A Low-Cost Robotic Assembly Method for Large Planar Workpieces Using Laser Displacement Sensors and Monocular Vision

Xiaobing Cao[1*], Yicen Xu[2], Yonghong Yao[1], Sheng Chen[3]
[1]School of Control Technology, Wuxi Institute of Technology, Jiangsu Wuxi, 214121, China
[2,]School of Intelligent Equipment and Automotive Engineering, Wuxi Institute of Commerce, Jiangsu Wuxi, 214153, China
[3]School of Communication and Electronic Engineering, East China Normal University, Shanghai, 200062, China
E-mail: xiaobingcao91@gmail.com
[*]Corresponding author

Keywords: side window glass, laser displacement sensor, monocular vision, robotic assembly, pose correction

Received: November 20, 2024

To address the challenges associated with large planar workpieces, such as the side window glass of a train, this paper proposes a practical, intelligent robotic assembly method that utilizes laser displacement sensors (LDSs) and monocular vision. The laser point cloud data is fitted to the equation of the plane of the side window optimally, allowing for the calculation of its plane coefficients and unit normal vector (UNV) using the Lagrange multiplier method. The robot's end flange is adjusted to ensure the camera imaging plane is along with the window plane at a specified distance. Monocular vision is employed to capture the pose features of the upper right corner of the window, facilitating a compensation method for precise assembly. The experimental results show that the robot achieved a positioning accuracy of less than 0.5mm and an orientation accuracy within 1 degree, confirming the effectiveness of the proposed method. In the experiment, the maximum displacement deviation in the X direction is about ± 0.4mm, the maximum offset in the Y and Z directions is about ± 0.3mm, the maximum rotation deviation in the X axis is about 0.6 °, and the minimum rotation deviation in the Y axis is about 0.8 °. All deviations meet the requirements of ± 1 ° attitude correction and ± 0.5mm displacement accuracy, and have good reliability. This method solves the practical problem of assembling large planar workpieces with integrated LDS and monocular vision. This technology involves using laser point cloud data to fit the precise plane equation of the window, determining alignment parameters through Lagrange multiplier method, and adjusting the posture of the end flange of the robot. Monocular vision further assists in extracting positional features, achieving precise real-time posture correction and alignment. This system proposes an effective and low-cost automation that reduces manual intervention and alleviates typical problems of remote processing of large workpieces.

Povzetek: Predstavljen je nizkocenovni robotski način sestavljanja velikih ravnih delov, ki združuje laserske merilnike premika in monokularni vid za natančno prilagajanje položaja in orientacije.

1 Introduction

Automatic assembly technology is extensively utilized in the automotive, marine, aerospace, and other manufacturing sectors to reduce labor costs and enhance production efficiency. The profound integration of machine vision technology with assembly technology represents a significant trend and a critical characteristic of the current evolution in assembly technology. Numerous advancements have been achieved in both theoretical research and practical applications [1], [2], [3]. With the rapid development of global manufacturing and continuous technological progress, automatic assembly technology has become an important means to improve production efficiency and reduce labor costs. In manufacturing industries such as automobiles, ships, and aerospace, automatic assembly technology not only improves product quality and consistency but also significantly enhances the competitiveness of enterprises. Especially in the train manufacturing industry, with the popularization of high-speed trains and urban rail transit, higher requirements have been put forward for the assembly accuracy and efficiency of train components. As an important component of the train, the assembly quality of the side window glass directly affects the safety and comfort of the train. Therefore, researching an efficient and accurate method for assembling train side window glass is of great significance.

Laser displacement sensors, as a high-precision and non-contact measurement tool, have also been widely used in the manufacturing industry. By emitting a laser beam and receiving reflected light, laser displacement sensors can achieve accurate measurements of the surface displacement of objects. During the assembly process of train side window glass, laser displacement sensors can monitor the gap between the glass and the window frame in real-time, ensuring the tightness and stability of the assembly. In addition, laser displacement sensors have the advantages of wide measurement range, high measurement accuracy, and fast response speed, which can meet various measurement needs during the assembly process of train side window glass.

Combining machine vision technology with laser displacement sensors can achieve comprehensive monitoring and control of the assembly process of train side window glass. By using machine vision technology to obtain image information of glass and window frames, and using image processing algorithms to process and analyze the images, the position, shape, size, and other information of the glass and window frames can be obtained. At the same time, real-time displacement data during the assembly process can be obtained by measuring the gap between the glass and the window frame through a laser displacement sensor. Combining these two technologies can achieve precise control over the assembly process of train side window glass, improving the accuracy and efficiency of assembly. Since the side window glass is a large flat workpiece, its assembly process faces many challenges, such as high positioning accuracy requirements and difficulty in ensuring surface flatness. To address these challenges, this article will focus on the deep integration of machine vision technology and assembly technology, particularly the combination of laser displacement sensors (LDS) and monocular vision. LDS, as a high-precision, non-contact measuring tool, can obtain real-time three-dimensional coordinate information on the workpiece surface. During the assembly process of train side window glass, LDS can accurately measure the displacement changes on the glass surface, providing reliable data support for subsequent assembly work. Meanwhile, monocular vision, as an economical and efficient image processing technique, can capture the pose features of the upper right corner of the window, providing necessary compensation information for precise assembly. Although machine vision technology has been widely applied in the field of assembly, there are still many urgent problems to be solved for the assembly of large flat workpieces such as train side window glass. For example, how to effectively integrate data from LDS and monocular vision to achieve high-precision assembly positioning and how to adjust the robot's motion trajectory based on measurement data to ensure the stability and accuracy of the assembly process. Therefore, this article aims to propose a practical intelligent robot assembly method through in-depth research on the combination of LDS and monocular vision to solve the challenges in the assembly process of train side window glass.

