

Improvement of visual servo system of industrial robot based on sliding mode control and deep reinforcement learning

Yunzhe Zhou

Department of Smeal Business, Pennsylvania State University, PA, USA

Yzz5886@psu.edu

Abstract. Visual servo system is more and more widely used in the field of industrial robots because it allows robots to sense external signals through sensors to convey control commands to themselves and complete tasks. However, the traditional visual servo has many limitations in the design of controller and image feature extraction, such as insufficient robustness and image extraction accuracy. This research focuses on the optimization of controller and image feature extraction, which can improve the overall performance and autonomy of the system by combining sliding mode control and Convolutional Neural Network (CNN). Sliding mode control performs well in terms of robustness and response speed, while CNN has excellent ability for image feature extraction. The research results show that the combination of the two and the visual servo has better performance in a variety of application scenarios, so this is also the development direction that industrial robots can adopt in the future.

Keywords: Visual servo system, sliding mode control, industrial robot.

1. Introduction

With the development of automation and robot technology in the industrial field, industrial robots have made great breakthroughs in the 21st century. The first industrial robot, designed by Griffith P. Taylor in 1935, set a fixed program to carry goods, and since then, the technology has developed into a variety of highly intelligent multi-functional robots [1]. The subsequent Unimation robot, designed by George and Joseph in 1956, using cash's servo motor and sensor technology, with more than six degrees of freedom, and with networking technology, capable of remote monitoring and collaborative work, was the world's first programmable industrial robot, marking the beginning of the era of robot automation [2]. After the 1980s, multi-axis robots became the standard for industrial robots, improving their usability. The introduction of offline programming technology made it possible for robots to be programmed and tested in a virtual environment, reducing debugging time in actual production. In the 21st century, many production lines began to move toward automation, including robot collaborative industry. Material handling, processing assembly, and product packaging are completed by integrated operations. The widespread application of artificial intelligence algorithms and big data can enable robots to learn and accumulate autonomously and constantly optimize themselves.

In the history of industrial robot development, the visual servo system has been widely used, and visual servo system is a technology that uses visual information to control robot movement, captures images through cameras, processes and extracts image features, and forms control signals according to features through error calculation, so as to adjust the position and attitude of the robot itself [3]. Visual

servo (VS) is mainly divided into image-based IBVS and position-based PBVS [3]. The former directly calculates image feature points, while the latter uses image feature points to calculate three-dimensional parameters for attitude estimation and then control. Since the system includes both perception and control aspects, each link can improve the performance of the visual servo by introducing more efficient and advanced algorithms and technologies.

As an important system of robots in the industrial field, the visual servo system brings great development for the autonomy and independence of robots in completing tasks, and effectively reduces labor costs and time costs. It is widely used in the industrial field and has a variety of application scenarios.

The visual servo system can accurately guide the robot to carry out complex assembly tasks. For example, in automobile manufacturing, the visual servo system can help the robot to identify and select small parts in the parts library, and then complete the installation, such as screws, metal blocks, baffles, etc. This greatly reduces the labor cost of the factory and improves the production efficiency of the production line.

In the welding process, the visual servo system can monitor the welding position, the completion of welding and the welding quality in real time. Through the images provided by the camera, the robot uses the miniature coordinates of each point on the surface of the welding object to make connection and analyzes the precise welding path to ensure the same weld accuracy.

Visual servo also has applications in product quality detection in industrial production. Through the extraction of object feature points by robots, it can carry out multiple magnification to find the missing feature points on the structure of the object, so as to identify plane scratches, dimensional deviations and other errors.

In addition, there are many improvement schemes based on visual servo system, which aim to improve the efficiency and accuracy of robots to complete tasks through the visual servo system. For example, by combining the advantages of position-based visual servo and image-based visual servo control methods, 2.5D VS establishes a connection between the image plane and three-dimensional space, builds a hybrid Jacobian matrix, and adjusts errors from the depth direction of the image [4]. The backward camera phenomenon and singularity problem of IBVS system are overcome, thus ensuring the stability of the control process.

2. Limitations analysis of traditional visual servo systems

There are still limitations in the traditional visual servo system. The shortcomings and improvements in achieving autonomous tasks can be analyzed in this section.

