

An Active Pose Estimation Framework for Robotic Assembly of Large-Scale Aircraft Components

Yuzhe Wang, Guijun Ma, *Member, IEEE*, Zidong Wang, *Fellow, IEEE*, Xiaoyou Duan, Long Yu, Weibo Liu, *Member, IEEE*, Yong Zhang, and Han Ding, *Senior Member, IEEE*

Abstract—Accurate pose estimation is essential for robotic docking assembly of large-scale aircraft components. Conventional passive approaches, such as exhaustive robotic scanning or reliance on external metrology systems, involve extensive data acquisition and processing, resulting in high computational cost, complex system deployment, and limited adaptability to dynamic assembly environments. This article presents a closed-loop active pose estimation method that integrates next-best-view (NBV) planning and hierarchical point cloud registration within a perception-action loop for robotic docking assembly. In the NBV planning stage, a geometric information gain-driven NBV strategy is formulated by jointly considering geometric stiffness, overlap consistency, and feature saliency, enabling the robot to autonomously select informative viewpoints for efficient data acquisition. In the point cloud registration stage, a sparse-to-dense hierarchical registration network is designed to achieve accurate alignment with reduced computational burden. The pose estimation process iterates in a closed-loop manner until predefined criteria on both alignment error and inlier fitness are satisfied. Experiments conducted on both simulated and real-world aircraft assembly platforms showcase the effectiveness of the proposed method on efficiency and accuracy compared to fixed-path and coverage-based methods. Ultimately, the proposed method achieves assembly success rates of 95% in simulation and 90% on the real-world platform, while maintaining reliable docking performance with a real-world gap error of 0.6 ± 0.2 mm. These results demonstrate the capability of the proposed method to enable high-precision and efficient robotic docking assembly in aircraft manufacturing.

Index Terms—robotic docking assembly, large-scale aircraft components, active pose estimation, point cloud registration, next-best-viewpoint

I. INTRODUCTION

AMONG numerous tasks involved in large-scale aircraft manufacturing, docking assembly is a critical step during final integration. The resulting alignment accuracy directly

Yuzhe Wang, Guijun Ma, Xiaoyou Duan, and Ye Yuan are with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China. Guijun Ma and Ye Yuan are also with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Wuhan 430074, China. (e-mail: yuzhe_wang@hust.edu.cn; mgj@hust.edu.cn; xiaoyouduan@hust.edu.cn; yye@hust.edu.cn).

Zidong Wang and Weibo Liu are with the Department of Computer Science, Brunel University of London, Uxbridge, Middlesex, UB8 3PH, United Kingdom (e-mail: Zidong.Wang@brunel.ac.uk; Weibo.Liu2@brunel.ac.uk).

Long Yu is with AVIC Chengdu Aircraft Company Limited, Chengdu, China (e-mail: 908452469@qq.com).

Yong Zhang is with the School of Artificial Intelligence and Automation, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: zhangyong77@wust.edu).

Han Ding is with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, and also with the School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: dinghan@hust.edu.cn).

affects the aerodynamic profile, flight performance, and overall structural safety of the aircraft [1]. Conventional docking assembly systems rely on rigid gantry structures, which is difficult to adapt across aircraft variants. In contrast, industrial robots have attracted increasing attention for docking assembly owing to their superior flexibility and operational efficiency [2]. In robotic aircraft docking assembly, variant adaptability can be achieved through end-effector replacement, while the high degrees of freedom of robotic manipulators enable flexible pose adjustment to accommodate diverse geometric configurations.

In fact, robotic aircraft docking assembly relies on accurate pose estimation of large-scale aircraft components. A commonly used approach estimates pose by measuring artificial targets attached to the fuselage (e.g., spherically mounted retroreflectors and reflective fiducial markers) [3]–[5]. Nevertheless, this approach is sensitive to both the placement and quantity of artificial targets, where suboptimal placement or an insufficient number of artificial targets may fail to adequately capture the complex geometry of large-scale aircraft structures. With recent advances in 3-D sensing technologies, dense point clouds can be acquired more conveniently and efficiently. Compared with target-based measurements, point clouds provide more comprehensive geometric information, offering a promising alternative sensing modality for pose estimation in large-scale aircraft docking assembly.

Existing strategies for point cloud acquisition can be broadly classified into passive and active measurement methods. Passive methods deploy external sensing devices, such as LiDAR or depth cameras, fixed around the workspace to capture global point clouds of the fuselage. To achieve sufficient pose estimation accuracy, robotic docking systems typically require more than three sensing units. Moreover, the layout of these devices is largely dependent on operator experience and must be adjusted to mitigate occlusions, leading to repeated calibration procedures [6], [7]. In contrast, active measurement methods mount sensing equipment on the robot end-effector and utilize manipulator motion to acquire multi-view point clouds of large-scale aircraft components [8], [9]. This configuration eliminates the need for multiple external sensors and requires only a one-time hand-eye calibration. Additionally, active measurement methods enable adaptive scanning by optimizing the observation path online to emphasize key regions, which improves point cloud quality and pose estimation accuracy.

Severing as a promising paradigm for active perception, next-best-viewpoint (NBV) planning has been widely adopted in active measurement methods [10]–[13]. During the NBV planning process, the robot evaluates candidate viewpoints

by simulating virtual observations at each pose and selects the best viewpoint that maximizes a predefined information gain criterion. Most existing NBV approaches are developed for reconstruction tasks, such as large-scale architectural modeling, where information gain is typically formulated to maximize the surface coverage of previously unseen regions [14]. Nevertheless, in aircraft docking assembly, maximizing surface coverage can drive the robot to scan geometrically flat regions with limited structural features, which provide a weak contribution to pose estimation [15]. Prior studies indicate that regions with high curvature or diverse normal distributions contain richer geometric information and impose stronger geometric constraints on pose alignment [16]–[19]. Motivated by these insights, a feasible strategy is to formulate a new information gain metric that integrates geometric indicators, enabling the NBV planner to actively prioritize regions rich in distinct geometric features.

It should be noted that acquiring geometrically informative point clouds is only a prerequisite; the subsequent challenge lies in efficiently processing large-scale point cloud data to achieve high-precision pose estimation. Point cloud registration methods can generally be classified into optimization-based and learning-based approaches. Optimization-based methods, such as the iterative closest point (ICP) algorithm [20], estimate relative pose by iteratively minimizing spatial discrepancies between source and target point clouds. Although these methods are effective under accurate initialization, they are sensitive to initial pose and susceptible to local minima. Learning-based approaches can be further divided into global feature regression and local feature matching methods. Global regression methods formulate registration as a direct pose prediction problem by encoding entire point clouds into global descriptors [21]. Nevertheless, such methods lack interpretability and discard local geometric details required for high-precision registration.

In comparison, local feature matching methods estimate relative pose by establishing point-to-point correspondences between source and target point clouds based on learned local descriptors [22]. To alleviate the computational complexity of correspondence search in large-scale point clouds, these methods typically perform matching on downsampled representations. Although computationally efficient, such sparse matching may not provide sufficient accuracy to satisfy the stringent tolerance requirements of robotic aircraft docking assembly. On the other hand, directly performing dense correspondence matching on full-resolution point clouds can improve alignment precision but incurs substantial computational cost, limiting real-time performance [23], [24]. Therefore, a hierarchical sparse-to-dense registration framework that progressively refines alignment provides a promising solution to balance computational efficiency and the precision demands of robotic docking assembly.

In this article, a novel active pose estimation framework is proposed for efficient and accurate pose estimation of large-scale aircraft components in robotic docking assembly. The framework integrates NBV planning and hierarchical point cloud registration within a closed-loop perception-action architecture. A new geometric information gain metric is formulated for NBV planning by incorporating multiple ge-

ometric indicators, enabling the robot to actively prioritize feature-rich regions that provide strong constraints for pose estimation. To efficiently process the acquired point clouds, a hierarchical sparse-to-dense registration network is developed to progressively refine alignment while maintaining computational efficiency. By integrating active data acquisition with efficient data processing, the proposed approach is conducive to a precise robotic assembly process. Extensive experiments on both simulated and real-world robotic assembly platforms validate the effectiveness of the proposed framework, demonstrating significant improvements in pose estimation accuracy and operational efficiency.

The main contributions of this paper are summarized as follows:

- An active pose estimation framework is established for the robotic docking assembly of large-scale aircraft components, which integrates NBV planning strategy and hierarchical registration network into a closed-loop system to achieve active data acquisition and reliable pose estimation.
- A new information gain integrating geometric indicators is proposed to actively prioritize feature-rich regions for high-precision pose estimation.
- A sparse-to-dense registration network is developed to balance computational efficiency and accuracy.
- Comprehensive validation on both simulated and real-world robotic assembly platforms, demonstrating significant performance improvements over baseline methods.

