defined by the candidate plane M_k. Subsequently, K-Nearest Neighbors (K-NN, K=4) [21] is employed to reconstruct the local topological connectivity of the points. As shown in **Figure 6b,** for each point, the nearest neighbors are identified, and the unit vectors of their connecting lines are computed. By analyzing the angular distribution of these local vectors, the two dominant principal directions θ_1 and θ_2 are extracted. The orthogonality score S_{ort} is then derived as:

$$S_{ort\_k} = \sin(|\theta_1 - \theta_2|) \qquad (5)$$

The higher the S_{ort}, the more M_k fits the orthogonal geometry of the rebar structure, indicating greater geometric regularity.

(3) Coverage (S_{cov}): Global completeness

To assess the global integrity, the area of the convex hull (A_k) of $S_k^{inliers}$ projected onto plane M_k is calculated. As shown in **Figure 6c**, although the global fit (green area) and the local fit (red area) share the same number of supporting points (12), the global fit possesses a significantly larger convex hull area, making it more representative of the complete spatial position and orientation of the rebar structure. To prevent the score from being skewed by extreme anomalies, S_{cov} is normalized using the median area (A_{median}) of all candidate planes in the current iteration:

$$S_{cov\_k} = \frac{A_k}{A_k + A_{median}} \qquad (6)$$

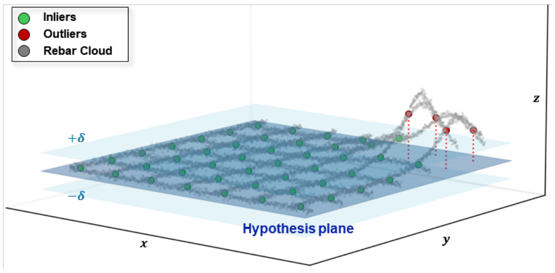
The higher the S_{cov}, the more M_k captures the global physical bounds of the target, indicating greater global completeness.

Candidates with A_k below a safety threshold (10% of the camera's theoretical FOV area at distance d_k) are pruned.

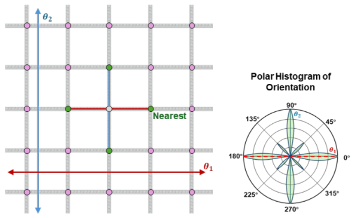
Finally, the structural congruence score for M_k is computed by integrating the three metrics. This formulation ensures that a high score is only achieved if the plane is supported by both strong data evidence (S_{cop}) and sound structural priors (S_{ort} and S_{cov}). The structural congruence score T is derived as:

$$T_k = S_{cop\_k} \times (\omega_1 S_{ort\_k} + \omega_2 S_{cov\_k}) \qquad (7)$$

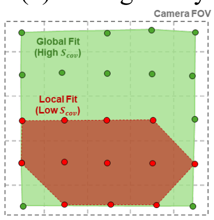
where ω_1 and ω_2 are weighting factors, empirically set to 0.5 to balance the contributions of geometric regularity and global completeness.



(a) Coplanarity



(b) Orthogonality



(c) Coverage

FIGURE 6. Illustration of the structural congruence scoring

2.2.3 Model Selection and Refinement

The candidate plane with the maximum structural congruence score is selected as the optimal interpretation. A final least-squares plane fit is then performed on the entire set of support inliers $S_{best}^{inliers}$ of this winning model to refine its initial three-point-based estimate, yielding the final plane pose.

3 IMPLEMENTATION DETAILS AND EXPERIMENTS

3.1 Implementation Details

In this study, all experiments were conducted on a self-developed rebar tying robot. The hardware architecture comprises a Ranger four-wheel differential drive chassis as the mobile platform. For perception, an Azure Kinect DK is mounted at the lateral midpoint, facing the rebar cage to facilitate the docking process. All onboard computations, including deep learning and SC-RANSAC inference, are executed on an Apex AD10 equipped with NVIDIA Jetson AGX Orin (32GB), as shown in **Figure 7**.

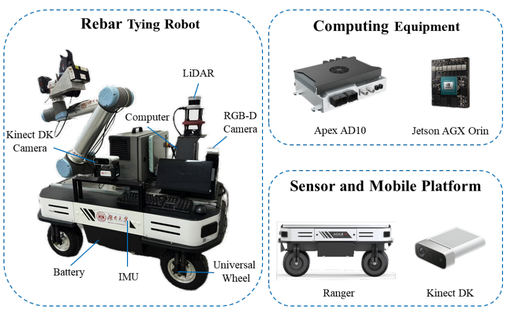


FIGURE 7. Experimental platform and configuration

3.2 Dataset and Scene Construction

This study constructed a physical dataset designed to encompass the stochastic nature of construction scenes.

