WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 6, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v6n1.06
Journal homepage: https://wepub.org/index.php/IJCSIT/index

VISAR: Vision-Based Robotic Arm System for Intelligent Industrial Inspection

Liwei Zeng[1, 2, 3], Zewei Ye[1, 2, 3], Lingling Shen[1, 2, 3, *], Junjian Sun[1, 2, 3], Zhicheng Wang[1, 2, 3], Mingyu Chen[1, 2, 3], Yaqi Cheng[1, 2, 3], Haowen Zheng[1, 2, 3], Qi Dong[1, 2, 3], Xiaojun Qian[1, 2, 3]

[1] School of Artificial Intelligence, Nanjing Normal University, Nanjing, China

[2] School of Computer and Electronic Information, Nanjing Normal University, Nanjing, China

[3] Artificial Intelligence Research Institute, Nanjing Normal University, Nanjing, China

*Corresponding Author: llshen509@163.com

ABSTRACT

VISAR (Vision-based Intelligent System for Automated Robotic Inspection) is an intelligent inspection system that leverages a vision-equipped robotic arm to address key challenges in real-time target detection, dynamic path planning, and precise spatial localization. The system integrates a lightweight YOLOv5-ShuffleNetv2 model for efficient object recognition, combined with Canny edge detector and HSV-based color segmentation for robust target localization. Adaptive path planning is achieved using the A* algorithm, while accurate coordination between the camera and robotic arm is ensured through a nonlinear optimization-based hand-eye calibration. Implemented on the Robot Operating System (ROS) with a QT5-based user interface, the system has been validated through both Gazebo simulations and real-world tests on industrial cabinets. Results demonstrate its high efficiency, accuracy, and reliability in complex environments, showcasing strong potential for advancing automation in industrial inspection and manufacturing applications.

KEYWORDS

Vision Robotic Arm; YOLOv5; ShuffleNetv2; A* Algorithm; Canny edge detector; Hand-eye Calibration; Intelligent Inspection System.

1. INTRODUCTION

With the rapid advancement of artificial intelligence and robotics, vision-equipped robotic arms, a key branch of intelligent robotics, hold significant potential in automated production, intelligent warehousing, and inspection systems. Inspection tasks demand high precision, efficiency, and autonomy, and vision robotic arms are capable of autonomously performing operations in complex environments. They can accurately assess the status of equipment and detect potential issues, thanks to their advanced visual perception systems, flexible mechanical structures, and intelligent control algorithms.

In recent years, advancements in technologies such as computer vision and deep learning have greatly enhanced the autonomous grasping capabilities of vision-equipped robotic arms. However, in complex and dynamic inspection environments, challenges such as target recognition, localization, dynamic obstacle avoidance, and path planning remain to be addressed. An in-depth exploration of these key technologies is crucial for improving inspection efficiency and ensuring equipment safety.

This study aims to develop VISAR (Vision-based Intelligent System for Automated Robotic Inspection), a reliable and intelligent visual robotic arm inspection system designed to address the

challenges mentioned above. The system provides theoretical support and technical guidance for autonomous grasping, promotes technological innovation, and facilitates its widespread application in industrial automation, intelligent manufacturing, and inspection systems. It also contributes to industrial upgrading and economic development.

2. SYSTEM DESIGN AND KEY TECHNOLOGIES

2.1. System Architecture

VISAR is built on ROS, and its overall architecture (as shown in Fig. 1) consists of three main modules: the target recognition and spatial localization module, the motion planning module, and the human-computer interaction module. The system captures visual information through an RGB-D camera and implements motion planning, control, and obstacle avoidance for the robotic arm using MoveIt!. Additionally, the robotic arm's control interface is developed using QT5, which provides users with visualizations of the robotic arm model, the camera feed, target recognition results, and motion control.