The remaining sections of this article are organized as follows. The background of point cloud registration, NBV and the problem description are presented in Section II. Section III introduces the proposed geometric information gain-driven NBV planning strategy, the hierarchical point cloud registration network, closed-loop active pose estimation network and their evaluation metrics. The self-developed robotic assembly experimental platform and the experimental dataset are described in Section IV. In Section V, experimental results and analysis of point cloud registration and robotic assembly are provided and discussed in detail. Finally, the conclusions are drawn, and the possible future topics are addressed in Section VI.

II. PRELIMINARY

A. Background of Point Cloud Registration

Given a source point cloud \mathcal{P} and a target point cloud \mathcal{Q} , the point cloud registration process is generally divided into two stages: correspondence estimation and rigid transformation estimation. In the first stage, a set of N correspondences $\mathcal{C} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$ is estimated using a correspondence estimation method, where each pair is associated with a confidence weight w_i . In the second stage, based on these predicted correspondences, the optimal rigid transformation $\mathbf{T} = \{\mathbf{R}, \mathbf{t}\}$ (where $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$) is derived to align \mathcal{P} to \mathcal{Q} . The parameters are determined by minimizing the following weighted alignment error:

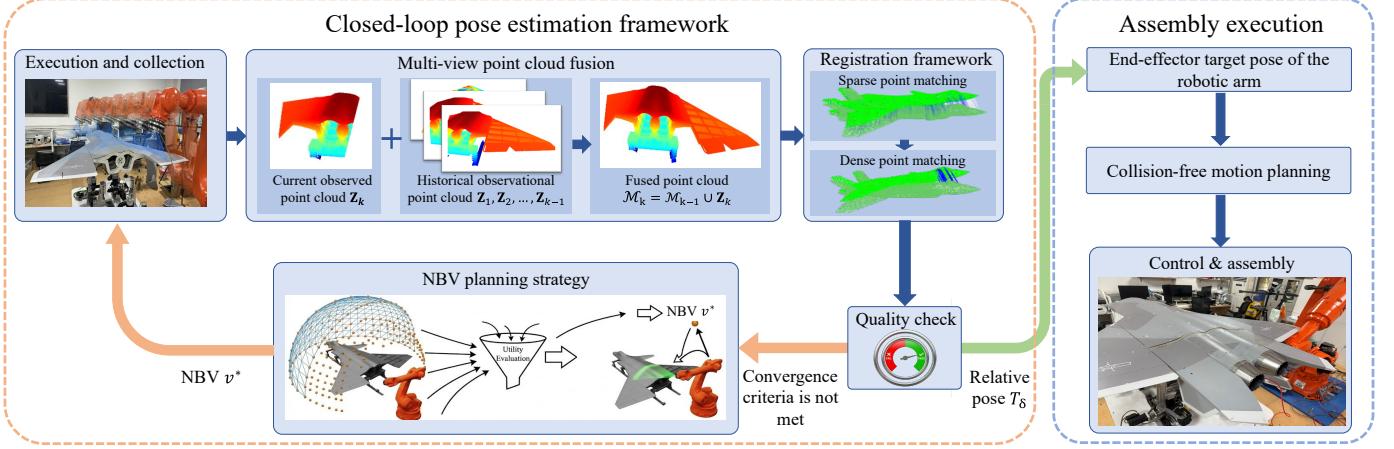


Fig. 1. The schematic overview of the proposed closed-loop pose estimation framework and the robotic assembly execution. The workflow consists of two coordinated phases. (Left) In the closed-loop pose estimation phase, scanning, multi-view fusion, hierarchical registration, and NBV planning are performed iteratively to update the 6D pose estimate until a predefined stopping criterion is met. (Right) In the assembly execution phase, the estimated pose is passed to the motion planning and control modules for component alignment and docking.

$$\min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^N w_i \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2. \quad (1)$$

Generally, to efficiently solve the optimization problem defined in (1), the singular value decomposition (SVD) solver is typically employed. This approach provides a closed-form solution to the orthogonal Procrustes problem [25] by decoupling the estimation of rotation and translation. Specifically, the optimal translation \mathbf{t} is determined by aligning the weighted centroids defined as $\bar{\mathbf{p}} = \sum w_i \mathbf{p}_i / \sum w_i$ and $\bar{\mathbf{q}} = \sum w_i \mathbf{q}_i / \sum w_i$. With the translation removed, minimizing the alignment error is mathematically equivalent to maximizing the trace of the correlation between the centered point sets [26]. This correlation is captured by the weighted cross-covariance matrix \mathbf{H} . By performing SVD on \mathbf{H} , the optimal rotation \mathbf{R} is derived to align the singular vectors of the two point sets

$$\mathbf{H} = \sum_{i=1}^N w_i (\mathbf{p}_i - \bar{\mathbf{p}})(\mathbf{q}_i - \bar{\mathbf{q}})^T, \quad (2)$$

$$\mathbf{R} = \mathbf{V} \mathbf{D} \mathbf{U}^T, \quad (3)$$

$$\mathbf{t} = \bar{\mathbf{q}} - \mathbf{R} \bar{\mathbf{p}}, \quad (4)$$

where $\mathbf{H} = \mathbf{U} \Sigma \mathbf{V}^T$ admits the SVD, and $\mathbf{D} = \text{diag}(1, 1, \det(\mathbf{V} \mathbf{U}^T))$ ensures a strictly proper rotation [27].

B. Background of NBV Planning

Let $\mathcal{V}_f = \{v^1, v^2, \dots, v^n\}$ denote a set of feasible candidate viewpoints sampled on a spherical surface, which strictly satisfy task-specific constraints, including kinematic reachability, collision avoidance, and visibility. In NBV perception, given the sensor pose $\mathbf{T}_k \in SE(3)$ at the k -th time step and the accumulated point cloud map \mathcal{M}_k fused from historical observations $\mathcal{Z}_{1:k} = \{Z_1, Z_2, \dots, Z_k\}$, the objective is to select the NBV $v_{k+1}^* \in \mathcal{V}_f$ such that the predicted observation $\hat{Z}_{k+1}(v_{k+1}^*)$ maximizes the information gain for point cloud registration. This optimization problem is formulated as

$$v_{k+1}^* = \arg \max_{v \in \mathcal{V}_f} G(v, \mathcal{M}_k, \mathcal{M}_{\text{prior}}), \quad (5)$$

where $\mathcal{M}_{\text{prior}}$ denotes the prior geometric model and $G(\cdot)$ is the information gain function designed to quantify the observation information.

C. Problem Description

In robotic docking assembly, pose estimation aims to determine the desired target pose $\mathbf{T}_{\text{target}}$ of the robot end-effector with respect to the robot base frame. Due to manufacturing tolerances and placement deviations, the real-world fuselage exhibits an unknown spatial transformation $\mathbf{T}_\delta \in SE(3)$ relative to its ideal pose. As a result, the actual target pose $\mathbf{T}_{\text{target}}$ must be computed by incorporating this deviation into the kinematic chain, which can be expressed as follows:

$$\mathbf{T}_{\text{target}} = \mathbf{T}_{\text{calib}} \cdot \mathbf{T}_\delta \cdot \mathbf{T}_{\text{calib}}^{-1} \cdot \mathbf{T}_{\text{nom}}, \quad (6)$$

where \mathbf{T}_{nom} denotes the nominal assembly pose calibrated under an ideal fuselage pose, and $\mathbf{T}_{\text{calib}}$ represents the hand-eye calibration matrix that describes the transformation from the camera frame to the end-effector frame [28].

To obtain $\mathbf{T}_{\text{target}}$, the problem is reduced to estimating the deviation transformation \mathbf{T}_δ , which can be formulated as a point cloud registration problem between the fused multi-view measured point cloud \mathcal{M} and the measured fuselage point cloud at ideal pose $\mathcal{M}_{\text{ideal}}$. The optimal \mathbf{T}_δ^* is defined as the rigid transformation that minimizes the weighted geometric alignment error:

$$\mathbf{T}_\delta^* = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^N w_i \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2, \quad (7)$$

where $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$ represent the rotational and translational components of \mathbf{T}_δ , respectively. $\mathcal{C} = \{(\mathbf{p}_i, \mathbf{q}_i)\}_{i=1}^N$ denotes the set of point correspondences predicted by a neural network, associating $\mathbf{p}_i \in \mathcal{M}$ with $\mathbf{q}_i \in \mathcal{M}_{\text{ideal}}$, and w_i denotes the confidence weight of each correspondence. The optimization problem in (7) can be efficiently solved in a closed form using the SVD solver introduced in Section A. Nevertheless, its accuracy fundamentally depends