2.1. Robustness

The traditional visual servo system often shows insufficient robustness when dealing with different environmental changes and external interference, in other words, the robot is affected by the external conditions of the system, resulting in reduced operating accuracy, or even severe control program error. The nonlinear control of visual servo is mainly completed by the closed-loop motion trajectory error calculation, and the condition interference outside the system will cause the system to deviate from the original path, and it needs manual parameter adjustment to restart the operation [5]. External interference is often manifested in noise, light, and obstruction problems. First of all, the electromagnetic interference generated by industrial equipment will affect the electronic components of the camera, resulting in distortion of the visual reception signal, and the vibration caused by noise will cause small changes in the position of the camera to continue to swing, affecting the extraction stability. For image extraction, noise will lead to different intensities of image feature edge construction, and some weak intensity features will be missed by the receiver, resulting in errors in image modeling, and eventually lead to operation errors.

The light and obstacle occlusion will also have a significant impact, strong light will lead to the camera lens's dazzling effect, affecting the image clarity. The light source into the camera is very obvious, thus affecting the judgment of the picture information, coupled with the vision sensor pixels

under strong light saturation phenomenon, resulting in the image blank, loss of detail information and finally affecting the image quality affection. The occlusion of obstacles will directly hinder the receiving line of sight, failing to obtain correct and complete information, or the extraction of obstacle features as the characteristics of the target object, which will cause large errors in the operation and deviate from the control trajectory.

2.2. Response time

Response time is also one of the limitations, meaning that the robot takes a long time from receiving the control signal to actually producing the output operation, which affects the efficiency and accuracy of the task. The control algorithm of the visual servo system needs to calculate the control instructions according to the image processing results, and the complexity of the control algorithm will directly affect the response time. The traditional visual servo system mainly uses PID controller. PID control is a classic feedback control algorithm, which is a common system in industrial automation [6]. Error correction is carried out by three parameters, proportion, integral and differential, so as to achieve a control effect [6]. First, proportional control generates a control signal according to the consistent error and changes the control response speed by adjusting the proportional gain, while the integral control generates a control signal according to a large number of accumulated values of errors to eliminate the steady-state error [6]. As the adjustment of the integral time constant and integral gain requires the accumulation of signals, the response speed will be too slow. Finally, differential control generates a control signal through the error change rate, so as to predict the future error change trend. However, the controller gain of PID is fixed, and the error accumulation in actual operation is also non-linear, so PID control needs to spend a certain amount of time in the error calibration process.

2.3. Image extraction accuracy

For the visual sensing part of the traditional visual servo system, most of the edge detection method is used, which is also the basic technology of image processing, for detecting the object boundary in the image, such as Sobel operator. The Sobel operator first grays the color image and uses two filters to calculate the gradients G_x and G_y in two directions respectively, so as to carry out convolution operations and judge the edge position of the object according to the gradient size and direction [7]. When the gradient size exceeds the set threshold, the layer is regarded as the edge layer. Although the calculation steps of Sobel operator are simple and the real-time processing ability is strong, its limitation and the volume of the convolution kernel are too large to determine the fine edge details, plus there are only two detection directions, and it also has a large error for variable objects. In addition to Sobel operator, Canny algorithm is also a common detection method. It smooths the image through the two-dimensional formula of the Gaussian filter, then calculates the gradients in both directions, then refines the edges using non maxima suppression, and finally sets the high and low thresholds for classification to facilitate the edges to be connected along the gradient size [7]. Canny has high robustness and stability, but the calculation steps are too complex, resulting in poor real-time performance, which is not conducive to industrial dynamic tasks.

Based on the limitations of the above visual servo, there are still considerable obstacles for the traditional visual servo to achieve full automation. In order to improve production efficiency and product quality, reduce unnecessary labor costs, and promote the progress of the overall technology of industrial robots, it is necessary to improve the visual servo system combined with sliding mode control and deep reinforcement learning technology.

3. Feasibility of optimization scheme

Sliding mode control is a nonlinear robust control method, which has the characteristics of fast response and strong robustness, and can effectively deal with the uncertainty and external interference of the system. It has two major design steps. The first is to design sliding surface, $s = ce + \dot{e}$ which is also the ideal motion state of the system; the second is to design control law $\dot{s} = -\epsilon \text{sgn}(s)$, $\epsilon > 0$, also called function switch, which is used to control the target object to approach the sliding mode surface

continuously [8, 9]. After the object reaches the sliding surface, only the control law will affect the motion trajectory of the object, which is not affected by external factors [8]. So that's why it's robust.