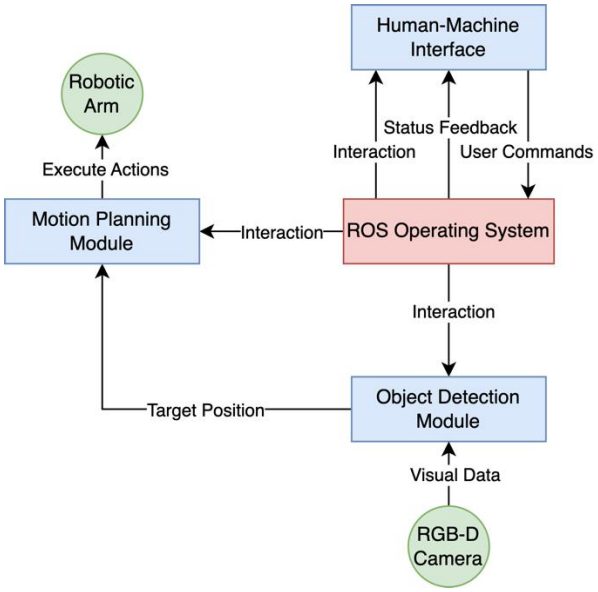

Figure 1. Framework of the VISAR System

2.2. Target Recognition and Spatial Localization Module

2.2.1. YOLOv5 target recognition algorithm based on ShuffleNetv2 optimization

In visual robotic arm inspection tasks, accurate and rapid recognition of target objects or equipment status is essential to ensuring inspection quality. Although the traditional YOLOv5 algorithm performs well in object detection, its high computational complexity and large number of parameters limit its applicability in resource-constrained environments, such as the robotic arm inspection scenario presented in this study [1]. To enhance computational efficiency without compromising detection accuracy, this study adopts a lightweight optimization strategy by replacing the Backbone layer of YOLOv5 with the ShuffleNetv2 architecture, resulting in a more efficient model design.

ShuffleNetv2 is an efficient convolutional neural network architecture optimized for mobile devices and edge computing. Its core design aims to reduce computational complexity while maintaining feature diversity through channel shuffling operations and group convolution. Specifically, ShuffleNetv2 combines pointwise convolution with group convolution to significantly reduce the model's computational complexity and parameter count. Additionally, by introducing the channel

mixing operation, ShuffleNetv2 effectively addresses the feature separation issue caused by group convolution, thereby enhancing feature interaction across different channels [2].

In this study, the combination of YOLOv5 and ShuffleNetv2 is primarily reflected in the following aspects: First, ShuffleNetv2 is employed as the Backbone layer of YOLOv5 for extracting deep features from the input image. Figure 2 illustrates the network structure of the optimized YOLOv5-ShuffleNetv2 model. This design not only preserves the multi-scale detection capability of YOLOv5 but also significantly reduces the model's computational complexity. Secondly, the model's inference speed is further enhanced by optimizing the network structure, making it more suitable for real-time inspection tasks.

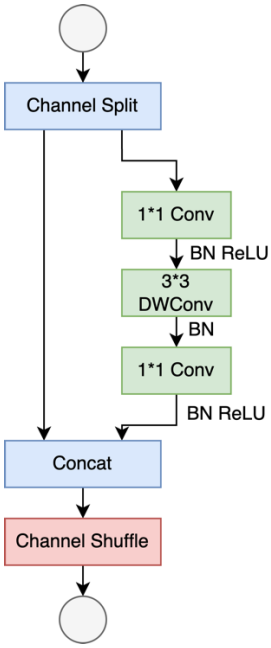

Figure 2. Network Structure

2.2.2. Color recognition algorithm based on Canny edge detector

In VISAR, color recognition technology is employed to identify color markings, such as red warning lights or yellow warning signs on equipment, to alert users to potential safety hazards. The core of color recognition involves converting the RGB color space to the HSV color space, followed by thresholding based on hue and saturation to convert the image into a binary format. This method distinguishes the target color from the background.