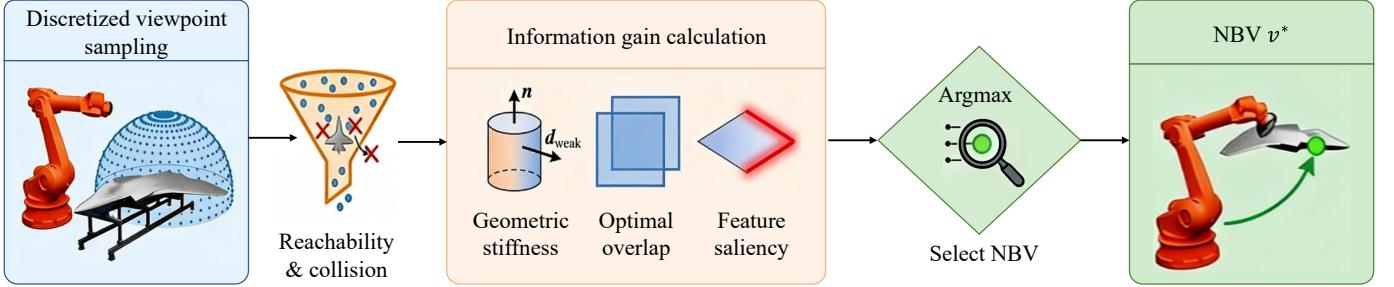


Fig. 2. Workflow of the active viewpoint planning strategy. Candidate viewpoints are first filtered by reachability and collision constraints. Subsequently, feasible viewpoints are evaluated based on a composite utility of geometric stiffness, overlap consistency, and feature saliency. The NBV v^* is optimized to maximize information gain.

on the reliability and geometric distribution of the established correspondences.

To enhance the quality of correspondences and geometric observability, an active perception strategy is introduced to acquire informative point clouds. Specifically, a NBV planning scheme selects optimal robotic viewpoints by maximizing a geometric information gain metric G_{geo} . The NBV decision process is formulated as

$$v_{k+1}^* = \arg \max_{v \in \mathcal{V}_f} G_{\text{geo}}(v, \mathcal{M}_k, \mathcal{M}_{\text{aircraft}}), \quad (8)$$

where v is a candidate viewpoint selected from the feasible set \mathcal{V}_f , and v_{k+1}^* denotes the optimal viewpoint for the $k + 1$ -th scan. The term \mathcal{M}_k represents the fused fuselage point cloud accumulated over the previous k scanning iterations, and $\mathcal{M}_{\text{aircraft}}$ denotes the prior CAD model of the aircraft component.

Finally, a closed-loop perception–action protocol is introduced to iteratively refine pose estimation by map updating and viewpoint re-planning. The iterative pose estimation terminates once the registration error falls within the prescribed assembly tolerance, providing an accurate estimation of \mathbf{T}_δ and thus $\mathbf{T}_{\text{target}}$.

III. METHODS

In this section, the proposed active pose estimation framework is introduced in detail. First, the geometric information gain-driven NBV planning strategy is presented. Next, the architecture of the hierarchical registration network is described. Furthermore, the closed-loop active pose estimation framework is detailed, followed by the definition of evaluation metrics. The schematic overview of the proposed framework is shown in Fig. 1.

A. Geometric Information Gain-Driven NBV Planning Strategy

To actively acquire sufficient point cloud for point cloud registration, the proposed strategy identifies the NBV v^* by maximizing a registration-oriented information gain defined as

$$v^* = \arg \max_{v \in \mathcal{V}_f} (w_s G_s(v) + w_o G_o(v) + w_f G_f(v)), \quad (9)$$

where $G_s(v)$, $G_o(v)$, and $G_f(v)$ denote the geometric stiffness gain, overlap consistency gain, and feature saliency gain, respectively; w_s , w_o , and w_f are the corresponding weights used to balance these gains.

1) *Geometric Stiffness Gain G_s :* Drawing upon the stability analysis of point cloud registration [16], reliable pose estimation necessitates regions with diverse normal orientations to fully constrain the pose. To quantify the strength of this constraint, the approximate Hessian matrix is utilized, defined as the sum of outer products of the Jacobians. The magnitudes of the eigenvalues quantify the geometric constraint strength along the corresponding directions, as determined by the distribution of surface normals. Following this criterion, the geometric stiffness gain is defined as the minimum eigenvalue of the approximate Hessian:

$$G_s(v) = \lambda_{\min} \left(\sum_{\mathbf{p} \in \mathcal{P}_{\text{vir}}} \mathbf{J}_{\mathbf{p}}^T \mathbf{J}_{\mathbf{p}} \right), \quad (10)$$

where \mathcal{P}_{vir} represents the virtual point cloud rendered from the prior model at viewpoint v , $\mathbf{J}_{\mathbf{p}} = [\mathbf{n}^T, (\mathbf{p} \times \mathbf{n})^T]$ is the point-to-plane Jacobian at \mathbf{p} , and \mathbf{n} is the unit surface normal. Maximizing $\lambda_{\min}(\cdot)$ therefore drives the robot toward viewpoints that enrich the diversity of surface normal orientations in the currently observed point cloud, thereby improving the accuracy of pose estimation.

2) *Overlap Consistency Gain G_o :* To balance efficient exploration and coverage continuity, the overlap consistency gain keeps the overlap ratio near τ_{ref} by penalizing deviations using the following Gaussian weight:

$$G_o(v) = \exp \left(-\frac{(\rho(v) - \tau_{\text{ref}})^2}{2\sigma_o^2} \right), \quad (11)$$

where $\rho(v)$ represents the raw overlap ratio calculated between the virtual point cloud \mathcal{P}_{vir} and the current map \mathcal{M}_k [29], and σ_o is a tuning parameter controlling the tolerance to overlap deviation.

3) *Feature Saliency Gain G_f :* To focus on geometric details that significantly constrain the pose, the feature saliency gain aggregates the following local curvature:

$$G_f(v) = \frac{1}{|\mathcal{P}_{\text{vir}}|} \sum_{\mathbf{p} \in \mathcal{P}_{\text{vir}}} \kappa(\mathbf{p}), \quad (12)$$

where $\kappa(\mathbf{p})$ denotes the local curvature at point \mathbf{p} , guiding the sensor toward salient geometric features [30].

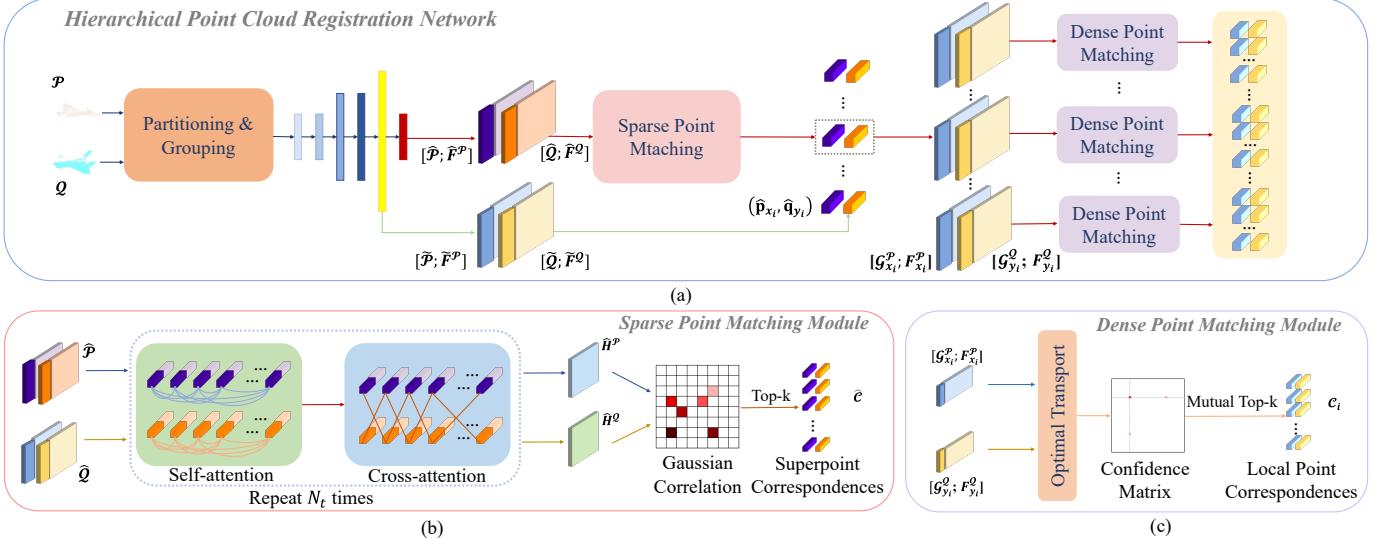


Fig. 3. Overview of the hierarchical registration framework. (a) The complete hierarchical registration pipeline, illustrating the flow from feature extraction to sparse-to-dense matching. (b) The sparse point matching module, utilizing attention mechanisms and Gaussian correlation to establish reliable superpoint correspondences. (c) The dense point matching module, which performs dense matching within local patches using optimal transport.