In recent years, some scholars have also carried out detailed studies on this aspect. For example, M. Parsapour et al used the robust estimator based on the Untraceless Kalman Observer (UKO) cascade and Kalman Filter (KF) to infer the physical parameters and placement posture of the target object, and established a sliding mode control model [10]. In this paper, the author uses Lyapunov theory to analyze the stability of the closed-loop control system [11]. By selecting the appropriate Lyapunov function $V = 1/2 S^T S$ and calculating its function derivative, it is proved that it is always negative under the action of the control signal, that is, the state of the system tends [11]. In the experiment, the authors used a 5-DOF RV robot. For the first experiment, the visual adjustment experiment, the target object does not move. The system needs to adjust the position and attitude of the end effector to make it reach the desired state, and control the signal through different types of switching functions. The results show that the controller converges the trajectory to the sliding surface within 0.2 seconds, and makes the position and velocity errors of the end-effector close to zero within 0.8 seconds [10]. In addition, the use of a sliding mode controller with a saturation function can effectively reduce the chattering phenomenon of sliding mode control. In the second experiment, the verification of the overall performance, the target object is independently moved along the X, Y, Z, A and B directions to verify the tracking performance of the system under the condition of target movement. The results show that the controller can respond quickly to the target movement. Through these experiments, it is also proved that sliding mode control can help the visual servo system overcome the shortcomings of slow response and poor robustness.

In addition, visual servo can also be combined with Convolutional Neural Network (CNN) to improve the system's performance of receiving information, that is, the ability to extract images. The CNN is a deep learning model with powerful image feature extraction ability [12]. CNN extracts pixel values through the convolutional kernel to generate a rough model and then extracts the image by dropout and pooling to reduce dimension and parameters and retain key features, aiming at improving computing efficiency [12]. Then the number of pooling is activated by the ruler function, and finally the matrix transformation of the full connection layer is carried out to convert countless input parameters into a single control signal through the function [12]. Therefore, CNN can process complex image features and extract them with higher precision. In an article based on the detection system combined with isolation security and deep learning, the author compared the accuracy of CNN and SVM, and the results showed that CNN always maintained a high and stable accuracy in the whole training process (shown in the Figure 1) [13].

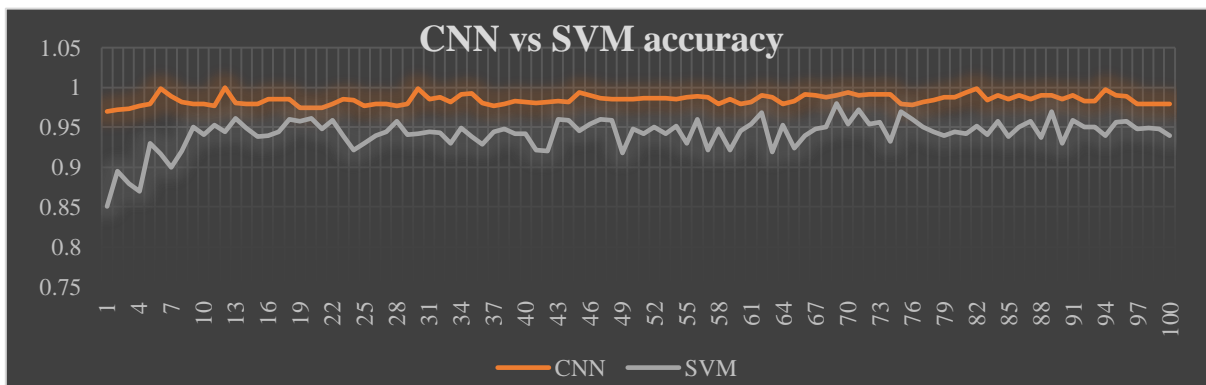


Figure 1. Experiment result accuracy performance of CNN and SVM [11].

4. Proposal of optimization program

Based on the two available improvements, visual servo systems can be used in different situations. Since the sliding mode control and CNN cannot meet all situations in terms of execution efficiency, implementation cost, and operational complexity, engineers need to carry out scenario analysis and choose the appropriate optimization scheme under the specific task.

4.1. Sliding mode control system

In the environment with high real-time requirements and relatively simple tasks, only the Sliding Mode Control (SMC) system can meet the requirements. Because the system can keep the object moving steadily on the sliding surface, to achieve fast response and complete the task with high robustness, this method is very suitable for the task requiring the robot to perform low latency and high real-time tasks. For example, complete the robot's production line assembly tasks and sorting tasks. In the task, the robot does not need to show high-intensity image extraction capability but only needs to lock the target object and perform the corresponding operation. In addition, such tasks require high stability, and if the system is vulnerable to external interference, it needs to be manually monitored and adjusted in real time.