The dataset was organized into 3 subsets to evaluate different robustness aspects:

1. Random Viewpoint Dataset (300 frames): Collected from a rebar cage (8mm vertical, 10mm horizontal diameter) with mixed grid spacings (10×10, 20×20 cm). The robot captured data at random poses with working distances of 0.490-1.234 m and incident angles of $-24.6°$ to $+16.9°$, simulating navigation docking errors;
2. Diameter Sensitivity Dataset (30 frames): To test the physical detection limit, data was collected on grid meshes (10×10 cm) constructed with three distinct rebar diameters: 6, 8, and 10 mm.
3. Occlusion Dataset (30 frames): Using the same cage as in Dataset 1. Foreground occlusion (10%, 30% and 50%) was introduced by positioning a white board between the camera and the rebar cage to simulate interference from personnel or materials.

Acquiring ground truth is challenging due to assembly imperfections in rebar cages. These deviations prevent the surface from forming an ideal plane. To address this, a distributed sampling strategy was employed. ChArUco boards were attached to the rebar at spatially distributed positions. The 6-DoF poses were computed using OpenCV algorithms. Finally, the board thickness was compensated for to ensure depth accuracy.

In total, 360 valid data frames were collected, spanning all combinations of the aforementioned rebar specifications and environmental variations. Each sample comprises synchronized RGB-D images, robot kinematic states, and the corresponding ground truth poses derived from ChArUco markers.

4 EXPERIMENTS AND RESULTS ANALYSIS

4.1 Evaluation Metrics

Performance is quantified using orientation error, translation error, and runtime cost. These metrics were selected because they directly correspond to the critical control variables for the docking servo loop: heading angle and depth, as well as the system's requirement for real-time execution. Performance is quantified using orientation error (angular deviation of estimated vs. ground-truth plane normal), translation error (depth-axis distance in mm), and runtime cost (mean execution time per iteration).

4.2 Robustness to Viewpoint Variations

This experiment evaluated the performance of the SC-RANSAC system, Basic RANSAC, and modern RANSAC variants: MLESAC [21], MAGSAC++ [22], LO-RANSAC [23], and NAPSAC [24] across the Random Viewpoint Dataset. **Table 1** summarizes the overall results.

SC-RANSAC achieved a 100% success rate across all 300 frames. In contrast, the best-performing baseline, NAPSAC, succeeded in only 51.7% of experiments, while MLESAC achieved only 36.3%. All baseline methods exhibited a consistent failure pattern: at distances exceeding 0.8 m, the increasing sparsity of rebar point clouds caused the algorithms to fit denser background structures, typically the rear wall or the ground plane**.** This failure is fundamentally semantic rather than algorithmic: these methods lack the mechanism to distinguish the target plane from geometrically valid background planes.

Table 1. Comparison with RANSAC variants on the Random Viewpoint Dataset (300 frames)

Method	Success Rate	Mean Angle Error (°)	Mean Distance Error (mm)	Mean Runtime (s)
SC-RANSAC	100%	0.71 ± 0.38	9.2 ± 2.9	2.55
Basic RANSAC	48.3%	1.03±0.56	10.3±4.1	6.09
MLESAC	36.3%	0.84 ± 0.55	8.4 ± 2.8	13.22
MAGSAC++	42.3%	0.98 ± 0.60	8.0 ± 3.4	25.70
LO-RANSAC	49.0%	1.00 ± 0.52	10.0 ± 3.6	1.53
NAPSAC	51.7%	1.17 ± 0.69	9.9 ± 4.9	0.82

On the subset where each baseline succeeded, SC-RANSAC still achieved lower orientation errors (0.58°–0.65° vs. 0.84°–1.17°), confirming that semantic keypoints yield tighter plane fits than dense point clouds even when the correct plane is found. Translation accuracy was comparable across methods on successful subsets. SC-RANSAC maintained orientation errors below 1.3° for 95% of the frames over the full 300-frame dataset, with translation errors below 15 mm. **Figure 9** presents the per-frame comparison across all methods, with the failure map showing the distance-dependent degradation of baseline methods.