B. Hierarchical Point Cloud Registration Network

A sparse-to-dense hierarchical registration network is presented for pose estimation. As shown in Fig. 3, a dual-resolution representation is first constructed and encoded to obtain sparse- and dense-level features. Sparse correspondences are then inferred through attention-based feature interaction, followed by patch-wise dense matching. The final pose $\{\mathbf{R}, \mathbf{t}\}$ is estimated from the aggregated correspondences via a weighted SVD layer.

1) Sparse-Dense Point Partitioning and Feature Extraction: Using farthest point sampling method [31], the source point cloud \mathcal{P} is downsampled into sparse points $\tilde{\mathcal{P}}$ and dense points $\tilde{\mathcal{P}}$. To pass sparse matches to the dense level for patch-wise refinement, each dense point is assigned to its nearest sparse based on Euclidean distance, forming local patches where each sparse point is associated with a specific group of dense points. Subsequently, a shared PointNet [32] extracts dense features $\tilde{\mathbf{F}}^{\mathcal{P}}$ via point-wise processing and aggregates sparse features $\tilde{\mathbf{F}}^{\mathcal{P}}$ via max-pooling within each group. The sparse features are fed into the sparse matching module for sparse correspondence, while dense features are retained for patch-wise dense refinement. The same procedure is applied to the target point cloud \mathcal{Q} .

2) Sparse Point Matching Module: As shown in Fig. 3(b), to make features more distinct in low-texture regions by adding global context, the initial sparse features are processed by an attention-based interaction module. To gather context step by step, the features are updated at the l -th layer through the following residual attention operation:

$$\mathbf{H}^{(l)} = \mathbf{H}^{(l-1)} + \text{Softmax} \left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V}, \quad (13)$$

where $\mathbf{H}^{(0)}$ represents the initial features, and \mathbf{Q} , \mathbf{K} , and \mathbf{V} are linear projections derived for either intra-cloud self-attention or inter-cloud cross-attention.

After obtaining the context-aware hybrid representations $\hat{\mathbf{H}}^{\mathcal{P}}$ and $\hat{\mathbf{H}}^{\mathcal{Q}}$, a Gaussian kernel is adopted to measure the feature-space similarity:

$$s_{i,j} = \exp \left(-\frac{1}{2\sigma^2} \left\| \hat{\mathbf{h}}_i^{\mathcal{P}} - \hat{\mathbf{h}}_j^{\mathcal{Q}} \right\|_2^2 \right), \quad (14)$$

where σ is a learnable scale parameter. To mitigate ambiguous matches and encourage mutual agreement, the raw scores are normalized into a confidence matrix via dual-normalization:

$$\bar{s}_{i,j} = \frac{s_{i,j}}{\sum_{k=1}^{|\tilde{\mathcal{Q}}|} s_{i,k}} \cdot \frac{s_{i,j}}{\sum_{k=1}^{|\tilde{\mathcal{P}}|} s_{k,j}}, \quad (15)$$

where $\bar{s}_{i,j}$ represents the normalized similarity score between the i -th source feature $\hat{\mathbf{h}}_i^{\mathcal{P}}$ and the j -th target feature $\hat{\mathbf{h}}_j^{\mathcal{Q}}$, $|\tilde{\mathcal{P}}|$ and $|\tilde{\mathcal{Q}}|$ denote the number of sparse points in the source and target clouds, respectively. Finally, the sparse correspondence set $\hat{\mathcal{C}} = \{(\hat{\mathbf{p}}_{x_i}, \hat{\mathbf{q}}_{y_i})\}_{i=1}^k$ is determined by selecting the k pairs with the highest confidence scores.

3) Dense Point Matching Module: As illustrated in Fig. 3(c), for each extracted sparse correspondence $(\hat{\mathbf{p}}_{x_i}, \hat{\mathbf{q}}_{y_i}) \in \hat{\mathcal{C}}$, the associated local patches are retrieved to solve a fine-grained matching problem. First, a similarity matrix \mathbf{M}_i is computed via the scaled dot product as follows:

$$\mathbf{M}_i = \left(\tilde{\mathbf{F}}_{x_i}^{\mathcal{P}} (\tilde{\mathbf{F}}_{y_i}^{\mathcal{Q}})^T \right) / \sqrt{\tilde{d}}, \quad (16)$$

where $\tilde{\mathbf{F}}_{x_i}^{\mathcal{P}}$ and $\tilde{\mathbf{F}}_{y_i}^{\mathcal{Q}}$ denote the dense feature matrices in the local patches centered at $\hat{\mathbf{p}}_{x_i}$ and $\hat{\mathbf{q}}_{y_i}$, respectively; \tilde{d} is the feature dimension. Then, the Sinkhorn–Knopp algorithm [33] is used to compute the soft assignment matrix \mathbf{Z}_i . To establish rigorous local correspondences, the dense match set \mathcal{C}_i is formed by selecting the k pairs with the highest confidence scores from \mathbf{Z}_i . Finally, all local matches are aggregated as $\mathcal{C} = \bigcup \mathcal{C}_i$ to estimate the rigid transformation $\{\mathbf{R}, \mathbf{t}\}$ via a weighted SVD layer.

4) *Loss Function*: To supervise correspondence prediction and pose estimation, a negative log-likelihood loss [34] is adopted for discrete matching, together with pose regression and geometric consistency terms. Given N_g randomly sampled ground-truth sparse correspondences, the local matching loss is designed to promote high confidence for valid matches while suppressing outliers. For the i -th patch pair, let m_i and n_i denote the number of features in the source and target patches, respectively. The sets of unmatched points in the source and target are represented by \mathcal{I}_i and \mathcal{J}_i . The matching loss is formulated as

$$\mathcal{L}_{m,i} = - \sum_{(x,y) \in \mathcal{M}_i} \log z_{x,y}^i - \sum_{x \in \mathcal{I}_i} \log z_{x,m_i+1}^i - \sum_{y \in \mathcal{J}_i} \log z_{n_i+1,y}^i, \quad (17)$$

where z^i denotes the predicted confidence matrix. The first term maximizes the log-likelihood of the correct matches in the set \mathcal{M}_i . The second and third terms enforce the unmatched points in \mathcal{I}_i and \mathcal{J}_i , indexed at $m_i + 1$ and $n_i + 1$, respectively, thereby handling non-overlapping regions effectively.

The overall objective is formulated as the following weighted sum to jointly supervise feature learning and pose regression:

$$\mathcal{L}_{\text{total}} = \lambda_m \mathcal{L}_m + \lambda_p \mathcal{L}_p + \lambda_g \mathcal{L}_g, \quad (18)$$

where $\mathcal{L}_m = \frac{1}{N_g} \sum \mathcal{L}_{m,i}$ is the averaged matching loss; \mathcal{L}_p penalizes deviations on the $SE(3)$ manifold, defined as $\|\mathbf{R}_{\text{est}}^T \mathbf{R}_{\text{gt}} - \mathbf{I}\|_F^2 + \|\mathbf{t}_{\text{est}} - \mathbf{t}_{\text{gt}}\|_2^2$, where $\{\mathbf{R}_{\text{est}}, \mathbf{t}_{\text{est}}\}$ and $\{\mathbf{R}_{\text{gt}}, \mathbf{t}_{\text{gt}}\}$ denote the estimated and ground-truth rotation and translation, respectively. Additionally, \mathcal{L}_g denotes a Chamfer-like geometric consistency loss [35] that measures spatial alignment error. The weights λ_m , λ_p , and λ_g are employed to balance these loss terms.

C. Closed-Loop Active Pose Estimation Framework

As shown in Fig. 1, a closed-loop active pose estimation framework is proposed which integrates the hierarchical registration network with the NBV planner. The feedback mechanism allows the robot to continuously refine the target pose estimation by accumulating measurements until the registration confidence meets strict manufacturing requirements.

An online assessment criterion is introduced for the closed-loop pose estimation. After each registration step, an inlier set \mathcal{K} is extracted from the accumulated source point cloud \mathcal{P} . A point $\mathbf{p} \in \mathcal{P}$ is classified as an inlier if its Euclidean distance to the nearest point in the target \mathcal{Q} falls within a preset threshold δ after transformation. i.e., $\min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{R}_{\text{est}} \mathbf{p} + \mathbf{t}_{\text{est}} - \mathbf{q}\|_2 < \delta$.