The Canny edge detector is then used for target segmentation on the binary image. First, a Gaussian filter is applied to smooth the image and reduce noise. Next, the gradient strength and direction of the pixel points are calculated using the Sobel operator, and the edges are refined through non-maximum suppression. The double thresholding technique is applied to filter the edge points: those with gradient strengths higher than the high threshold are classified as strong edges, those lower than the low threshold are discarded, and those in between are evaluated based on connectivity. Finally, the mean coordinates of the valid pixel points are calculated to determine the center of mass of the target color region [3]. Figure 3 shows the results of edge detection and color segmentation.

This method demonstrates high robustness and accuracy in complex backgrounds, providing a critical spatial reference for visual robotic arm localization and operation, while meeting the real-time and reliability requirements of inspection systems.

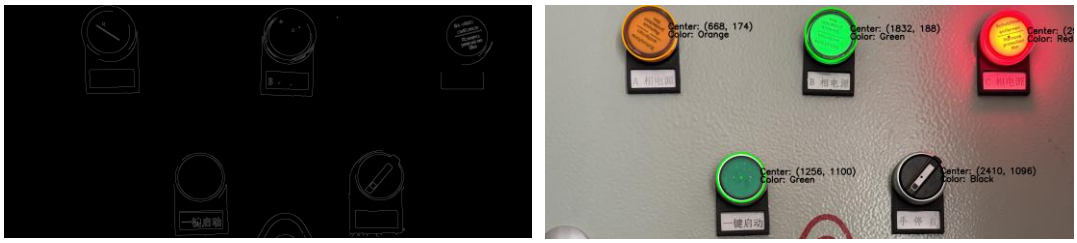

Figure 3. Edge detection and color detection results

2.2.3. Hand-eye calibration algorithm and spatial localization based on nonlinear optimization algorithm

In VISAR, the camera is fixed at the end of the robotic arm, adopting the "eye-in-hand" configuration. The core of hand-eye calibration is to solve the transformation relationship between the camera's coordinate system and the coordinate system of the robotic arm's end effector. By collecting data from the robotic arm at different positions—including the position and orientation of the end effector in the base coordinate system, and the position and orientation of the calibration plate in the camera's coordinate system—a mathematical model is constructed to solve the transformation matrix between the two systems. During this process, a nonlinear optimization algorithm is applied to iteratively optimize the objective function, minimizing the influence of noise and errors, thereby improving calibration accuracy [4].

The mathematical model for hand-eye calibration can be expressed as the following equation:

$$T_{base}^{end} \cdot T_{end}^{cam} = T_{base}^{obj} \cdot T_{obj}^{cam} \tag{1}$$

Where T_{base}^{end} denotes the pose transformation matrix of the end effector of the robotic arm in the base coordinate system, T_{end}^{cam} represents denotes the pose transformation matrix of the camera in the coordinate system of the robotic arm's end effector, T_{base}^{obj} denotes the pose transformation matrix of the calibration plate in the base coordinate system, and T_{obj}^{cam} denotes the pose transformation matrix of the calibration plate in the camera's coordinate system. By collecting multiple sets of robotic arm position and calibration plate data, the following optimization problem can be formulated:

$$min \sum_{i=1}^{N} \left\| T_{base}^{end,i} \cdot T_{end}^{cam} = T_{base}^{obj,i} \cdot T_{obj}^{cam,i} \right\|^2 \tag{2}$$

Iteratively solving the above objective function using a nonlinear optimization algorithm, the optimal T_{end}^{cam} can be obtained, achieving high-precision calibration of the camera and the robotic arm's end effector. Figure 4 illustrates the hand-eye calibration process, including the camera placement and calibration plate setup.

Figure 4. Hand-eye calibration process

2.3. Motion Planning Module

In the inspection process, the robotic arm must efficiently and smoothly move to the specified position. Traditional Rapidly-exploring Random Tree (RRT) algorithm struggles with adaptability in dynamic environments, which are often characterized by factors such as equipment movement, lighting changes, and other variables. RRT algorithms have difficulty adjusting path planning in real time, potentially leading to collisions with the environment or deviations from the target. To address these limitations, this system employs the A* algorithm for robotic arm path planning.