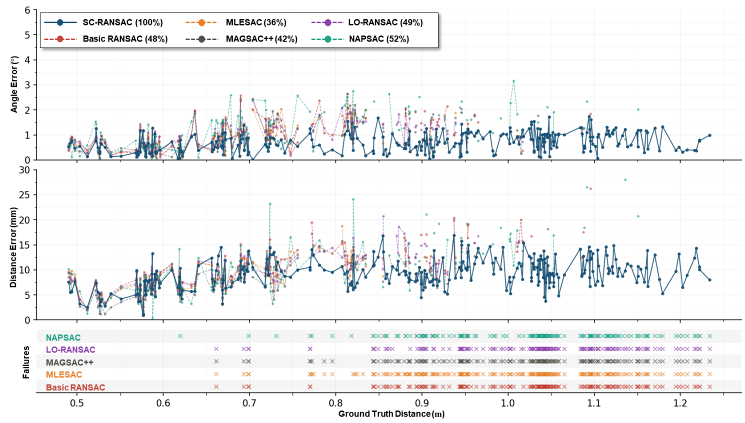


FIGURE 9. Evaluation results of SC-RANSAC and modern RANSAC variants

4.3 Robustness to Rebar Diameter

To evaluate the system's adaptability to varying structural dimensions, rebar cages with diameters of 6, 8, and 10 mm were tested. The working distance was maintained at approximately 0.5 m to ensure all methods had sufficient point density to potentially yield valid results.

As presented in **Table 2**, five modern RANSAC variants failed completely with 6 mm rebars, where the rebar dimension approaches the sensor's noise threshold and produces insufficient point cloud density for reliable plane fitting. With 8 mm and 10 mm rebars at close range, most baselines achieved high success rates, as the shorter working distance increases point cloud density on the rebar surface and reduces the dominance of background structures. SC-RANSAC successfully estimated poses across all diameters, despite a slight increase in error margins as the diameter decreased. These findings demonstrate that the proposed hierarchical filtering and semantic abstraction modules extract structural keypoints, confirming the efficacy of structural priors in perceiving thin, information-sparse elements.

Table 2. Quantitative results with different rebar diameters

Method	6mm			8mm			10mm		
	Success Rate	Distance Error (mm)	Angle Error (°)	Success Rate	Distance Error (mm)	Angle Error (°)	Success Rate	Distance Error (mm)	Angle Error (°)
SC-RANSAC	100%	15.3	1.3	100%	9.8	0.7	100%	9.3	0.4
Basic RANSAC	0%	-	-	93.3%	11.4*	1.3*	96.7%	9.2*	0.9*
MLESAC	0%	-	-	93.3%	8.8*	0.6*	100%	9.6	1.1
MAGSAC++	0%	-	-	100%	11.3	1.0	100%	8.5	0.6
LO-RANSAC	0%	-	-	96.7%	10.4*	0.7*	100%	10.2	0.9
NAPSAC	0%	-	-	100%	10.1	0.6	100%	9.1	0.4

*Precision evaluated on each method's successful subset only.

4.4 Efficacy in Occluded Scenarios

This experiment evaluated SC-RANSAC and modern RANSAC variants under foreground occlusion levels of 10%, 30%, and 50%. As presented in **Table 3**, all baseline methods failed in all scenarios, consistently fitting the foreground whiteboard. This failure is attributed to the continuous planar geometry and high infrared reflectance of the obstruction, which generated a denser point cloud that dominated the inlier count relative to the sparse rebar structure. Conversely, SC-RANSAC maintained a 100% success rate across all conditions. While the occlusion case showed a increase in error due to the reduction of valid data, the method correctly anchored the pose estimation to the top-layer rebar plane by leveraging semantic constraints.

Table 3. Quantitative results with occluded environments

Obstruction Level	Method	Success Rate	Distance Error	Angle Error
10%	SC	100%	8.1 mm	0.7°
	All modern RANSAC variants	0%	Failure	Failure
30%	SC	100%	13.5 mm	1.0°
	All modern RANSAC variants	0%	Failure	Failure
50%	SC	100%	13.3 mm	1.4°
	All modern RANSAC variants	0%	Failure	Failure

5 CONCLUSION

This research addresses the challenge of autonomous docking for rebar tying robots by establishing a robust perception framework capable of operating in unstructured construction sites. By leveraging the inherent geometric regularity of rebar cages, the proposed methodology integrates deep learning-based semantic abstraction with the SC-RANSAC algorithm. This approach effectively resolves perceptual ambiguity and sensor noise that hinder traditional depth-based methods.

Physical experiments confirm that the proposed framework maintains high accuracy and stability across varying working distances, diverse rebar specifications, and occluded conditions. Comparative evaluation against five modern RANSAC variants (MLESAC, MAGSAC++, LO-RANSAC, NAPSAC, and Basic RANSAC) reveals that the performance gap is fundamentally semantic: all baseline methods achieve comparable precision when they identify the correct plane, but fail in 48–64% of experiments by selecting background structures. SC-RANSAC eliminates this failure mode entirely through its structural congruence scoring. These findings indicate that integrating structural priors into the perception loop is a viable and necessary strategy for reliable pose estimation in scenes containing multiple geometrically valid planes.