Based on the identified inliers, the alignment accuracy is quantified by the root mean square error (RMSE)

$$\mathcal{E}_{\text{rmse}} = \sqrt{\frac{1}{|\mathcal{K}|} \sum_{\mathbf{p} \in \mathcal{K}} \min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{R}_{\text{est}} \mathbf{p} + \mathbf{t}_{\text{est}} - \mathbf{q}\|_2^2}. \quad (19)$$

The geometric overlap ratio is further defined as $S_{\text{fit}} = \frac{|\mathcal{K}|}{|\mathcal{P}|}$, which measures the proportion of well-aligned source points. To guarantee assembly precision, the registration loop terminates only when both geometric accuracy and overlap consistency simultaneously satisfy

$$\mathcal{E}_{\text{rmse}} < \tau_{\text{rmse}} \quad \text{and} \quad S_{\text{fit}} > \tau_{\text{fit}}, \quad (20)$$

Algorithm 1 Closed-Loop Active Pose Estimation Method

Input: Target point cloud \mathcal{Q} ; initial viewpoint v_1 ; termination thresholds τ_{rmse} and τ_{fit}
Output: Estimated pose $\mathbf{T}_{\text{est}} = \{\mathbf{R}_{\text{est}}, \mathbf{t}_{\text{est}}\}$

- 1: Initialize the accumulated point cloud map $\mathcal{M}_0 = \emptyset$ and iteration index $k = 1$.
- 2: Initialize the loop state Loop = True.
- 3: **while** Loop **do**
- 4: Execute robot motion to viewpoint v_k and capture a new point cloud observation Z_k .
- 5: Update the accumulated point cloud map by fusing the new observation: $\mathcal{M}_k = \mathcal{M}_{k-1} \cup Z_k$.
- 6: Estimate the 6D pose $\mathbf{T}_k = \{\mathbf{R}_k, \mathbf{t}_k\}$ by registering the accumulated map \mathcal{M}_k to the target \mathcal{Q} via the hierarchical registration network.
- 7: Extract the inlier set \mathcal{K} from \mathcal{M}_k , and compute the alignment RMSE $\mathcal{E}_{\text{rmse}}$ via (19), as well as the overlap ratio.
- 8: **if** the termination criteria in (20) are satisfied **then**
- 9: Update the final pose estimation $\mathbf{T}_{\text{est}} = \mathbf{T}_k$.
- 10: Loop = False.
- 11: **else**
- 12: Plan the NBV v_{k+1}^* by maximizing the information gain via (9).
- 13: Update the next viewpoint $v_{k+1} = v_{k+1}^*$ and $k = k + 1$.
- 14: **end if**
- 15: **end while**
- 16: **return** \mathbf{T}_{est}

where τ_{rmse} and τ_{fit} denote predefined tolerance thresholds.

If either condition is not met, the NBV planner is triggered via (8) to acquire the next optimal view and update the accumulated map. The overall execution process of the closed-loop active pose estimation method is summarized in Algorithm 1.

D. Evaluation Metrics

To comprehensively assess the performance of the proposed framework, three categories of metrics are considered: point cloud registration performance and assembly performance.

1) *Evaluation Metrics for Registration*: Following the standard protocols adopted in recent benchmarks [36], [37], the hierarchical registration network is evaluated using three quantitative metrics, namely inlier ratio (IR), feature matching recall (FMR), and registration recall (RR). IR is defined as the fraction of putative correspondences whose residuals are below a strict threshold (e.g., 2 mm) under the ground-truth transformation, and it is used to characterize the precision of correspondence generation. FMR is defined as the fraction of point cloud pairs whose inlier ratio exceeds a minimum confidence threshold (e.g., 5%), which characterizes the likelihood of recovering a valid pose hypothesis. RR is treated as a system-level metric that quantifies the probability of successful pose recovery; a registration is regarded as successful only if the pose error, including both rotational and translational deviations, lies within the acceptance tolerances.

2) *Evaluation Metrics for Assembly*: To assess the effectiveness of the closed-loop planning strategy, metrics associated with robustness and time-cost efficiency are employed, including success rate (SR) and average steps (AS). SR is defined as the percentage of experimental trials in which convergence to the target assembly tolerance is achieved within a permissible observation budget, thereby reflecting the capability to cope with initial uncertainties. For successfully converged trials, AS is computed as the mean number of viewpoints required to satisfy the termination criteria, and it is used as a proxy for sensing efficiency in the assembly process.

IV. EXPERIMENTAL PLATFORM AND DATASET

To comprehensively validate the effectiveness and reliability of the proposed active pose estimation framework, an experimental platform scaled down from the real robotic docking assembly scenario was constructed. Based on this platform, an aircraft dataset tailored for registration tasks in robotic docking assembly industrial scenarios was established.

A. Experimental Platform Setup

The experimental platform consists of two functional sub-platforms: the active pose estimation platform and the docking assembly platform. As shown in Fig. 4, the active pose estimation platform is built upon a KUKA KR210 industrial robot, which offers a large workspace and high positioning repeatability suitable for assembling large-scale aircraft components. An Intel RealSense D455 depth camera is attached to the robot's last joint via a custom-designed bracket. To replicate realistic misalignment scenarios, the fuselage is supported by four distributed 3-DoF numerical control locators. Each unit allows independent actuation along the x , y , and z axes, enabling the precise generation of any pose deviations relative to the ideal pose for validation purposes.

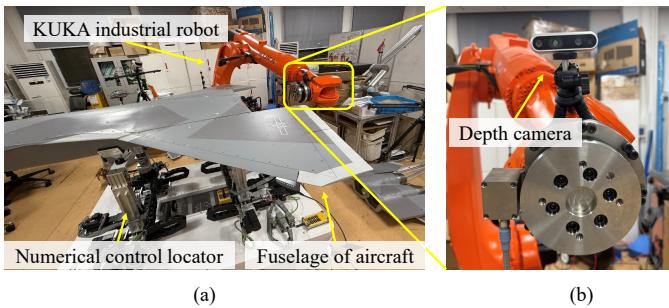


Fig. 4. Experimental platform for the active pose estimation: (a) the overall workspace containing the KUKA robot, the aircraft fuselage, and the numerical control locators; (b) the eye-in-hand depth camera configuration used for point cloud acquisition.

The docking assembly platform, shown in Fig. 5, is designed to evaluate the assembly accuracy of large-scale aircraft components using the pose estimated by hierarchical point-cloud registration. The robot end-effector integrates a custom clamp and a connecting flange, where the clamp securely holds the aircraft tail and enables the robotic arm to dock it with the fuselage.

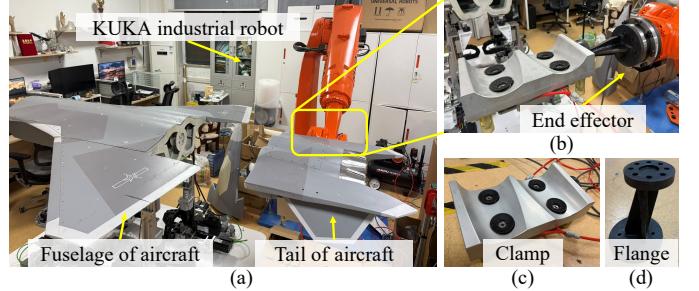


Fig. 5. Experimental platform for the docking assembly: (a) the docking scenario between the tail and fuselage sections; (b) the custom-designed end-effector; (c) the clamp; and (d) the connecting flange.

B. Data Generation

To facilitate the training of the hierarchical registration network, a synthetic dataset was constructed derived from real-world scans [38]. Raw point clouds of the aircraft fuselage collected via the experimental platform served as the source data. The corresponding target point clouds were synthesized by applying random rigid body transformations, which served as the ground-truth pose labels. The transformation parameters were sampled from $[0.2, 0.6]$ m for translation and $[5^\circ, 25^\circ]$ for rotation. To enhance model robustness and prevent trivial correspondences, two data augmentation strategies were implemented: 1) independent random downsampling was applied to the source and target clouds; and 2) random cropping was performed to maintain an overlap ratio between 40% and 80%. The final dataset comprises 8,000 pairs, divided into training, validation, and testing sets in a 7:1:2 ratio.

V. EXPERIMENTS AND ANALYSIS

This section presents the experimental evaluation of the proposed hierarchical point cloud registration network and the active pose estimation framework in robotic aircraft assembly. In the first stage, the registration performance is benchmarked against representative optimization-based and learning-based methods. In the second stage, the closed-loop performance of the active pose estimation is assessed within a simulation environment. Finally, real-world validation experiments conducted on a KUKA KR210 platform are illustrated to demonstrate the system's feasibility.

A. Evaluation of Point Cloud Registration Framework

1) *Experimental Setting*: To evaluate the performance of the proposed hierarchical registration network, it was benchmarked against four representative methods: the optimization-based iterative closest point (ICP) [20] and fast global registration (FGR) [39], and the learning-based maximal cliques (MAC) [40] and reliable inlier evaluation network (RIENet) [41]. All learning-based models were trained on the constructed aircraft dataset using an Intel Core i9-13900K CPU and a NVIDIA RTX 4090 GPU. The loss weights were set to $\lambda_p = \lambda_g = 1$ and $\lambda_m = 0.5$.