1.1 Literature review

A precision assembly system for tiny components was created by Ma et al. [4] using a force sensor, three cameras, and an industrial robot. They proposed a pose alignment strategy based on binocular coordination and employed the differential principle to compensate for actuator position offsets. Experimental results demonstrated a high level of pose assembly accuracy for the system. Using real-time visual servoing, Wen-Chung Chang [5] created an automated method for assembling the shell of smartphones. This system utilized binocular vision technology to ascertain the position and orientation of the workpiece from a distance, subsequently applying monocular visual servoing control to complete the assembly task at close range autonomously. Experimental results indicated that the system provided high efficiency, flexibility, and reliability. However, this assembly approach necessitated that the entire assembly space remains fully visible, making it suitable for small workpieces but challenging to implement for larger workpieces with long-distance and expansive requirements. Su et al. [6] proposed a method for designing orientation strategies for locating irregular parts based on constraints formed by fixtures. Therefore, only through the squeezing action of robots can irregular parts such as motors be accurately located. Then, we developed a bolt hole insertion strategy using force constraints formed by the environment, which requires less robot action to insert the bolt, thereby achieving efficient and accurate bolt hole insertion. In contrast, Jinzhou et al. [7] developed a monocular vision-based alignment technology for an industrial robot. It estimates the target pose of the workpiece with the Perspective-N-Point (PNP) method by leveraging some feature points on a dedicated calibration board. Following that, it repeatedly modifies the end effector's position until the workpiece and robot's alignment axis are precisely aligned. However, this method requires setting a special calibration board on the target workpiece, which to some extent constrains its generality. Wang et al. [8] presented a holistic robotic assembly system including two vision sensors and three one-dimensional laser sensors. They have presented a method that includes multiple sensors to achieve precise step-by-step alignment and positioning. Experimental findings showed that this approach offered high accuracy and efficiency but, on the other hand, is associated with the assembly scheme having high costs and involving a complex algorithm. Hu et al. [9] studied side window glass designs using the Large Eddy Simulation (LES) model, computational fluid dynamics (CFD) simulation, and acoustic finite element analyses to investigate its impact on in-vehicle noise. They found that polyvinyl butyral (PVB) laminated glass effectively reduces noise transmission caused by low-frequency turbulence, with the inner and outer thicknesses varying in their effects across different frequency bands; however, low frequencies were found to be the most influential. Furthermore, material properties and geometric design are critical factors in noise control. Zhang et al. [10] have studied the application of an external circular array of LDSs which detect membrane penetration to enhance the precision of the volume change as well as pore pressure in triaxial gravelly sandy soils. The results indicated that unit membrane compliance increases linearly with effective confining pressure on a semi-logarithmic scale. Additionally, it was demonstrated that increased gravel content leads to greater membrane penetration; however, this effect diminishes at higher relative densities. It was also observed that broadly graded soils exhibit higher membrane penetration. The errors associated with membrane thickness increased as the diameter-to-thickness ratio decreased, highlighting the importance of sensors in detecting subtle variations to provide consistent measurements. Shao et al. [11] proposed a monocular vision-based system for vibration displacement measurement in civil structures to solve several of the major problems associated with multiple cameras used in structural health monitoring (SHM) systems. Deep neural networks have been utilized to enable substantial reduction of costs and ease in estimating depth information from images captured by a single camera. Laboratory tests verified the accuracy of the system output that provided three-dimensional displacement measurements and demonstrated the potential of the system for practical applications in SHM. Chen et al. [12] developed an algorithm for applying monocular vision to quantify construction and demolition waste from individually taken images. The results of the research were promising: in this system, highly accurate identification was realized for truckloads of materials like rock, gravel, and wood. Accompanied by a comparative error of 0.065 for truck bucket measurements and 0.169 in approximating material volume, the algorithm analyzes an image in only 3.3 seconds, which enables fast and non-intrusive assessments. The methods were adopted for over 2,900 truckloads in a facility in Hong Kong to estimate that the daily intake of construction and demolition waste (CDW) was at 800.0 m³, along with identifying rejected loads containing non-inert materials. Thus, this algorithm will be able to work effectively on different industrial challenges in the quantification of waste. Sun et al. [13] present a monocular vision-based approach for measuring 3D movement regarding enabling single-camera SHM applications. This high-level method captures data from 3D displacement with high accuracy by merging deep learning techniques through pose estimation from the Dense Pose Object Detector. It doesn't require multi-view systems or sophisticated targets, so it would be quite in line with urban surveillance cameras. The experimental validation showed that it is effective for detecting 3D displacement and principal frequencies. It has also been proven to be workable regarding the applicability of the method as a reachable SHM tool. Sleaman et al. [14] developed an algorithm that allowed a mobile robot to move around in an unfamiliar environment, leveraging the power of only monocular vision and low-cost cameras for depth perception. This system combined convolutional neural network (CNN) layers with decision-making processes that modulate real-valued, kinematic movement control without any preprocessing or pre-mapping. It was capable of training on a set of diverse data and achieved an accuracy indoor setting of up to 77%. This model proved effective enough to facilitate efficient navigation of the robot compared to more expensive sensor systems. Sun et al. [15] proposed an algorithm that developed a monocular vision-based method, which allowed a mobile robot to navigate an unknown environment autonomously by low-cost depth camera data. They have introduced the integration of CNN layers with decision-making processes to guide robot movement without any pre-mapping. The system has been trained on various datasets and showed an accuracy of 77% during indoor tests to prove that the technique can be as effective in allowing efficient navigation by the robot as compared to more expensive sensor systems. Gao et al. [16] proposed an algorithm, Perspective 3-Point Rammer Pose Estimation (P3P-RPE), derived from monocular vision for real-time ramming settlement monitoring in dynamic compaction construction. This approach overcomes the substantial limitations that can compromise efficiency and safety in the construction process through manual monitoring. The use of the P3P-RPE algorithm will allow for simultaneous construction and monitoring of processes without disrupting the ongoing operations. Lou et al. [17] proposed a SLAM solution using monocular vision, integrated with LiDAR, which could afford better placement of unmanned vehicles. They integrated semantic images with LiDAR point clouds to give dense depth mapping and improved positioning accuracy with in-depth embedded ORB（Oriented FAST and Rotated BRIEF）features. When their proposed approach was tested on CityScapes and KITTI datasets, it could reduce the positioning error by 87% compared with other approaches in the SLAM category. Results showed the huge impact monocular vision could have when applied to applications involving autonomous vehicles. A self-calibration method based on monocular vision was introduced by Li et al. [18] for the accurate assessment of 3D displacement in structures. In this regard, the approach is based on a personalized marker to accurately recognize images for the automatic estimation of the object's 3D position and orientation with the purpose of solving problems arising in calibration for long-term monitoring. Monte Carlo simulations and experiments were carried out for verification, obtaining that the system achieved an accuracy of about 0.049 mm. It follows that 3D displacement measurements may be efficiently carried out in fully automated ways for applications like structural tests. Ahn et al. [19] propose a robotic assembly that fuses both force and visual information over large position and orientation errors. This is realized through the use of dual neural networks that process both force and image inputs, optimized through deep reinforcement learning, thus giving adaptive assembly capabilities to a variety of shapes. By this approach, an assembly can be effectively realized despite the initial misalignment. The experimental validation is performed for peg-in-hole tasks and provides a general approach to a wide range of robotic assembly applications. A robotic assembly system for precision drilling and fastening in aircraft assembly has been proposed by Mei et al. [20]. Optimum hand-eye configuration and on-machine calibration are considered to minimize positioning errors in the system. The maximum error of 0.08 mm was obtained, hence meeting the strict accuracy criteria required in aircraft assembly. Zhang et al. [21] presented the residual reinforcement learning approach for robotic autonomous assembly, in which policy visualization and force-based methods are well integrated. The approach makes it possible to be adaptive in an unstructured environment without prior data. Efficiency is pointed out in multiple assembly tasks and configurations verified by simulations and real-world tests. Champatiray et al. [22] developed a methodology of robotic assembly planning that comprises the development of assembly interference matrices" and an axis-aligned bounding box" to assess any interference and determine the optimized assembly sequences of the assemblies. This will enhance feasibility and efficiency in industrial applications by taking into consideration geometry from the involved tools and parts, hence saving time and costs. Machine vision technology is a technique that uses computer vision and image processing algorithms to simulate human visual functions. In the experiment of X-ray free electron laser, machine vision technology can monitor the stability of sample transportation by real-time acquisition and processing of images of liquid jet samples [23]. The collected images need to undergo a series of image processing algorithms, including image filtering, edge detection, feature extraction, etc. These algorithms can help us identify key information such as the shape, velocity, and position of liquid jet samples. By further analyzing the processed images, we can calculate the stability indicators of the liquid jet sample, such as fluctuations in jet velocity and deviations in jet direction. These indicators can be fed back to the control system in real time to adjust and optimize the parameters of sample delivery promptly [24].

Based on the automatic assembly task for the side window glass of a subway train, this study investigates the low-cost spatial assembly problem of large planar workpieces. A monocular vision-guided assembly method is proposed, assisted by a LDS. An experiment uses a specific case to provide a reference for designing automatic assembly schemes for similar workpieces.

1.2 Research gaps and novelties

Table 1: Method comparison table

Method	Accuracy	Complexity	Cost	Limitations
Multi camera setup	Medium to high	high	high	Requires complex calibration and synchronization, and is sensitive to lighting conditions
Specialized sensors	high	Medium to high	high	High cost and may not be suitable in certain environments (such as outdoor or complex industrial environments)
Manual monitoring	Low to moderate	low	low	Dependent on the operator's experience and skills, not suitable for high-precision or long-distance work
Visual based method (sensorless)	Low to moderate	secondary	Low to moderate	Highly affected by lighting, shadows, and occlusion, with limited accuracy
Hybrid method (multi-sensor fusion)	high	high	high	Although accuracy has been improved, system complexity and cost remain high
The proposed method (LDS+monocular vision)	high	secondary	Low to moderate	Relying on the fusion quality of LDS and visual data, but reducing system complexity and cost

The main innovation of this study lies in proposing a hybrid approach that combines LDS and monocular vision for robust real-time pose correction and component alignment. Compared with multi-sensor systems, this method reduces system complexity and cost through integration. Specifically, this study utilized point cloud data generated by LDS to determine alignment parameters by fitting the side window plane equation, thereby improving assembly accuracy. Meanwhile, utilizing monocular vision effectively extracts the pose features of the upper right corner of the window, providing necessary compensation information for precise assembly. This method not only solves the research purpose

The requirement for precise positioning and orientation with displacement tolerance of less than 0.5mm and directional tolerance of less than 1 degree also fills a key gap in providing practical and low-cost solutions for the assembly of large planar components.

The objective of this study is to validate the effectiveness and feasibility of the proposed hybrid method in practical industrial environments. Specifically, this study aims to demonstrate the following points:

The proposed hybrid method can reduce system complexity and cost while maintaining high accuracy.

Compared with existing multi-sensor systems, this method has higher efficiency and lower cost.

This method has broad application prospects and can be extended to similar tasks in various industries.

1.3 Paper organization

The work is categorized in the manner that: Section 2 offers and discusses the constitution and calibration of the empirical system to assemble subway train side window glass. Section 3 provides a detailed description of the visual alignment method for the LDS. Section 4 develops an optical positional measuring technique. The analysis of defects encountered during assembly and the experimental validation of the proposed assembly methodology for subway train side window glass are outlined in Section 5. Ultimately, Section 6 concludes the investigation.

2 Composition and calibration of experimental system for side window glass assembly of train

The automated assembling of glass for windows in trains requires an accurately defined experimental setup. This section details the basic elements that form the experimental setup, their functions, and the necessary calibration processes toward achieving maximum accuracy in assembling processes. The system includes an industrial robot with six axes, an industrial camera, an LDS (Laser distance sensor), and an aligning fixture that allows accurate placement and alignment.

2.1 Composition of experimental system and definition of coordinate system

The experimental system for the train-side window glass assembly described in this paper comprises a mounting bracket, a six-axis industrial robot, an industrial camera, an LDS, and a fixture, as illustrated in Fig. 1. This experiment uses a high-precision laser displacement sensor with a measurement accuracy of micrometers and a measurement range of meters. LDS calculates the distance between an object and a sensor by emitting a laser beam and receiving the reflected light signal. In this study, LDS was used to accurately measure the displacement changes on the surface of the train side window glass, providing key data for subsequent assembly work. The fixture is firmly connected to the tail flange of the robot, with its plane parallel to the flange plane.

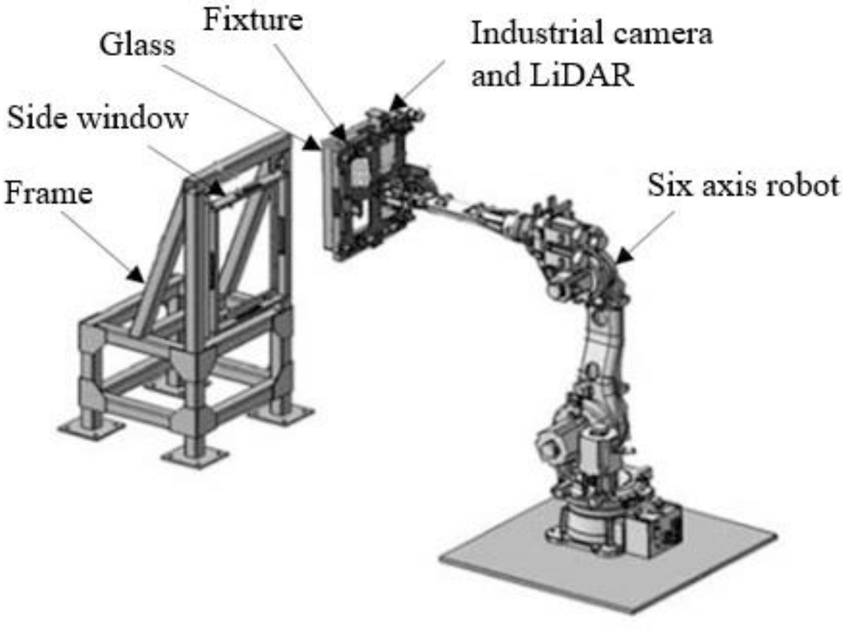


Figure 1: Schematic diagram of automatic assembly system for side window glass of train.