The core concept of the A* algorithm is to guide the search process by evaluating the cost function of each node. The cost function consists of two components: the actual cost from the starting point to the current node (g(n)) and the estimated cost from the current node to the goal node (h(n)). By selecting the node with the smallest total cost for expansion, the A* algorithm efficiently finds an optimal path from the start to the goal. In robotic arm path planning, the A* algorithm effectively avoids issues such as path non-smoothness that can arise from randomness, while also adapting to changes in a dynamic environment.

In practice, the workspace of the robotic arm is first modeled as a grid graph, with the arm's motion parameters mapped onto the grid nodes. By designing a suitable heuristic function, the A* algorithm can quickly identify the optimal path in a complex environment. Compared to traditional RRT algorithms, the path generated by A* is smoother and more computationally efficient, greatly enhancing the robotic arm's motion planning capabilities in dynamic settings[5].

2.4. Interaction Module

VISAR is built on QT5 and features a human-computer interface (Fig. 5), offering functionalities such as user login, robotic arm monitoring and control, camera screen monitoring, and real-time target detection screen monitoring. This interface allows the operator to intuitively monitor the robotic arm's operation status and manually intervene when necessary, enhancing both the system's usability and reliability.

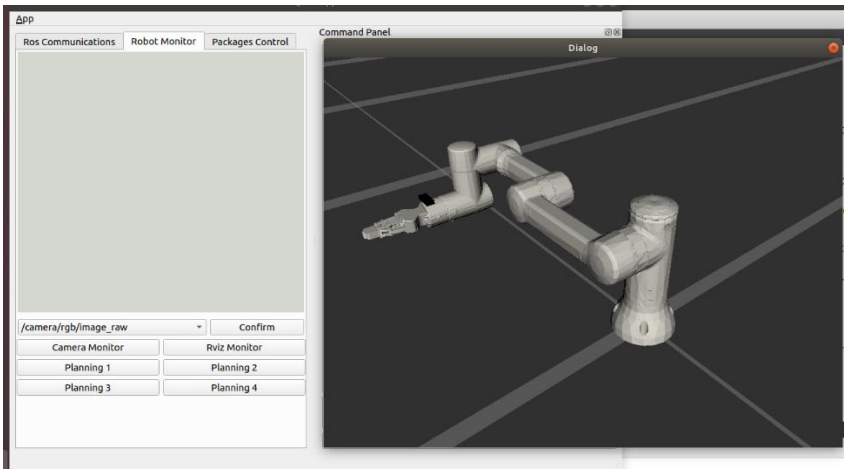

Figure 5. Human-computer interaction interface

3. EXPERIMENTAL VALIDATION

3.1. Simulation Analysis

The simulation test leverages the Gazebo simulation environment and the designed human-computer interface to simulate and validate VISAR. Through target recognition, path planning, and system

integration tests within the simulation environment, we comprehensively assessed the system's performance. The experimental results are as follows:

3.1.1. Target Identification Performance

The experimental results (summarized in Table 1 and visualized in Fig. 6) demonstrate that the optimized model performs faster than the original YOLOv5 on the dataset, with only a slight reduction in accuracy. This performance fully meets the dual requirements of real-time operation and accuracy for the inspection system.