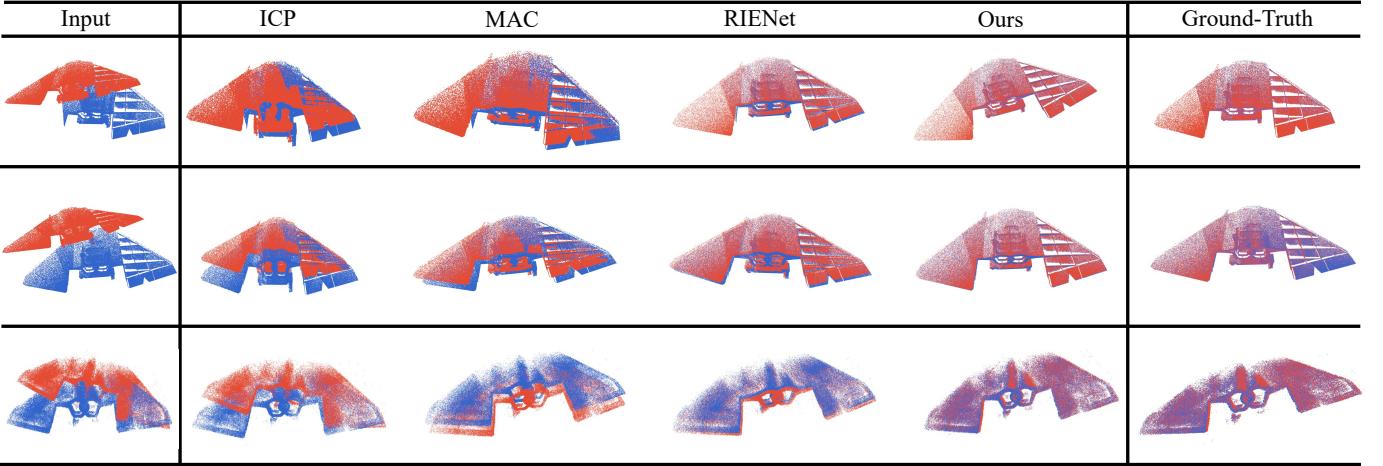


Fig. 6. Visualizations of registration results on the aircraft dataset. In the Input column, red and blue point clouds represent the initial source and target scans. In the registration results and ground-truth columns, red denotes the transformed source point clouds aligned to the target (blue). Compared to the optimization-based methods (ICP [20] and FGR [39]) and the learning-based methods (MAC [40] and RIENet [41]), our method achieves better registration results.

2) *Prediction Results:* Quantitative comparisons are reported in Table I. The highest performance across all metrics is achieved by the proposed method. Specifically, our method attains an IR of 71.9%, outperforming the second-best method (RIENet) by 7.6% and the ICP by over 23%. Regarding FMR, the proposed method records a score of 98.0%, while the learning-based baselines MAC and RIENet achieve 96.6% and 96.2%, and the optimization-based methods ranging from 95.0% to 95.6%. Finally, for the RR, the proposed method achieves 93.2%. This represents a notable improvement over RIENet (90.8%) and MAC (88.6%), while the optimization-based ICP and FGR yield lower recall rates of 78.4% and 85.1%. Qualitative registration results are shown in Fig. 6, where the proposed method achieves better performance than the baselines.

TABLE I
PERFORMANCE COMPARISON OF IR, FMR AND RR ACROSS DIFFERENT POINT CLOUD REGISTRATION METHODS ON THE AIRCRAFT DATASET

Method	IR (%)	FMR (%)	RR (%)
ICP	48.6	95.0	78.4
FGR	52.2	95.6	85.1
MAC	57.8	96.6	88.6
RIENet	64.3	96.2	90.8
Ours	71.9	98.0	93.2

An ablation study was further conducted to validate the effectiveness of the attention modules, as detailed in Table II. The baseline model, devoid of any attention mechanisms, yields the lowest performance with an IR of 52.8%, FMR of 84.2%, and RR of 80.7%. Incorporating either self-attention or cross-attention independently results in moderate improvements. Specifically, the self-attention module increases the IR to 59.5% and RR to 85.9%, while the cross-attention module achieves an IR of 60.2% and RR of 86.7%. The most significant performance gain is observed when both modules are employed simultaneously. The full configuration reaches an IR of 71.9%,

FMR of 98.0%, and RR of 93.2%, indicating that the attention modules are beneficial for high-precision registration.

TABLE II
PERFORMANCE COMPARISON OF IR, FMR AND RR ACROSS DIFFERENT ATTENTION MODULE CONFIGURATIONS ON THE AIRCRAFT DATASET

Components	Metrics				
	Self-Att.	Cross-Att.	IR (%)	FMR (%)	RR (%)
–	–	–	52.8	84.2	80.7
✓	–	–	59.5	90.8	85.9
–	✓	–	60.2	90.3	86.7
✓	✓	✓	71.9	98.0	93.2

B. Simulation Validation of Robotic Aircraft Assembly

1) *Experimental Settings:* A high-fidelity simulation environment was constructed in NVIDIA Isaac Sim by integrating the aircraft fuselage model, the KUKA KR210 robotic arm, and the RGB-D camera. To validate the proposed active pose estimation framework, the system operates in a continuous perception-decision-execution cycle. Specifically, following each perception update, the NBV planning strategy determines the next optimal viewpoint. The corresponding goal is converted into a collision-free trajectory using the IKFast solver, enabling the robot to move to the viewpoint and perform the next scan. Upon convergence of the pose estimation, the target end-effector pose is derived through kinematic chain transformations, and a collision-free trajectory to this pose is then planned using the TRAC-IK solver [42] for the final precision docking assembly. For each experimental trial, the fuselage was placed with a random initial pose deviation within a range of ± 5 cm in translation and $\pm 15^\circ$ in rotation relative to the ideal pose.

In the simulation, the proposed active pose estimation framework was compared with a fixed-path method and a coverage-driven NBV method [43]. For the fixed-path method, a static sequence of viewpoints was pre-defined based on

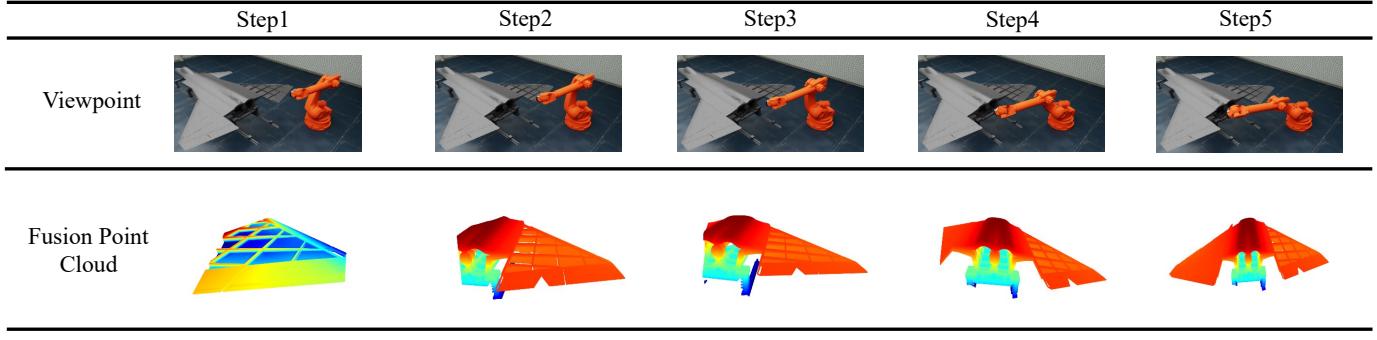


Fig. 7. Visualization of the active pose estimation process in simulation. The sequence from Step 1 to Step 5 illustrates the robot motion toward planned viewpoints (top row) and the corresponding incremental point-cloud fusion (bottom row). Within five iterations, the fused point cloud becomes geometrically sufficient for high-precision registration, consistent with the rapid convergence reported in Fig. 9.

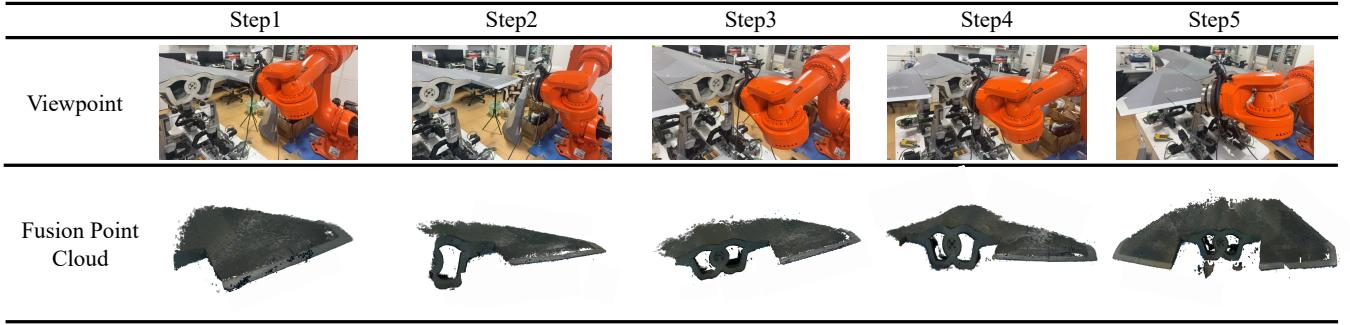


Fig. 8. Visualization of the real-world active pose estimation. The robot adaptively plans viewpoints (top) to capture point cloud, updating the fused point cloud (bottom) to achieve high-precision registration within five steps.

experience, and the same trajectory was executed in all trials without feedback from the current observation. For the coverage-driven NBV method, the NBV was selected to maximize the surface coverage of previously unobserved regions.