Installation bracket: This bracket is the cornerstone of the experimental system, used to simulate the structure of train carriages and ensure accurate assembly testing of glass in real environments.

Six axis industrial robot: As the core power source of the system, the six-axis industrial robot is responsible for grasping, moving, and positioning glass. Its flexible six axis design enables the robot to execute complex motion trajectories in three-dimensional space, thereby achieving precise assembly operations.

Industrial camera: Industrial cameras are installed at the edge of fixtures to capture image information of glass and side window positions. Through image processing algorithms, edge features of glass and side windows can be extracted, and their relative positions and orientations can be calculated.

Laser Ranging Sensor (LDS): LDS is also installed at the edge of the fixture to measure the precise distance between the glass and the side window. The high precision and stability of laser ranging enable the system to maintain a high degree of accuracy during assembly.

Fixture: The fixture is firmly connected to the tail flange of the robot, with its plane parallel to the flange plane. The design of the fixture takes into account the size and shape of the glass to ensure that it can be stably grasped and positioned. The establishment of the flange coordinate system (CS) takes place at the center point of the flange end plane, the same as the location of the X_H O_H Y_H plane. In this case, X_H and Y_H are the flange coordinate system principal axes, while O_H refers to the origin. Coordination by this system is necessary to facilitate accuracy in robot assembly processes. These same coordinates are specifically referred to in Equation (X) when calculating transformations, while being indicated by Figure (Y) so as to give an idea of their location in three-space.

The base coordinate system (CS) of the robot is defined as $O_BX_BY_BZ_B$, and that of the workpiece as $O_TX_TY_TZ_T$. Besides them, other CSs are defined for the end flange, camera, and image pixels, as $O_HX_HY_HZ_HO_CX_CY_CZ_C$, and ouv, respectively. The origin and axis directions of the base CS of the robot should be determined in settings specified by the factory. The operator defines a CS for the workpiece at a position close to where the work will be done. During an assembly operation, this becomes a dynamic reference frame that replaces the base frame. The source of the flange CS is positioned at the core of the end plane of the flange, coincident with the $X_HO_HY_H$ plane. In this system, the Z-axis points outward and is vertical to the plane of the flange. In the picture pixel CS, the u-axis points horizontally to the right, the v-axis points vertically downward, and the origin of the coordinates is situated in the upper-left corner. The axes X_C and Y_C are parallel to the u- and v-axes of the image, respectively; the Z_C-axis points out of the imaging plane.

2.2 System calibration

First, the system must be calibrated to achieve its pose correction and positioning functions. The calibration process includes laser sensors, cameras, and hand-eye calibrations.

To calibrate the position of the laser range finder, solve for the laser source coordinate matrix in the robot's end flange CS, r, and the unit direction vector matrix of the laser line, l. Planar constraints can be used to resolve it. It is then combined with the interval measured by the Laser Sensor and the positioning of the robot's end to determine the coordinates of the laser point in the workpiece CS [25], [26].

In experimental systems, calibration is a key step in ensuring the precise operation of laser sensors, cameras, and hand eye systems. The calibration process includes laser sensor calibration, camera calibration, and hand eye calibration.

The calibration of laser sensors is mainly to ensure the accuracy of their measurements. The specific calibration steps include:

Installation and preheating: Install the laser sensor on the fixture and preheat it to ensure its stable operation.

Zero-point calibration: Place a reference surface at a known distance in front of the sensor, adjust the sensor to read zero, and complete the zero-point calibration.

Linearity calibration: Move points at different distances on the reference surface, record the sensor readings, compare them with the actual distance, and adjust the sensor parameters to ensure linearity.

A pinhole camera model can be used to represent how a camera captures images. Based on a calibration process, the internal camera parameters are derived, represented as a matrix A in terms of radial distortion parameters K_1 , K_2 , K_3 , and tangential distortion parameters p_1 and p_2, along with the external parameters $[R\ t]$. Internal parameters describe the relation of pixel coordinates with image coordinates, while external parameters describe the relation of the camera's CS with the world's CS. Images of the checkerboard calibration board are taken by the camera at different angles so that both the world coordinates and the corresponding pixel coordinates of the checkerboard corners can be obtained. Calibration parameters are computed using the Zhang Zhengyou method [27].

Hand-eye calibration is performed to estimate the homogeneous transformation matrix between the camera's CS and the robot's end flange CS, denoted by ${}^{H}_{C}T$. Estimates of the parameters of such calibration are obtained with the TSAI method [28], which combines the extrinsic parameters attained from camera calibration with the corresponding pose parameters of the robot's end effector.

To calibrate the position of the laser rangefinder, we need to solve the laser source coordinate matrix r and the unit direction vector matrix l of the laser line in the end flange coordinate system (CS) of the robot. Select a known plane as the reference plane, which intersects the laser line at a point. By moving the laser rangefinder to multiple known positions and recording the laser measurement values and robot posture at each position, a series of equations can be established to solve r and l. By combining the interval measured by the laser rangefinder with the positioning information of the robot end, the coordinates of the laser point in the workpiece coordinate system (CS) can be further determined. This usually involves converting laser measurement values into coordinates in the robot coordinate system and then converting them into the workpiece coordinate system.

3 Visual alignment on the basis of LDS

Accurately positioning the side window through visual alignment relies on a precise characteristic image of the side window. Therefore, the first step is to achieve vision alignment r, ensuring that the camera's imaging plane is along with the side window plane to prevent imaging distortion. In this study, a laser range finder is employed to project a laser point onto the side window plane, and the coordinate data of the laser point are utilized to fit the equation of the side window plane. The normal vector (NV) parameters are then obtained, allowing for the establishment of the camera coordinate system (CCS). The equation of the axis Z_{C} and the plane's NV of the side window determines the posture correction angle of the robot's end flange. The axis Z_{C} is aligned with the NV of the side window plane to achieve visual alignment. Additionally, the interval between the camera's imaging plane and the side window plane must be adjusted using an LDS to meet the preset value, ensuring consistency in the imaging scale. Conduct visual alignment tests under different environmental lighting and reflective surface conditions to verify the robustness and accuracy of the alignment algorithm. By enhancing the recognition and fitting of laser points, improving the alignment between the camera coordinate system and the side window plane, optimizing the use and adjustment of LDS, and enhancing the robustness testing and verification of visual alignment, the robustness of LDS based visual alignment can be significantly improved under environmental lighting or reflective surface changes.

3.1 Unit normal vector of the side window plane

A laser range finder is used to project multiple laser points onto a side window plane r, and the laser line's unit direction vector coordinate matrix l. The coordinate matrix of the i-th laser point in the workpiece CS can be listed as P_i. The expression is:

$$P_i = {}_{H}^{T}T \begin{pmatrix} r + d_i l \\ 1 \end{pmatrix} \tag{1}$$

Where ${}_{\mathrm{H}}^{\mathrm{T}}T$ is the transformation matrix of the robot end flange CS about the workpiece CS, which is determined from data transformation of the teaching pendant, and d_{i} is the distance value measured by a laser sensor.

Due to various errors, the coordinates of the measured laser points cannot be strictly located in the same plane, so applying the coordinate data to fit the plane equation of the side window to minimize the distance between each laser point and the plane is essential. According to the theory of spatial analytic geometry, the ideal equation of the side window plane in the workpiece CS is:

$$Ax + By + Cz - D = 0 \tag{2}$$

In this relation, A, B, and C are components of the UNV of the side window plane in the workpiece CS, and D is the interval between the origin of the workpiece CS and the side window plane. Let the overall amount of laser points be n; the components of the i-th laser point in the workpiece CS are x_i, y_i, z_i, at a interval from the side window plane $|Ax_i + By_i + Cz_i - D|$. The problem is then reduced to finding the coefficients of the plane equation such that:

$$\begin{cases} min(f) = min \sum_{\mathrm{i}=1}^{n} (Ax_i + By_i + Cz_i - D)^2 \\ A^2 + B^2 + C^2 - 1 = 0 \end{cases} \tag{3}$$

Coefficient A, B, C, D： These are known constants that appear in the formula as weights or offsets. Variable xi, yi, zi： These are variables that need to be optimized, and for each i (from 1 to n), there is a set of such variables. They represent a point in the problem space. The summation symbol $\sum$ represents the summation of all i values from 1 to n. That is to say, function f is the sum of all n terms. Each term in the square operation is squared. This is to make the function f non negative, and there will be a positive penalty for any deviation from the optimal solution. For the minimization problem described by Equation (3), the function is constructed according to the Lagrange multiplier method:

$$\begin{aligned} F = \sum_{i=1}^{n} (Ax_i + By_i + Cz_i - D)^2 \\ + \lambda(A^2 + B^2 + C^2 - 1) \end{aligned} \tag{4}$$