However, the current specific metrics for geometric regularity are optimized for common orthogonal planar cages. Future work involves generalizing these specific indicators to curved or irregular configurations.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFC3806800) and the National Natural Science Foundation of China (Grant No. 52308312). DeepSeek

was used to improve the language and readability of this manuscript.

References

[1] S.B.S. Ahmad, M.U. Mazhar, A. Bruland, B.S. Andersen, J.A. Langlo, O. Torp. Labour productivity statistics: a reality check for the Norwegian construction industry. *International Journal of Construction Management,* 20(1): 39–52, 2020.

[2] M. Abdel-Wahab and B. Vogl. Trends of productivity growth in the construction industry across Europe, US and Japan. *Construction Management and Economics*, 29(6):635–644, 2011.

[3] T. Bock. The future of construction automation: Technological disruption and the upcoming ubiquity of robotics. *Automation in Construction*, 59:113–121, 2015.

[4] R. Bogue. The role of robots in the construction industry. *Industrial Robot: The International Journal of Robotics Research and Application*, 52(2):157–164, 2025.

[5] C.A. Issa and A. Nasr. An experimental study of welded splices of reinforcing bars. *Building and Environment*, 41(10):1394–1405, 2006.

[6] X. Zhao and C.C. Cheah. BIM-based indoor mobile robot initialization for construction automation using object detection. *Automation in Construction,* 146:104647, 2023.

[7] W. Umer, H. Li, G.P.Y. Szeto and A.Y.L. Wong. Low-cost ergonomic intervention for mitigating physical and subjective discomfort during manual rebar tying. *Journal of Construction Engineering and Management*, 143(10):04017075, 2017.

[8] J. Meng, S. Wang, G. Li, L. Jiang, X. Zhang, C. Liu and Y. Xie. Iterative-learning error compensation for autonomous parking of mobile manipulator in harsh industrial environment. *Robotics and Computer-Integrated Manufacturing,* 68:102077, 2021.

[9] Z. Chen, Y. Feng, T. Li and Y. Jiang. Object-based terminal positioning solution within task-boosted global constraint for improving mobile robotic stacking accuracy. *Advanced Engineering Informatics,* 60:102521, 2024.

[10] L. He, B. Li and G. Chen. Noise model-based line segmentation for plane extraction in sparse 3-d lidar data. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024.

[11] E. Kaartinen, K. Dunphy and A. Sadhu. Lidar-based structural health monitoring: applications in civil infrastructure systems. *Sensors*, 22(12):4610, 2022.

[12] Z. Liao, X. Zhang, T. Zhang, Z. Li, Z. Zheng, Z. Wen and Y. Li. A real-time degeneracy sensing and compensation method for enhanced LiDAR SLAM. *IEEE Transactions on Intelligent Transportation Systems*, 26(3):4202–4213, 2025.

[13] M. Liu, J. Guo, L. Deng, S. Wang and H. Wang. Enhanced vision-based 6-DoF pose estimation for robotic rebar tying. *Automation in Construction*, 171:105999, 2025.

[14] J. Li, W. Gao, Y. Wu, Y. Liu and Y. Shen. High-quality indoor scene 3D reconstruction with RGB-D cameras: A brief review. *Computational Visual Media*, 8(3):369-393, 2022.

[15] D. DeTone, T. Malisiewicz and A. Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 337–349, Salt Lake City, USA, 2018.

[16] P.E. Sarlin, D. DeTone, T. Malisiewicz and A. Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pages 4937–4946, Seattle, USA, 2020.

[17] M.-K. Kim, J.P.P. Thedja and Q. Wang. Automated dimensional quality assessment for formwork and rebar of reinforced concrete components using 3D point cloud data. *Automation in Construction*, 112:103077, 2020.

[18] H. Sarbolandi, D. Lefloch, A. Kolb, Kinect range sensing: Structured-light versus Time-of-Flight Kinect, *Computer Vision and Image Understanding*, 139: 1-20, 2015.

[19] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[20] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[21] P.H.S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1): 138–156, 2000.

[22] D. Baráth, J. Noskova, M. Ivashechkin and J. Matas. MAGSAC++, a fast, reliable and accurate robust estimator. *In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1301–1309, Seattle, USA, 2020.

[23] O. Chum, J. Matas and J. Kittler. Locally optimized RANSAC. *In Proceedings of the 25th DAGM Symposium on Pattern Recognition*, pages 236–243, Magdeburg, Germany, 2003.

[24] D.R. Myatt, P.H.S. Torr, S.J. Nasuto, J.M. Bishop and R. Craddock. NAPSAC: High noise, high dimensional robust estimation – it's in the bag. *In Proceedings of the 13th British Machine Vision Conference (BMVC)*, pages 45.1–45.10, Cardiff, UK, 2002.