To analyze the efficiency of each method, two experiments were designed based on 20 independent trials. The fixed step-setting restricts the observation budget to a maximum of 5 steps to evaluate the SR under strict resource constraints. Meanwhile, the adaptive termination setting executes the strategy until the termination criteria are met, utilizing the AS to quantify the convergence efficiency required to guarantee solution quality.

2) Assembly Results: The comparative results of the closed-loop system performance are summarized in Table III. Under the fixed step experiment, the fixed-path strategy fails frequently (SR 70%), as pre-defined trajectories cannot adapt to large random initial deviations. The coverage-driven strategy performs moderately (SR 80%) but often wastes steps on non-informative regions. In contrast, a dominant success rate of 95% (19/20) is achieved by the proposed method, demonstrating that key geometric constraints can be locked onto by the stiffness-aware planner significantly faster than the baselines. Regarding the adaptive termination experiment, high time costs are incurred by the baselines to guarantee convergence. The fixed-path strategy requires an average of 9 steps to identify the correct constraints, while the coverage-driven strategy needs 7 steps to accumulate sufficient map coverage. Superior efficiency is demonstrated by the proposed method, requiring only 5 steps

on average to reach the same high-precision standard.

To further investigate the convergence behavior, the RMSE defined in (19) was recorded after each step to monitor the step-wise uncertainty reduction. Fig. 9 illustrates the convergence trends. The fixed-path strategy (grey line) starts with a relatively low error but quickly plateaus because the static path lacks adaptivity. The error is reduced gradually by the coverage-driven strategy (blue line), but approximately 8 steps are required to reach a low error level, as map completeness is prioritized over geometric constraints. In contrast, a sharp error reduction is achieved by the registration-oriented strategy (red line). By explicitly optimizing for geometric stiffness, observations that tightly constrain the degrees of freedom are actively sought, reducing the alignment residual to a convergent state within only 5 steps.

TABLE III
PERFORMANCE COMPARISON OF SR UNDER THE FIXED STEP EXPERIMENT AND AS UNDER THE ADAPTIVE TERMINATION EXPERIMENT ACROSS DIFFERENT STRATEGIES ON A SIMULATION ASSEMBLY PLATFORM

Method	Fixed Step	Adaptive Termination
Fixed-Path	70.0%	9
Coverage-Driven	80.0%	7
Ours	95.0%	5

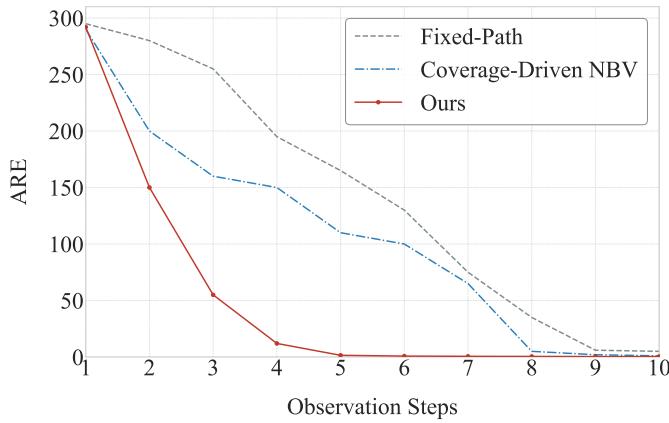


Fig. 9. Comparison of alignment residual convergence for different viewpoint planning strategies. Rapid convergence is achieved by our strategy (red), significantly outperforming coverage-driven (blue) and fixed-path (grey) strategies.

C. Real-World Validation of Robotic Aircraft Assembly

1) *Experimental Settings:* To evaluate the feasibility of the proposed framework in the real-world domain, experiments were conducted on the KUKA KR210 platform described in Section IV. Consistent with the simulation, the system operated under the same closed-loop perception-decision-execution architecture. Specifically, the analytical IKFast solver was utilized to efficiently compute collision-free joint configurations for the planned viewpoints. To replicate realistic assembly misalignment, the initial pose of the fuselage component was randomized before each trial. The four distributed numerical control locators supporting the fuselage were actuated independently along the X, Y, and Z axes to generate random translational deviations ranging from 1 cm to 5 cm relative to the nominal pose. Ten independent trials were conducted for each of the three strategies (fixed-path, coverage-driven, and Ours) to ensure statistical reliability. The performance was assessed using the same metrics (SR and AS) and experiments (fixed step and adaptive termination) as established in the simulation validation.

2) *Assembly Results:* The quantitative results from the real-world trials are summarized in Table IV. Consistent with the simulation trends, the fixed-path strategy achieved a SR of only 60%, primarily failing in cases where large initial deviations caused the pre-defined trajectory to miss critical geometric features. The coverage-driven strategy improved the SR to 80% by maximizing surface coverage but incurred a higher time cost, requiring an average of 8 steps to accumulate sufficient data. In contrast, the proposed strategy achieved the highest SR of 90% with a significantly reduced average of 5 steps. As visualized in Fig. 8, the system actively directed the sensor toward geometrically salient regions, verifying that the proposed framework effectively transfers to the real-world environment to rapidly reduce pose uncertainty.

3) *Assembly Quality Verification:* Beyond perception efficiency, the final docking quality was validated to ensure compliance with manufacturing tolerances. The real-world gap between the mating interfaces was measured using a feeler

gauge. As indicated in Table IV, the fixed-path strategy resulted in a significant misalignment with an average gap of 1.8 ± 1.2 mm, while the coverage-driven baseline improved the accuracy to 1.1 ± 0.4 mm. In comparison, the proposed method achieved a uniform mating interface with a minimum average gap error of 0.6 ± 0.2 mm, successfully satisfying the sub-millimeter accuracy requirements for large-scale aircraft assembly.

TABLE IV
PERFORMANCE COMPARISON OF SR UNDER THE FIXED STEP EXPERIMENT,
AS UNDER THE ADAPTIVE TERMINATION EXPERIMENT AND FINAL GAP
ACROSS DIFFERENT PERCEPTION STRATEGIES ON A REAL-WORLD
ASSEMBLY PLATFORM

Method	Fixed Step	Adaptive Termination	Final Gap
Fixed-Path	60%	8	1.8 ± 1.2
Coverage-Driven	80%	8	1.1 ± 0.4
Ours	90%	5	0.6 ± 0.2

VI. CONCLUSION

In this article, a closed-loop active pose estimation framework has been developed to achieve high-precision pose estimation for the robotic assembly of large-scale aircraft components. The proposed approach introduces a geometric information gain-driven NBV planning strategy, which integrates geometric stiffness, overlap consistency, and feature saliency to actively capture feature-rich point clouds. On the basis of the actively acquired data, a hierarchical sparse-to-dense registration network has been put forward to accurately estimate the 6D pose and efficiently evaluate the alignment reliability online. The proposed closed-loop framework has been validated on a simulation environment and a real-world robotic assembly platform. Experimental results demonstrate two key findings: (1) the proposed hierarchical registration network outperforms representative optimization-based and learning-based methods in terms of registration recall and inlier ratio; and (2) the geometric information gain-driven NBV strategy enhances the convergence speed and assembly success rate compared to the fixed-path and coverage-driven baselines. Finally, the sub-millimeter assembly accuracy has been successfully verified on our real-world robotic assembly platform.

In the future, we plan to: 1) incorporate multi-modal sensory fusion by combining visual pose estimation with force/torque feedback, thereby facilitating compliant control during the contact-rich docking phase; and 2) investigate a hybrid active perception strategy that synergizes the interpretability of optimization-based methods with the generalization capability of learning-based approaches, further enhancing the robustness of the viewpoint planning against unstructured environmental uncertainties.

REFERENCES

- [1] G. Wang, W. Li, C. Jiang, and H. Ding, Machining allowance calculation for robotic edge milling an aircraft skin considering the deformation of assembly process, *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 3350–3361, 2022.
- [2] H. Jia, S. Bai, and P. Chirarattananon, Aerial manipulation via modular quadrotors with passively foldable airframes, *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 4, pp. 1930–1938, 2023.