Assuming the objective function is f (x) and the constraint condition is g (x)=0 (where x can be a multidimensional vector). The basic idea of Lagrange multiplier method is to introduce a Lagrange multiplier λ and integrate the constraints into the objective function, thereby forming a new function - Lagrange function L (x, λ). The Lagrange multiplier method is usually very effective in dealing with optimization problems with simple constraints. However, as the constraints become more complex or the number increases, the computational complexity may significantly increase. In addition, if the objective function or constraint conditions are nonlinear, the solving process may involve complex numerical methods. For optimization problems with complex constraints or nonlinear objective functions, other optimization methods such as interior point method, penalty function method, or trust region method can be considered. These methods may be more effective or easier to implement than the Lagrange multiplier method in certain situations. In applications such as attitude estimation, noise measurement is a common problem. In order to improve the efficiency and robustness of eigenvalue calculation, some numerical stability techniques can be used, such as regularization methods, singular value decomposition (SVD), etc. In addition, for large-scale datasets or high-dimensional problems, efficient numerical algorithms and parallel computing techniques can be considered to accelerate the computation process. For a minimal value, there are $\frac{\partial F}{\partial A} = \frac{\partial F}{\partial B} = \frac{\partial F}{\partial C} = \frac{\partial F}{\partial D} = 0$. After derivation and simplification, the following equations are achieved:

$$\begin{cases} \sum_{i=1}^{n} x_i(Ax_i + By_i + Cz_i - D) + \lambda A{=}0 \\ \sum_{i=1}^{n} y_i(Ax_i + By_i + Cz_i - D) + \lambda B{=}0 \\ \sum_{i=1}^{n} z_i(Ax_i + By_i + Cz_i - D) + \lambda C{=}0 \\ D = \frac{1}{n}\sum_{i=1}^{n}(Ax_i + By_i + Cz_i) \end{cases} \tag{5}$$

$$\begin{bmatrix} \sum_{i=1}^{n} x_i\, v_{x_i} & \sum_{i=1}^{n} x_i\, v_{y_i} & \sum_{i=1}^{n} x_i\, v_{z_i} \\ \sum_{i=1}^{n} y_i\, v_{x_i} & \sum_{i=1}^{n} y_i\, v_{y_i} & \sum_{i=1}^{n} y_i\, v_{z_i} \\ \sum_{i=1}^{n} z_i\, v_{x_i} & \sum_{i=1}^{n} z_i\, v_{y_i} & \sum_{i=1}^{n} z_i\, v_{z_i} \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = (-\lambda) \begin{bmatrix} A \\ B \\ C \end{bmatrix} \tag{6}$$

$$\begin{cases} v_{x_i} = x_i - \overline{x}, \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ v_{y_i} = y_i - \overline{y}, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \\ v_{z_i} = z_i - \overline{z}, \quad \overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \end{cases} \tag{7}$$

$$\begin{cases} \sum_{i=1}^{n} \overline{x} v_{x_i} = \sum_{i=1}^{n} \overline{x} v_{y_i} = \sum_{i=1}^{n} \overline{x} v_{z_i} = 0 \\ \sum_{i=1}^{n} \overline{y} v_{x_i} = \sum_{i=1}^{n} \overline{y} v_{y_i} = \sum_{i=1}^{n} \overline{y} v_{z_i} = 0 \\ \sum_{i=1}^{n} \overline{z} v_{x_i} = \sum_{i=1}^{n} \overline{z} v_{y_i} = \sum_{i=1}^{n} \overline{z} v_{z_i} = 0 \end{cases} \tag{8}$$

$$\begin{cases} \sum\limits_{i=1}^{n}(x_i - v_{x_i})v_{x_i} = \sum\limits_{i=1}^{n}(x_i - v_{x_i})v_{y_i} = \sum\limits_{i=1}^{n}(x_i - v_{x_i})v_{z_i} = 0 \\ \sum\limits_{i=1}^{n}(y_i - v_{y_i})v_{x_i} = \sum\limits_{i=1}^{n}(y_i - v_{y_i})v_{y_i} = \sum\limits_{i=1}^{n}(y_i - v_{y_i})v_{z_i} = 0 \\ \sum\limits_{i=1}^{n}(z_i - v_{z_i})v_{x_i} = \sum\limits_{i=1}^{n}(z_i - v_{z_i})v_{y_i} = \sum\limits_{i=1}^{n}(z_i - v_{z_i})v_{z_i} = 0 \end{cases} \tag{9}$$

$$\begin{bmatrix} \sum_{i=1}^{n} v_{x_i}\, v_{x_i} & \sum_{i=1}^{n} v_{x_i}\, v_{y_i} & \sum_{i=1}^{n} v_{x_i}\, v_{z_i} \\ \sum_{i=1}^{n} v_{y_i}\, v_{x_i} & \sum_{i=1}^{n} v_{y_i}\, v_{y_i} & \sum_{i=1}^{n} v_{y_i}\, v_{z_i} \\ \sum_{i=1}^{n} v_{z_i}\, v_{x_i} & \sum_{i=1}^{n} v_{z_i}\, v_{y_i} & \sum_{i=1}^{n} v_{z_i}\, v_{z_i} \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \lambda' \begin{bmatrix} A \\ B \\ C \end{bmatrix} \tag{10}$$

The coefficient matrix on the left side of equation (10) is a third-order real symmetric matrix, which we represent as RC. The elements of this matrix are related to the variables in the equation, and some important properties, such as eigenvalues and eigenvectors, will play a crucial role in subsequent analysis. Specifically, we can consider λ, a, B, and C (where a, B, and C may represent elements of the eigenvector. However, their specific meanings need to be determined based on the context, as this was not explicitly stated in your original description) as elements related to the eigenvalues and eigenvectors of RC. To find the eigenvalues of R C, we set the characteristic polynomial to zero, i.e. ｜ R C − λ′ I ｜=0 (Equation 11), where I is the identity matrix. In this way, we can find all the eigenvalues of R C by solving this equation Where λ'=-λ. It is clear that the coefficient matrix on the left side of Eq. (10) is a real symmetric matrix of order 3, while λ', A, B, and C can be regarded as the eigenvalues and eigenvector elements of the coefficient matrix. This coefficient matrix is denoted as R_c, and its eigenvalues can be obtained by Eq. (11) as follows:

$$|\boldsymbol{R}_c - \lambda'\boldsymbol{I}| = 0 \tag{11}$$

According to the eigenvalue method [29], when the eigenvalue is the smallest, the interval between each laser point and the plane is minimized, and the corresponding eigenvector is the UNV of the side window plane, thus obtaining the values of A, B and C. To satisfy the subsequent robot end flange posture adjustment requirements, the NV must point to the inside of the side window. Considering the bi-directionality of the plane NV, if the obtained NV points to the outside of the side window, its reverse vector is taken.

3.2 Pose correction of the robot end flange

To make the camera imaging plane parallel to the side window plane, the pose of the robot end flange needs to be adjusted. In this paper, the XYZ fixed angle method is used for rotating around the X, Y, and Z axes of the workpiece CS, with angles γ, β, and α, so that the CCS axis Z_C is in the same direction as the side window UNV obtained in Section 3.1. In certain fields or applications, specific rotation sequences may be widely adopted due to historical reasons or conventions. For example, in aerospace, the ZYX sequence (also known as yaw pitch roll) may be preferred as it corresponds to the natural motion of the aircraft. In other fields such as robotics or automation, XYZ order may be more common. For different rotation orders, even if the rotation angle is the same, the resulting rotation matrix will be different. This is because each rotation is relative to a different axis.

Some rotation sequences are more susceptible to the influence of universal joint locks, that is, when two rotation axes coincide, one dimension of rotational ability will be lost. This requires special attention when using Euler angles to represent rotation. In Eq. (12**)**, ${}^{H}_{C}\boldsymbol{R}$ refers to the rotation transformation matrix, through which the camera coordinate system (C) gets transformed into the robot end flange coordinate system (H). The matrix depicts the directional correlation between the camera and robot end flange, thus providing the accurate alignment necessary for successful pose correction. Suppose that the rotation transformation matrix from the CCS to the robot end flange is as follows:

$$ {}^{H}_{C}\boldsymbol{R} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{12} $$

Then, the Z_C component of the CCSs axis's unit direction vector in the flange CS is $[a_{13} \quad a_{23} \quad a_{33}]^T$. The rotation matrix is projected to the workpiece CS using the XYZ fixed angle method.

$$ \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} c\,\alpha\, c\,\beta & c\,\alpha\, s\,\beta\, s\,\gamma - s\,\alpha\, c\,\gamma & c\,\alpha\, s\,\beta\, c\gamma + s\,\alpha\, s\gamma \\ s\,\alpha\, c\,\beta & s\,\alpha\, s\beta\, s\,\gamma + c\alpha\, c\,\gamma & s\alpha\, s\,\beta\, c\gamma - c\alpha s\gamma \\ -\,s\,\beta & c\,\beta\, s\gamma & c\,\beta\, c\,\gamma \end{bmatrix} \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} \tag{13} $$

Where c stands for cos and s stands for sin. The angle data in the robot teach pendant at the time of projecting the last laser point used to fit the plane can be used as an initial value α_0, β_0, and γ_0. Using the LM method for iterative optimization, solve the nonlinear equation (13) and finally obtain α, β, and γ. Based on this, adjust the flange attitude. Construct a nonlinear equation based on the rotation angles (α, β, and γ) of the robot flange and the actual measured laser point data (or other relevant data). This equation may describe the difference between the projection of the laser point on the fitting plane and the expected position. Use LM method to iteratively solve nonlinear equations. In each iteration, a new angle value is calculated based on the current angle value to reduce errors in the equation. This process will continue until a certain stopping criterion is met (such as an error less than a certain threshold or reaching the maximum number of iterations).