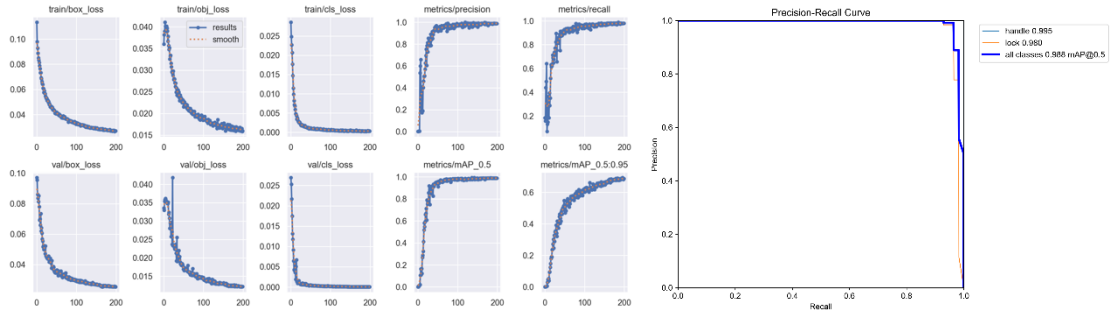

Figure 6. Loss reduction curve and P-R curve

Table 1. Comparison of indicators between the original model and the optimized model

metric	Original YOLOv5s	Optimized Model	Δ Change
mAP@0.5	62.1%	59.6%	↓ 2.5%
Params (M)	7.2	3.4	↓ 52.8%
FLOPs (G)	16.4	4.7	↓ 71.3%
Inference Speed (FPS)	158	208	↑ 31.6%
Model Size (MB)	14.3	6.9	↓ 51.7%
Training Time (h/epoch)	1.3	0.9	↓ 30.8%

3.1.2. Path planning performance

To test the path planning of the robotic arm, obstacles were set up in the simulation environment, and the A* algorithm was used to calculate the specific path. A test scenario, as shown in Fig. 7, was constructed in the simulation. The test requirement was for the robotic arm to bypass two columns and return to its initial position. The robotic arm successfully completed the task.

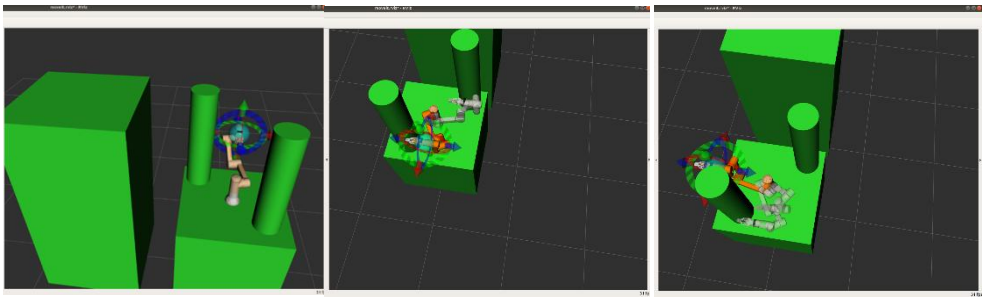

Figure 7. Path planning testing

3.1.3. System integration testing

VISAR successfully completed inspection tasks in the simulation environment, including target detection, path planning, and robotic arm motion control. Through the human-computer interaction interface, the operator can monitor the robotic arm’s operation status in real time and intervene manually. The simulation test verifies the system's functional integrity and performance reliability.

3.2. Live Testing

To verify the performance of VISAR in a real-world environment, we conducted a real machine test. The test scenario was set in an industrial cabinet, where the robotic arm's task was to open the cabinet door and perform an inspection of the equipment inside. The test environment included a standard industrial cabinet, a six-degree-of-freedom robotic arm, an RGB-D camera, and a ROS-based control system.

During the test, VISAR identifies the position of the cabinet door handle using the RGB-D camera and quickly locates it with the YOLOv5 algorithm optimized based on ShuffleNetv2. Through the hand-eye calibration algorithm, VISAR calculates the handle's position in its coordinate system and plans a smooth path to approach it. Once the actuator at the end of the robotic arm grips the handle, the cabinet door is smoothly opened using force feedback control, ensuring safe operation.