- [3] M. Laimer, D. Wertjanz, P. Gsellmann, G. Schitter, and E. Csencsics, High-precision 3-D measurements on moving objects, *IEEE/ASME Transactions on Mechatronics*, vol. 30, no. 2, pp. 1514-1525, 2025.
- [4] J. Lv, Z. Zhang, Y. Liu, and S. Shen, LCE-Calib: Automatic LiDAR-frame/event camera extrinsic calibration with a globally optimal solution, *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 5, pp. 2988-2999, 2023.
- [5] G. Ma, Z. Wang, W. Liu, Z. Yang, D. Huang, and H. Ding, Closed-loop parameter optimization for robotic machining using physics-informed machine learning and multiobjective optimization, *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 22410-22422, 2025.
- [6] W. Zhao, Y. Wang, J. Zhang, and Y. Liu, Parameter identification and scanning pose optimization for robotic vision measurement system, *IEEE/ASME Transactions on Mechatronics*, vol. 30, no. 1, pp. 541-553, 2025.
- [7] T. Hao and D. Xu, Trajectory imitation with visual guidance and compliance control for robotic bolt grasping and assembly, *IEEE/ASME Transactions on Mechatronics*, vol. 30, no. 6, pp. 4745-4753, 2025.
- [8] T. Jiang, H. Cui, and X. Cheng, A calibration strategy for vision-guided robot assembly system of large cabin, *Measurement*, vol. 163, art. no. 107991, 2020.
- [9] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, and K. Harada, Learning force control for contact-rich manipulation tasks with rigid position-controlled robots, *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5709-5716, 2020.
- [10] Z. Jiang, R. Su, Y. Liu, Z. Long, B. Song, and X. Tang, Accurate relative measurement of multitarget poses by monocular vision for nonmodel-based real-time calibration of industrial robot, *Measurement*, vol. 235, art. no. 114979, 2024.
- [11] K. Bekris, K. Hauser, S. Herbert, J. Yu, J. Lv, Y. Feng, C. Zhang, S. Zhao, L. Shao, and C. Lu, SAM-RL: Sensing-aware model-based reinforcement learning via differentiable physics-based simulation and rendering, *The International Journal of Robotics Research*, vol. 44, no. 10-11, pp. 1767-1783, 2024.
- [12] G. Ma, Z. Wang, Z. Yang, R. Chen, W. Liu, Y. Zhang, and S. Yan, A novel pairwise domain-adaptation-assisted dual-task learning approach to coprediction of robotic machining efficiency and quality in new parameter spaces, *IEEE Transactions on Industrial Informatics*, vol. 21, no. 7, pp. 5150-5159, 2025.
- [13] G. Ma, X. Yang, S. Xu, C. Cheng, and X. He, ERMN: An enhanced meta-learning approach for state of health estimation of lithium-ion batteries, *Journal of Energy Storage*, vol. 72, art. no. 108628, 2023.
- [14] D. Yan, J. Liu, F. Quan, H. Chen, and M. Fu, Active implicit object reconstruction using uncertainty-guided next-best-view optimization, *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6395-6402, 2023.
- [15] S. Aravecchia, A. Richard, M. Clausel, and C. Pradalier, Next-best-view selection from observation viewpoint statistics, *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems & Data Mining*, Detroit, USA, Oct. 2023, pp. 10505-10510.
- [16] N. Gelfand, L. Ikeimoto, S. Rusinkiewicz, and M. Levoy, Geometrically stable sampling for the ICP algorithm, in *Fourth International Conference on 3-D Digital Imaging and Modeling & Data Mining*, Banff, Canada, Oct. 2003, pp. 260-267.
- [17] S. Rusinkiewicz and M. Levoy, Efficient variants of the ICP algorithm, in *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, Quebec City, Canada, Aug. 2001, pp. 145-152.
- [18] R. Liu, Z. Li, X. Fan, C. Zhao, H. Huang, and Z. Luo, Learning deformable image registration from optimization: Perspective, modules, bilevel training and beyond, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7688-7704, 2022.
- [19] L. Yan, P. Wei, H. Xie, J. Dai, H. Wu, and M. Huang, A new outlier removal strategy based on reliability of correspondence graph for fast point cloud registration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 7, pp. 7986-8002, 2023.
- [20] P. Besl and N. D. McKay, A method for registration of 3-D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, 1992.
- [21] Y. Di and R. Zhang, GPV-Pose: Category-level object pose estimation via geometry-guided point-wise voting, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, New Orleans, USA, Jun. 2022, pp. 6771-6781.
- [22] F. Gorschlüter, P. Rojberg, and T. Pöllabauer, A survey of 6D object detection based on 3D models for industrial applications, *Journal of Imaging*, vol. 8, no. 3, p. 53, 2022.
- [23] C. Wang, D. Xu, Y. Zhu, R. Martin-Martin, C. Lu, L. Fei-Fei, and S. Savarese, DenseFusion: 6D object pose estimation by iterative dense fusion, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Long Beach, USA, Jun. 2019, pp. 3338-3347.
- [24] S. Peng, Y. Liu, Q. Huang, X. Zhou, and H. Bao, PVNet: Pixel-wise voting network for 6DoF pose estimation, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Long Beach, USA, Jun. 2019, pp. 4556-4565.
- [25] P. H. Schönemann, A generalized solution of the orthogonal Procrustes problem, *Psychometrika*, vol. 31, no. 1, pp. 1-10, 1966.
- [26] K. S. Arun, T. S. Huang, and S. D. Blostein, Least-squares fitting of two 3-D point sets, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698-700, 1987.
- [27] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376-380, 1991.
- [28] X. Qiao, C. Xu, Y. Wang, and G. Ma, SDI: A sparse drift identification approach for force/torque sensor calibration in industrial robots, *Neurocomputing*, vol. 620, art. no. 129292, 2025.
- [29] R. Pito, A solution to the next best view problem for automated surface acquisition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 1016-1030, 1999.
- [30] W. Peng, Y. Wang, Z. Miao, M. Feng, and Y. Tang, Viewpoints planning for active 3-D reconstruction of profiled blades using estimated occupancy probabilities (EOP), *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4109-4119, 2021.
- [31] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, PointNet++: Deep hierarchical feature learning on point sets in a metric space, in *Advances in Neural Information Processing Systems & Data Mining*, Long Beach, USA, Dec. 2017, pp. 5105-5114.
- [32] C. R. Qi and H. Su, PointNet: Deep learning on point sets for 3D classification and segmentation, in *2017 IEEE Conference on Computer Vision and Pattern Recognition & Data Mining*, Honolulu, USA, Jul. 2017, pp. 652-660.
- [33] R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, *Pacific Journal of Mathematics*, vol. 21, no. 2, pp. 343-348, 1967.
- [34] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, SuperGlue: Learning feature matching with graph neural networks, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Seattle, USA, Jun. 2020, pp. 4938-4947.
- [35] Y. Wang and J. M. Solomon, Deep Closest Point: Learning representations for point cloud registration, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Long Beach, USA, Jun. 2019, pp. 3523-3532.
- [36] X. Bai, Z. Luo, L. Zhou, H. Fu, L. Quan, and C.-L. Tai, D3Feat: Joint learning of dense detection and description of 3D local features, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Seattle, USA, Jun. 2020, pp. 6359-6367.
- [37] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler, Predator: Registration of 3D point clouds with low overlap, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Nashville, USA, Jun. 2021, pp. 4267-4276.
- [38] M. Kaspar, J. D. Muñoz Osorio, and J. Bock, Sim2Real transfer for reinforcement learning without dynamics randomization, in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems & Data Mining*, Las Vegas, USA, Oct. 2020, pp. 4383-4388.
- [39] Q.-Y. Zhou, J. Park, and V. Koltun, Fast global registration, in *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, Oct. 2016, pp. 766-782.
- [40] X. Zhang, J. Yang, S. Zhang, and Y. Zhang, 3D registration with maximal cliques, in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition & Data Mining*, Vancouver, Canada, Jun. 2023, pp. 17745-17754.
- [41] Y. Shen, L. Hui, H. Jiang, J. Xie, and J. Yang, Reliable inlier evaluation for unsupervised point cloud registration, in *Proceedings of the European Conference on Computer Vision*, Tel Aviv, Israel, Oct. 2022, pp. 392-409.
- [42] P. Beeson and B. Ames, TRAC-IK: An open-source library for improved solving of generic inverse kinematics, in *2015 IEEE-RAS 15th International Conference on Humanoid Robots*, Seoul, South Korea, Nov. 2015, pp. 928-935.
- [43] Y. Liu, W. Zhao, H. Liu, Y. Wang, and X. Yue, Coverage path planning for robotic quality inspection with control on measurement uncertainty, *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 3482-3493, 2022.