As mentioned above, after the camera imaging plane is aligned with the side window plane, it is essential to adjust the robot's end flange position to keep the interval between the camera imaging plane and the side window plane at a preset value. The actual interval between the two planes can be obtained using the laser range finder, and then the difference between the exact interval and the preset value, Δd, can be computed. Using the transformation matrix from the CCS to the robot end flange ${}^{\mathrm{H}}_{\mathrm{C}}R$, Δd is decomposed into the flange CS. The displacement correction component in the three-axis direction of the flange CS is:

$$ \begin{bmatrix} \Delta x_H \\ \Delta y_H \\ \Delta z_H \end{bmatrix} = {}^{H}_{C}\,R \begin{bmatrix} 0 \\ 0 \\ \Delta d \end{bmatrix} \tag{14} $$

Calculate the displacement correction of the robot's end flange using formula (14) and adjust the flange's position so that the interval between the camera imaging plane and the side window plane meets the preset value.

4 Vision-based positioning measurement

The combination of LDS and monocular vision has significant advantages in the assembly of train side window glass, such as relatively low cost and easy operation. LDS can provide high-precision distance measurement, while monocular vision can capture and process two-dimensional image information, and the combination of the two can achieve a certain degree of assembly accuracy. However, this combination also has limitations, such as the lack of depth information acquisition in monocular vision and the impact of lighting conditions on visual processing.

4.1 Pixel coordinate and physical coordinate conversion

After the alignment function is completed, the checkerboard is placed on the assembly plane, the camera captures the checkerboard image, and the proportional relation between the pixel and physical coordinates is computed using the real physical interval between the checkerboard corners and the corresponding pixel interval. Fig. 2 shows an acquired checkerboard image, and the proportional relation between the two CSs is:

$$ s = \frac{1}{n}\sum_{i=1}^{n}\frac{d_{xyi}}{d_{uvi}} \tag{15} $$

In the formula, d_{xyi} and d_{uvi} are the Euclidean intervals of the i-th pair of checkerboard corner points in the checkerboard CS and the Euclidean interval in the pixel CS, respectively. The accurate proportional coefficient between these two CSs can be obtained by averaging the proportional relationships of n pairs of corner points, s.

Visual data (usually from industrial cameras) plays a crucial role in the assembly process. In order to extract useful information from images, this article adopts an advanced image processing algorithm. These algorithms can identify and locate the edge features of glass and side windows, and then calculate their relative positions and orientations.

Specifically, the algorithm first improves image quality through preprocessing steps such as denoising and contrast enhancement. Then, edge detection algorithms such as Canny edge detectors are used to identify the edges of the glass and side windows. Next, use feature matching algorithms such as SIFT or SURF to find the corresponding relationships between these edges. Finally, use these corresponding relationships to calculate the relative position and orientation between the glass and the side windows.

To demonstrate this process more intuitively, we can provide a series of image processing result graphs, including the original image, preprocessed image, edge detection results, and feature matching results. These charts will help readers better understand the working principle of the algorithm.

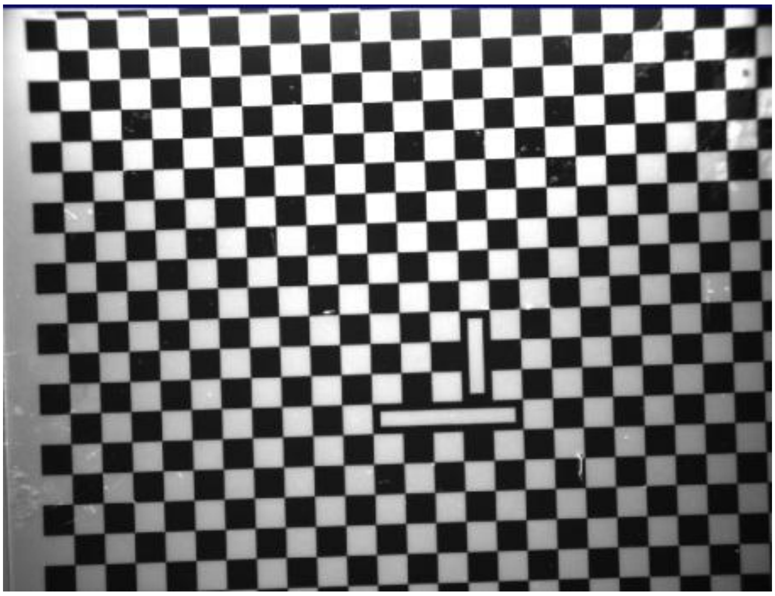


Figure 2: Checkerboard image

4.2 Feature identification

The arc part in the upper right corner of the side window, as seen in Fig. 3, is selected as the target feature to obtain the target's pose information. The straight-line equations of the two sides of the circular arc are obtained using the Hough transform. Hough transform is a classic algorithm used to detect specific shapes in images, such as lines, circles, arcs, etc. For arc detection, Hough Circle Transform or its variants can be used. Although Hough Circle Transform is mainly used to detect complete circles, it can also be used to detect arc parts by adjusting parameters and algorithms. Another method is to first detect the endpoints of the arc, then use these points to fit the arc, and then derive the equations for the lines on both sides based on the geometric properties of the arc, such as center, radius, and tangent direction. RANSAC (Random Sample Consensus) is an iterative algorithm used to estimate the parameters of a mathematical model from data containing a large amount of noise and outliers. It fits the model by randomly selecting a subset of data and evaluates the quality of the model by calculating the error between the remaining data and the fitted model. By combining the Hough transform and RANSAC least squares method, we can effectively identify the curved part of the upper right corner of the side window from the image and accurately fit the linear equations on both sides of the arc. These linear equations can be used for subsequent target pose information extraction and robot pose correction. Then, the accurate straight-line equation is obtained using the RANSAC least squares method [30].

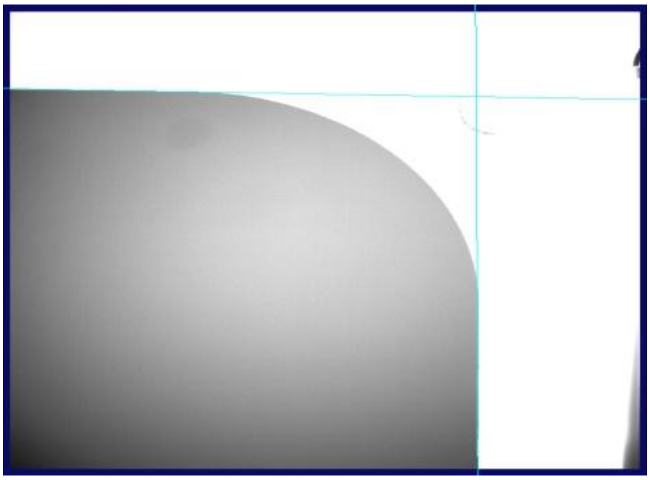


Figure 3: Target feature image

When the two lines intersect, that point serves as the displacement feature parameter for vision positioning, and the angle of the u-axis between them is used as a rotation characteristic parameter for visual positioning. The expected positioning characteristic parameters have been obtained by the above image processing method during offline teaching, and the positioning deviation can be determined by the feature elements of the current target image as follows:

$$\begin{aligned}\Delta u &= u_{\mathrm{d}} - u\\ \Delta v &= v_{\mathrm{d}} - v\\ \Delta\theta_{\mathrm{z}} &= \theta_{\mathrm{d}} - \theta\end{aligned} \tag{16}$$

Where $(u_{\mathrm{d}}, v_{\mathrm{d}})$ is the image coordinate of the intersection of the straight lines during teaching, (u, v) is the coordinate of the linear corner point in the current picture, θ_{d} is the angle of the straight-line during teaching, and θ is the line angle in the current image. Then $(\Delta u, \Delta v)$ is converted into physical coordinates using a proportional coefficient, and the offset matrix $\boldsymbol{T}'$ is obtained by combining $\Delta\theta_z$ as the primary data for the subsequent assembly process. θ_{d} and θ are the angle of the line in the teaching and the angle of the line in the current image, respectively. These angles may represent the angle between the line and a certain axis of the image coordinate system. The change in angle Δ θ=θ - θ_{d}can be used to describe the rotation of a straight line in an image. The offset matrix T 'is commonly used to describe the translation and rotation of an object in three-dimensional space. However, in your scenario, we may only be concerned with translation and rotation within the two-dimensional image plane. Therefore, T 'may be a 2x2 matrix (if only considering transformations within the plane) or a 3x3 matrix (if considering potential scaling and affine transformations, but keeping the z-coordinate constant).

4.3 Calculation and compensation of assembly pose

Visual positioning is divided into two steps: the calibration process and the assembly process.

(1) Calibration Process

Establish a workpiece CS T_0 at the upper right corner of the side window and record its pose as, T_0. With the assistance of the LDS, set the standard shooting pose of the camera and record the pose of the flange as C_Pos_0. The conversion relationship between C_Pos_0 and T_0 is $\left({}^{(T_0)}_{C\_Pos_0}H\right)$. The collected image of the upper right corner of the side window is used as the standard template, and a nine-point calibration is carried out to set up the mapping relation between the pixel and the workpiece CSs. The robot is manually guided to place the glass accurately in the workpiece CS T_0, and the flange pose is recorded as g_Pos_0. The mapping relationship between pixels and workpiece CS (T0) established through nine-point calibration should have a certain degree of accuracy. This can be evaluated by comparing the differences between the standard template images collected during the calibration process and the images collected during actual installation [31]. The difference should be within the preset accuracy threshold, for example, the pixel error should not exceed a certain amount. If LDS is used to fit the equation of the side window plane, the fitting result should be consistent with the actual plane. The fitting accuracy can be evaluated by comparing the difference between the fitted plane and the actual plane [32].