Next, VISAR inspects the equipment inside the cabinet. Using the color recognition algorithm, VISAR identifies the red warning light or yellow warning sign on the equipment and, in combination with the Canny edge detector, precisely locates the target area (Fig. 8). Based on the inspection task requirements, VISAR autonomously plans the path and avoids obstacles to complete a comprehensive inspection of the equipment inside the cabinet. The experimental results show that the system can efficiently and accurately complete the inspection task in a real environment, confirming its potential for industrial automation applications (Fig. 8).

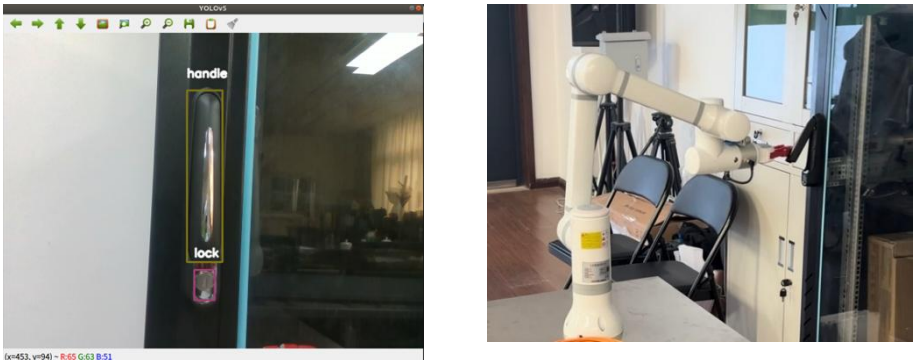

Figure 9. Live testing

4. CONCLUSION

VISAR is designed to significantly enhance the system's real-time performance and robustness. This is achieved by optimizing the YOLOv5 algorithm for improved object detection and adopting the A* algorithm for efficient path planning. Experimental results demonstrate that the system can effectively and accurately carry out inspection tasks within complex environments, showcasing its potential for applications in industrial automation and intelligent manufacturing.

Although the system performs well in both simulation and real-world testing, its design is not yet fully comprehensive. On one hand, while the system operates based on the images captured by the camera, the accuracy of the target recognition and attitude estimation algorithms may be compromised in more complex environments—especially in the presence of occlusions, lighting variations, and other environmental factors. To address these challenges, additional hardware support such as infrared sensors or LiDAR might be required. On the other hand, the end-effector currently used in this system is limited in functionality. To enhance the system's capabilities, various types of end-effectors could be integrated to support a wider range of tasks.

With the growing use of the internet and the rapid advancement of the artificial intelligence industry, the demand for automated inspection systems is expected to rise. Consequently, the number of

applications utilizing vision-guided robotic arms will increase, offering significant prospects for the development of these systems in the future.

ACKNOWLEDGMENTS

Thanks are due to the Institute of Artificial Intelligence Research at Nanjing Normal University for providing the experimental equipment and technical guidance, which laid a solid foundation for the development of this system. Gratitude is also extended to the Huai'an Science and Technology Program Frontier Technology R&D Project, “Key Technology Development of CV-based Automotive Wiring Harness Intelligent Inspection System” (HAG202416), for the financial support and provision of industry resources, which greatly contributed to the development of multimodal data and hardware.

REFERENCES

[1] Ma Y. Target tracking and detection based on YOLOv5 algorithm [C]. Proceedings of the 5th International Conference on Computing and Data Science. Haide College, Ocean University of China, 2023:418-428.

[2] Ma, N, Zhang, X, Zheng, H, Sun, J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018:116–131.

[3] Durgadevi P, Akilan T, Pradhan A, Shariff AM, Yadav N, Uppal P. Canny edge detection techniques for image segmentation [C]. 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022:986-989. IEEE.

[4] Kang M, Fenglei N, Zhaoyang C, et al. A unified calibration method for robot manipulators: hand-eye parameters, kinematic parameters and TCP position calibration [J]. Robotic Intelligence and Automation, 2024, 44(6):897-909.

[5] Fu X, Huang Z, Zhang G, et al. Research on path planning of mobile robots based on improved A* algorithm [J]. PeerJ. Computer science, 2025, 11e2691.