(2) Assembly Process

Adjust the robot flange with the laser sensor to make the camera plane along with the side window plane, and record the flange pose as C_Pos_1 . Based on the transformation relationship $\left({}^{T_0}_{C\_Pos_0}H\right)$, the pose of the workpiece CS T_1 can be calculated as $T_1 = C\_Pos_1 \cdot \left(\left({}^{T_0}_{C}H\right)^{-1}\right)$. Then, the product offset in the 2D plane is identified by the camera and the pose of the workpiece CS T_2 is calculated using the offset data, where $T_2 = T_1$.

Execute the installation pose g_Pos_0 in the T_2 CS, and the glass can be assembled accurately.

5 Experimental test and analysis of the result

5.1 Experimental test platform

Based on the assembly task requirements for the side window glass of a subway train in an enterprise, a platform for experimental testing was constructed, as displayed in Fig. 4. The size of the side window glass is 2 m × 1.5 m, and the working interval of the vision system is 400 mm. The attitude accuracy of the glass installation must be controlled within ±1°, and the positioning accuracy should be controlled within ±0.5 mm. The industrial robot is the ABB IRB7600-325, with a working range of 3.1 m and a maximum load capacity of 325 kg. The accuracy of repetitive positioning is 0.1 mm. The industrial camera is a CAM-CIC-4000 from Cognex, with a resolution of 2048 × 2048. The LDS is the LK-G500 from Keyence, with a range of 250 mm to 450 mm and a linear error of ±0.02%.

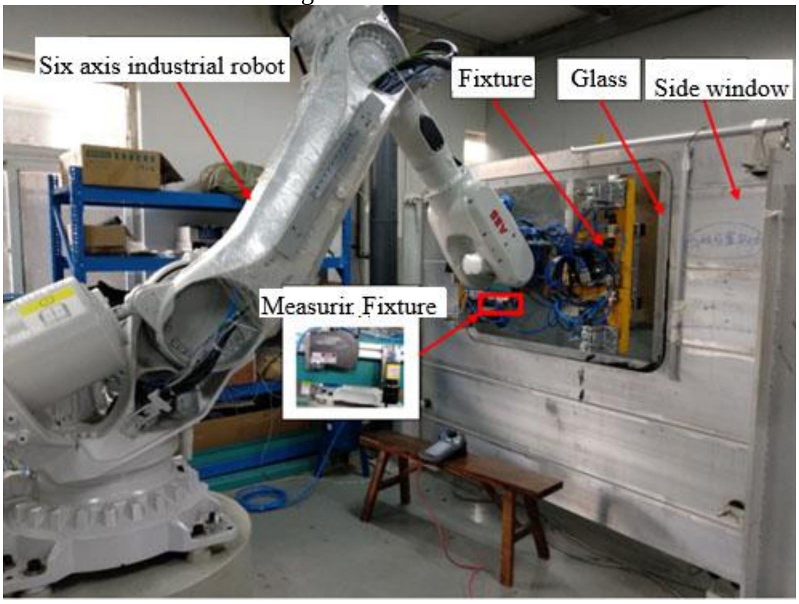


Figure 4: Assembly experiment site

5.2 System calibration results

Using the calibration method described above, the three types of calibration parameters of the system are first solved, and the final camera calibration result is as follows:

$$\boldsymbol{A}=\begin{bmatrix}5249.574 & 0 & 1014.700\\ 0 & 5248.587 & 1011.988\\ 0 & 0 & 1\end{bmatrix}$$

$$k_1=-0.158,\ k_2=-0.0249,\ k_3=0.0412,\ p_1=-0.0003,\ p_2=0.0004 \tag{17}$$

The hand-eye calibration result is:

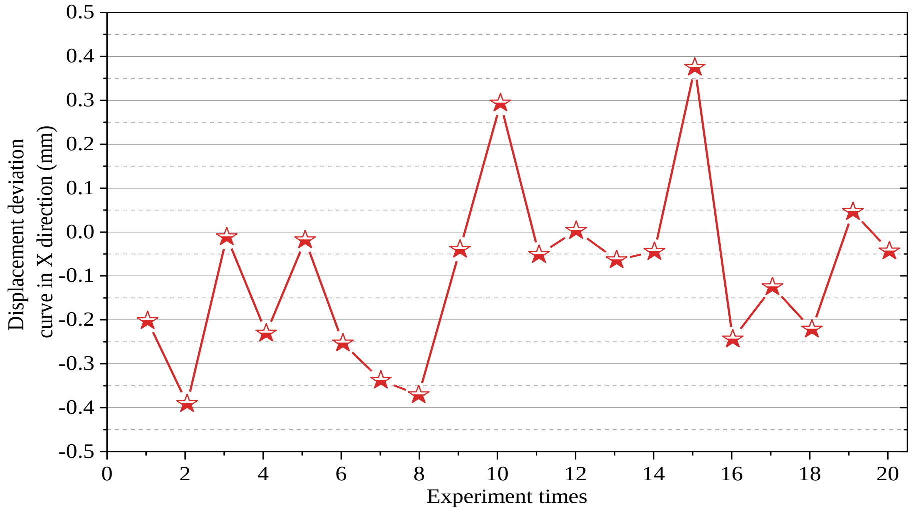
$${}^{C}_{H}\boldsymbol{T}=\begin{bmatrix}0.007 & -0.036 & -0.999 & -54.874\\ 1 & 0.0074 & 0.006 & -75.193\\ 0.007 & -0.998 & 0.036 & 102.960\\ 0 & 0 & 0 & 1\end{bmatrix} \tag{18}$$

The calibration results of the LDS are as follows:

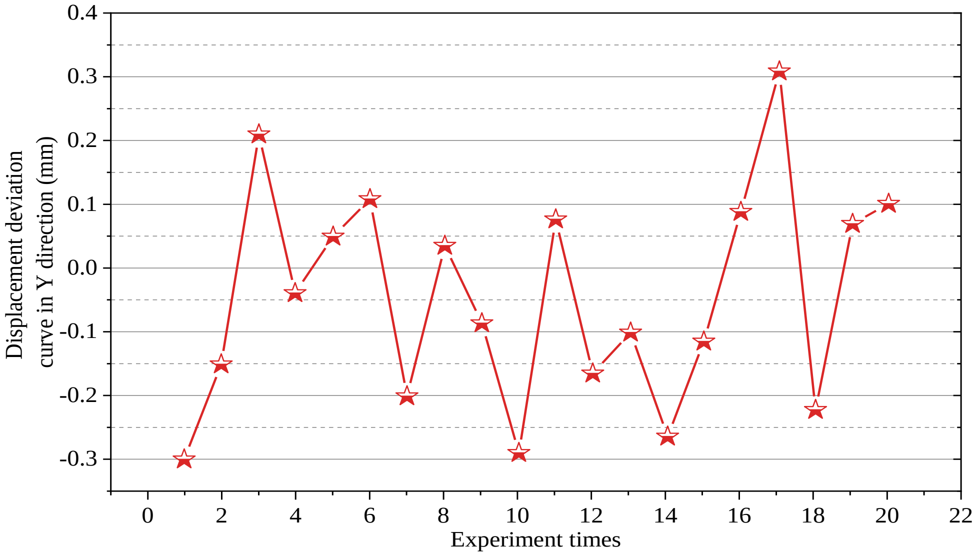
$$\begin{aligned}\boldsymbol{r}&=[43.6839,-40.4931,90.1015]\\ \boldsymbol{l}&=[0.9977,0.0626,0.0265]\end{aligned} \tag{19}$$

5.3 Analysis of assembly test results

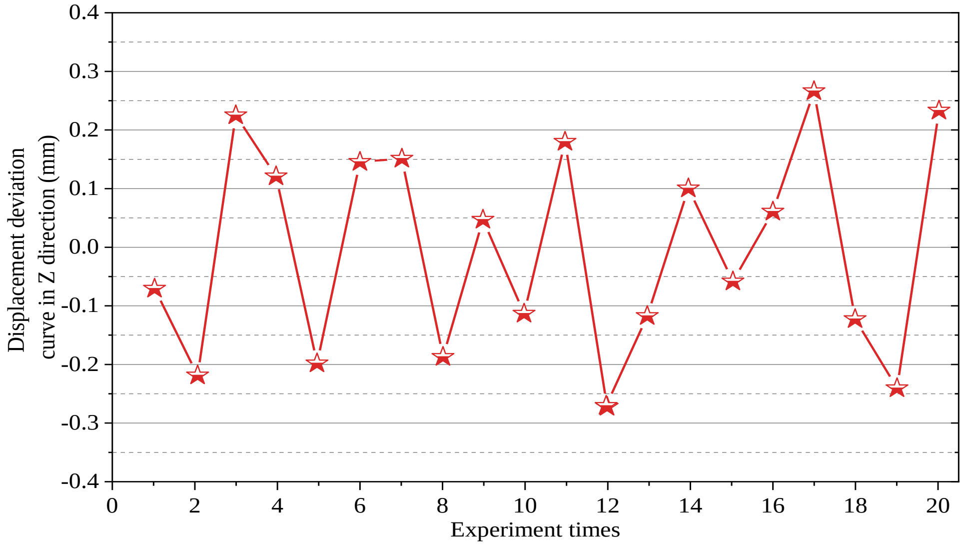
The manual method was first used in the experiment to guide the robot in accurately installing the glass and recording the robot flange pose data. Then, the process described above was used to adjust the robot flange automatically to perform the installation action under the guidance of the laser sensor and the camera. Here, 'posture' refers to the position and/or orientation of the side window glass relative to the robot or assembly platform. The robot flange pose data was also recorded. Finally, the data obtained through automatic guidance were compared with the flange pose data obtained by the manual method. Glass installation experiments on side windows with different postures were performed 20 times. The displacement deviation and rotation differences between automatic and manual assembly are illustrated in Figs. 5 and 6, respectively.



(a) Displacement deviation curve in X direction

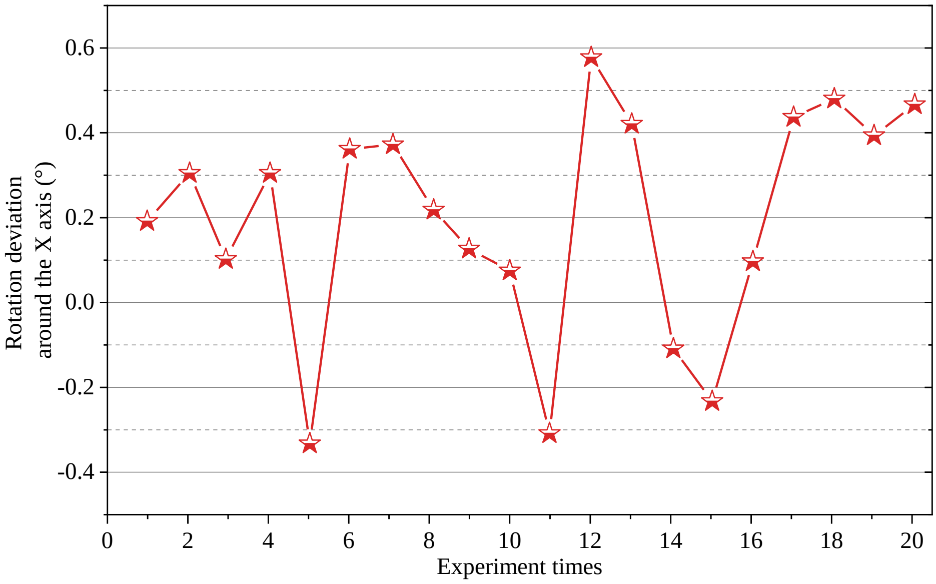


(b) Displacement deviation curve in the Y direction

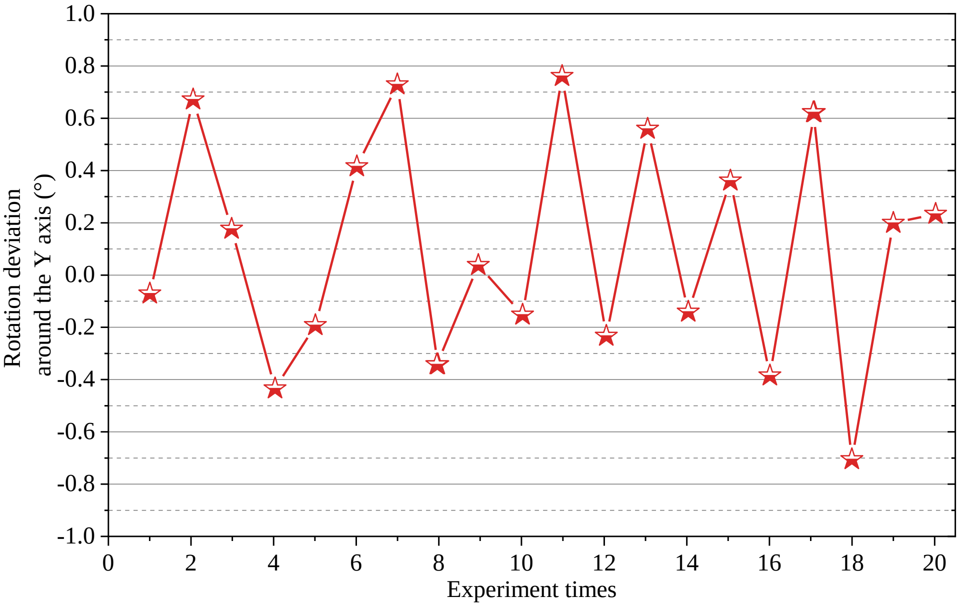


(c) Displacement deviation curve in Z direction

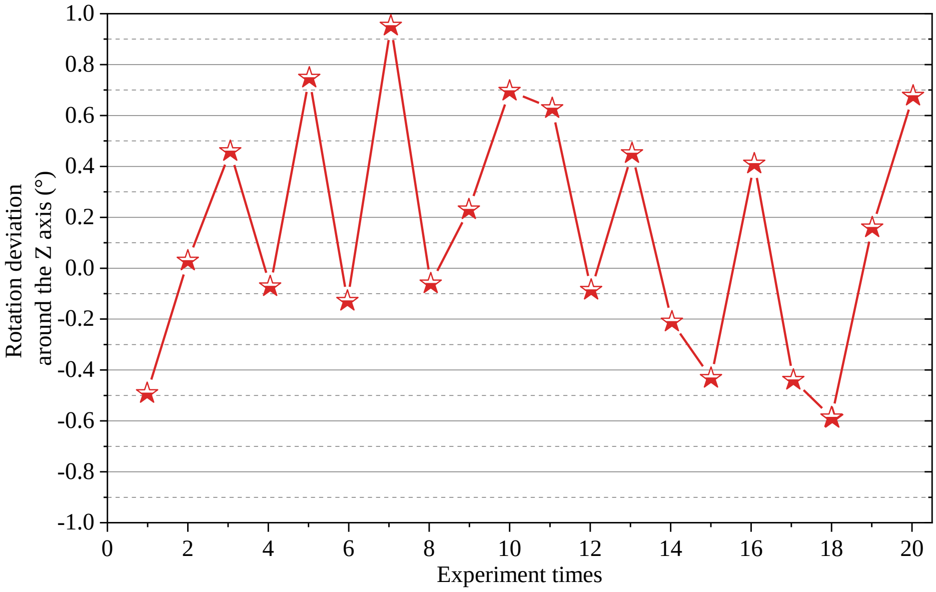
Figure 5: Displacement deviation curve



(a) Plot of rotation deviation around the X-axis



(b) Rotation deviation curve around the Y axis



(c) Plot of rotation deviation around the Z-axis

Figure 6: Rotational deviation curve

Figs. 5 and 6 show that in 20 experiments, the maximum displacement deviation in the X direction is about ±0.4 mm, the maximum displacement deviations in the Y and Z directions are about ±0.3 mm, the maximum rotation deviation of the X axis is about 0.6°, and the minimum rotation deviation of the Y axis is about 0.8°. All deviations meet the requirements of ±1° attitude correction and ±0.5 mm displacement accuracy, showing good reliability. Stereoscopic vision technology also faces some challenges. For example, the calibration and calibration process of a camera is relatively complex, requiring high-precision equipment and technical support. In addition, stereoscopic vision is sensitive to lighting conditions and environmental noise, which may lead to a decrease in image quality and a loss of computational accuracy. Therefore, when choosing stereoscopic vision as an alternative solution, we need to balance the relationship between its high accuracy and complexity. In addition to LDS, we can also consider using additional laser measurement systems to improve assembly accuracy. For example, a 3D laser scanner can be used to comprehensively scan the assembly area and generate accurate 3D point cloud data. These data can be used to generate a digital model of the assembly area and compare it with the expected model to calculate assembly deviations and make corrections.

6 Conclusion

It proposes an integrated assembly method guided by LDSs and monocular vision for the automatic assembly of train-side window glass. The core of the proposed approach is that, from the laser-projected point data, the plane equation of the window is derived, and its NV is used for adjustment of the robot's end flange. The camera attached to the robot shares a parallel plane with the window pane, and the distance between the two planes is adjusted to achieve accurate visual alignment. The monocular vision system identifies edge features of the side window to estimate the precise pose, which guides the robot in completing the glass installation. Experimental results have shown that this assembly approach consistently meets the required pose accuracy for glass installation. It offers the benefits of being inexpensive, having a simple structure, and providing high precision in positioning and orientation calculation, making it suitable for the automatic detection, positioning, guidance, and assembly of other large-scale planar workpieces. This method is mainly suitable for automatic assembly of large flat workpieces. In the previously proposed integrated method for automatic assembly of train side window glass based on a laser ranging system and monocular vision, this article emphasizes the close interaction between system calibration and assembly accuracy. System calibration is a crucial step in ensuring assembly accuracy, as it directly affects the accuracy of laser projection point data and the recognition ability of monocular vision systems for edge features. The calibration of the monocular vision system includes both intrinsic and extrinsic calibration of the camera. Internal calibration ensures that the geometric characteristics of camera imaging, such as focal length and distortion, are accurately recorded, while external calibration criteria ensure the accurate correspondence between the camera and the world coordinate system. Through calibration, the recognition accuracy of the monocular vision system for side window edge features can be improved, thereby more accurately estimating the posture of the glass and guiding the robot to complete precise installation.

However, in practical industrial applications, workpieces may have more complex geometric shapes. How to adjust this method to handle these complex shapes and ensure the accuracy and reliability of assembly is a challenging task. The monocular vision system relies on the recognition of edge features to estimate accurate poses. However, in some cases, such as when the edge of the side window is obstructed or severely worn, feature recognition may become difficult. Prospects include adapting the method to handle complex geometries or materials, applying advanced image processing techniques for better feature recognition, and developing fully autonomous assembly lines, particularly in the automotive and aerospace industries, where similar demands for high accuracy and efficiency exist. We will explore the application effects of this method on different types of large planar workpieces, and how to improve the adaptability and robustness of the method by improving the algorithm or adding sensors.

Nomenclature

Abbreviations		f	Objective function to be minimized
CDW	Construction and Demolition Waste	F	Function constructed
CFD	Computational Fluid Dynamics	H	Transformation matrix
CNN	Convolutional Neural Network	l	Laser line’s unit direction vector coordinate matrix
LES	Large Eddy Simulation	P	Coordinate matrix
LM	Levenberg-Marquardt method	r	Side window plane
ORB	Oriented FAST and Rotated BRIEF	R	Coefficient matrix
PnP	Perspective-n-Point	s	Proportional coefficient
PVB	Polyvinyl Butyral	x	Components of the ith

SHM	Structural Health Monitoring	Greek symbol	
SLAM	Simultaneous Localization and Mapping	α	Rotation angles
WRMSD	Work-Related Musculoskeletal Disorder	β	Rotation angles
symbols		γ	Rotation angles
A	Components	λ	Lagrange multiplier
B	Components	**subscripts**	
C	Components	C	the camera coordinate system
D	Distance	H	the robot end flange coordinate system
$\mathrm{d}_{\mathrm{xyi}}$	Euclidean distance	i	Index for each laser point
$\mathrm{d}_{\mathrm{uvi}}$	Euclidean distance	T	Transformation from one coordinate system to another
d	Distance measured		

Funding

This work was sponsored in part by the Jiangsu Qing Lan Project, the Wuxi Industrial AGV Technology Application and Promotion Public Service Platform Project (WX03-07D0337-052300-07).

Competing interests

No competing interests have been disclosed by the authors.

Authorship contribution statement

Xiaobing Cao: Project administration, Supervision, Conceptualization, Writing-Original draft preparation.
Yicen Xu: Language review.
Yonghong Yao: Methodology.
Sheng Chen: Validation.

Data availability

Available upon request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author atatement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

References

[1] B.-J. You, Y. S. Oh, and Z. Bien, "A vision system for an automatic assembly machine of electronic components," *IEEE Transactions on Industrial Electronics*, 37(5): pp. 349–357, 1990. DOI: 10.1109/41.103429

[2] Y. Ma, W. Zhu, and Y. Zhou, "Automatic grasping control of mobile robot based on monocular vision," *The International Journal of Advanced Manufacturing Technology*, 121(3): pp. 1785–1798, 2022. https://doi.org/10.1007/s00170-022-09438-z

[3] H. A. O. Renjie, W. Zhongyu, and L. I. Yaru, "Error analysis method for monocular vision pose measurement system," *Journal of Applied Optics*, 40(1): pp. 79–85, 2019. DOI: 10.5768/JAO201940.0103001

[4] Y. Ma, X. Liu, J. Zhang, D. Xu, D. Zhang, and W. Wu, "Robotic grasping and alignment for small size components assembly based on visual servoing," *The International Journal of Advanced Manufacturing Technology*, 106(1), pp. 4827–4843, 2020. https://doi.org/10.1007/s00170-019-04800-0

[5] W.-C. Chang, "Robotic assembly of smartphone back shells with eye-in-hand visual servoing," *Robot Comput Integr Manuf*, 50(1), pp. 102–113, 2018. https://doi.org/10.1016/j.rcim.2017.09.010

[6] Su, J., Shen, L., Peng, Z., & Qu, X. (2024). Design of a robot system for reorienting and assembling irregular parts. International Journal of Advanced Robotic Systems, 21(2), 17298806241229271. DOI: 10.1177/17298806241229271

[7] L. Jin-zhou, Z. Ling-bin, and Y. Nan, "Research on industrial robot alignment technique with monocular vision," *Optics & Precision Engineering*, 26(3): pp. 733–741, 2018. https://doi.org/10.1016/j.measurement.2022.112049

[8] Çapunaman, Ö. B., & Gürsoy, B. (2024). Vision-augmented robotic fabrication (V-aRF): systematic review on contemporary approaches and computational methods in architectural fabrication and assembly using machine vision. Construction Robotics, 8(2), 27. DOI: 10.1007/s41693-024-00139-5

[9] X. Hu, K. Shi, J. Mao, P. Guo, T. Yu, and J. Wang, "Optimization study on the influence of rearview mirror and side window glass on interior noise," *Journal of Vibration and Control*, 30(3–4): pp. 516–530, 2024. https://doi.org/10.1177/10775463231172070

[10] J. Zhang *et al.*, "Method for measuring the membrane penetration of triaxial specimens based on an external circular array of laser displacement sensors," *Measurement*, 217, p. 112905, 2023. https://doi.org/10.1016/j.measurement.2023.112905

[11] Y. Shao, L. Li, J. Li, Q. Li, S. An, and H. Hao, "Monocular vision-based 3D vibration displacement measurement for civil engineering structures," *Eng Struct*, 293, p. 116661, 2023. https://doi.org/10.1016/j.engstruct.2023.116661

[12] J. Chen, W. Lu, L. Yuan, Y. Wu, and F. Xue, "Estimating construction waste truck payload volume using monocular vision," *Resour Conserv Recycl*, 177(1), p. 106013, 2022. https://doi.org/10.1016/j.resconrec.2021.106013

[13] C. Sun, D. Gu, and X. Lu, "Three-dimensional structural displacement measurement using monocular vision and deep learning-based pose estimation," *Mech Syst Signal Process*, 190(1), p. 110141, 2023. https://doi.org/10.1016/j.ymssp.2023.110141

[14] W. K. Sleaman, A. A. Hameed, and A. Jamil, "Monocular vision with deep neural networks for autonomous mobile robots' navigation," *Optik (Stuttg)*, 272(1), p. 170162, 2023. https://doi.org/10.1016/j.ijleo.2022.170162

[15] Y. Sun, X. Wang, Q. Lin, J. Shan, S. Jia, and W. Ye, "A high-accuracy positioning method for mobile robotic grasping with monocular vision and long-distance deviation," *Measurement*, 215(1), p. 112829, 2023. https://doi.org/10.1016/j.measurement.2023.112829

[16] Q. Gao *et al.*, "Monocular vision measurement technology applied in dynamic compaction ramming settlement monitoring," *Measurement*, 216(1), p. 112941, 2023. https://doi.org/10.1016/j.measurement.2023.112941

[17] L. Lou, Y. Li, Q. Zhang, and H. Wei, "Slam and 3d semantic reconstruction based on the fusion of lidar and monocular vision," *Sensors*, 23(3) p. 1502, 2023. https://doi.org/10.3390/s23031502

[18] D. Li, B. Cheng, and K. Wang, "Self-calibrating technique for 3D displacement measurement using monocular vision and planar marker," *Autom Constr*, 159(1), p. 105263, 2024. https://doi.org/10.1016/j.autcon.2023.105263

[19] K.-H. Ahn, M. Na, and J.-B. Song, "Robotic assembly strategy via reinforcement learning based on force and visual information," *Rob Auton Syst*, 164(1), p. 104399, 2023. https://doi.org/10.1016/j.robot.2023.104399

[20] B. Mei, Z. Liang, Y. Xie, Y. Fu, and Y. Yang, "Positioning accuracy enhancement of a robotic assembly system for thin-walled aerostructure assembly," *J Ind Inf Integr*, 35(1), p. 100518, 2023. https://doi.org/10.1016/j.jii.2023.100518

[21] Z. Zhang, Y. Wang, Z. Zhang, L. Wang, H. Huang, and Q. Cao, "A residual reinforcement learning method for robotic assembly using visual and force information," *J Manuf Syst*, 72(1), pp. 245–262, 2024. https://doi.org/10.1016/j.jmsy.2023.11.008

[22] C. Champatiray, M. V. A. R. Bahubalendruni, R. N. Mahapatra, and D. Mishra, "Optimal robotic assembly sequence planning with tool integrated assembly interference matrix," *AI EDAM*, 37, p. e4, 2023. DOI: https://doi.org/10.1017/S0890060422000282

[23] Patel, J., Round, A., de Wijn, R., Vakili, M., Giovanetti, G., Melo, D. F. M. V. E., ... & Abbey, B. (2024). "Real-time analysis of liquid jet sample delivery stability for an X-ray free-electron laser using machine vision". *Applied Crystallography*, 57(6), 1. DOI: 10.1107/S1600576724009853

[24] Zhang, L., Zhou, C., Liu, B., Ding, Y., Ahn, H. J., Chang, S., & Ni, X. (2024). "Real-time machine learning–enhanced hyperspectro-polarimetric imaging via an encoding metasurface". *Science Advances*, 10(36), eadp5192. DOI: 10.1126/sciadv.adp5192

[25] B. Zeng, F. Meng, H. Ding, W. Liu, D. Wu, and G. Wang, "A plane-based hand-eye calibration method for surgical robots," *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, 34(2): pp. 200–207, 2017. https://doi.org/10.7507/1001-5515.201605050

[26] M. Li, Z. Du, X. Ma, W. Dong, and Y. Gao, "A robot hand-eye calibration method of line laser sensor based on 3D reconstruction," *Robot Comput Integr Manuf*, 71(1), p. 102136, 2021. https://doi.org/10.1016/j.rcim.2021.102136

[27] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans Pattern Anal Mach Intell*, 22(11) pp. 1330–1334, 2000. DOI: 10.1109/34.888718

[28] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3 d robotics hand/eye calibration," *IEEE Transactions on robotics and automation*, 5(3): pp. 345–358, 1989. DOI: 10.1109/70.34770

[29] G. Yunlan, C. Xiaojun, and S. Guigang, "A robust method for fitting a plane to point clouds," *Journal of tongji university (natural science)*, 36, (7): pp. 981–984, 2008. DOI: 10.1109/ICCC56324.2022.10065838

[30] Lettera, G., & Natale, C. (2024). An Integrated Architecture for Robotic Assembly and Inspection of a Composite Fuselage Panel with an Industry 5.0 Perspective. Machines, 12(2), 103. DOI: 10.3390/machines12020103

[31] Rawat, R., & Rajavat, A. (2024). Illicit Events Evaluation Using NSGA-2 Algorithms Based on Energy Consumption. Informatica, 48(18), 1.

[32] El Ghazi, M., & Aknin, N. (2024). Optimizing Deep LSTM Model through Hyperparameter Tuning for Sensor-Based Human Activity Recognition in Smart Home. Informatica, 47(10), 1.