4.2. Convolutional neural network

In specific complex visual tasks, it is suitable to use CNN alone for feature extraction and processing, which can improve the accuracy of image recognition. The most common example is the slim sensor for autonomous vehicles, where the vehicle needs to obtain real-time images of the road environment through the camera and analyze the images to identify obstacles such as pedestrians and other vehicles and convey danger signals to the car. In addition, this solution is also suitable for product testing, for complex structural frames, each corner and geometric slope need to ensure high precision so that there are no accidents in application.

4.3. Combination of SMC and CNN

In the task of high precision visual servo control in complex dynamic environment, the optimization scheme combining SMC and CNN is needed. The system not only needs high-precision image extraction but also needs to resist the interference of changing environment. For example, in the robot welding process, the system needs to complete the operation with high precision in the high-intensity noise environment generated by the welding. The CNN system provides virtually error-free details of the object's appearance such as edges and dioramas. The sliding mode control can make the system stably execute and input instructions on the sliding surface, which perfectly integrates the points of both sides (shown in Figure 2).

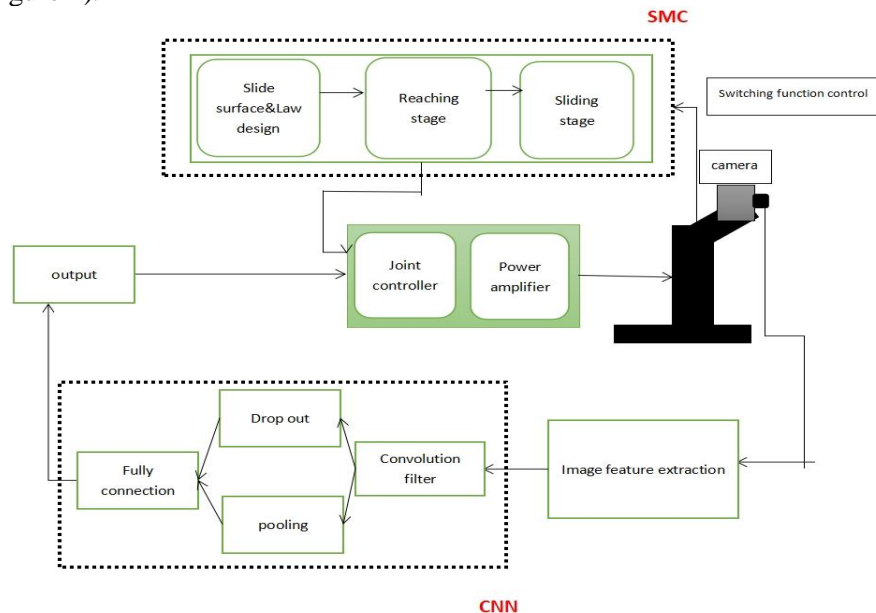


Figure 2. Comprehensive diagram.

The future robot vision servo system will pay more attention to the integration of multiple sensors, one purpose is to reduce the acceptance pressure of a single camera, the other is to analyze the target image from more angles. The vision sensor will be combined with lidar, ultrasonic, thermal imaging,

etc. The system will obtain more comprehensive environmental information and enhance its autonomous decision-making ability. With the continuous progress in the field of AI, industrial robots will pay more attention to intelligence. Deep learning will be a big picture of future research, and the development of multi-modal sensing will provide more powerful computing power for robots.

5. Conclusion

In this study, the optimization of the visual servo system is deeply discussed. By combining sliding mode control and CNN, the system robustness, response speed and image processing accuracy are effectively improved, which is of great help to the improvement of traditional visual servo systems.

The sliding surface and control law are designed to make the system move stably in the calibrated trajectory. The strong robustness to illumination changes, noise and occludes will be made and the correct path processing under uncertain conditions will be realized. By using the process of convolutional extraction and pooling as well as dropout, the CNN help system can recognize and extract the full range of pixels of the image, which is more precise and can extract more parameters than the traditional edge detection and corner detection. In the future, with the continuous development of artificial intelligence and deep learning technology, the industrial robot visual servo system will be further intelligent and efficient. The application of a multi-sensor fusion technology robot will expand the application range of the robot, improve its adaptability and autonomy in complex environments, and achieve a high degree of automation.

References

- [1] Grace, J. (1937). environment and nation. Griffith Taylor. The Journal of Geology, 45(5), 571–572. <https://doi.org/10.1086/624573>
- [2] Gasparetto, A., & Scalera, L. (2019). From the unimate to the delta robot: the early decades of industrial robotics. In *Explorations in the History and Heritage of Machines and Mechanisms: Proceedings of the 2018 HMM IFToMM Symposium on History of Machines and Mechanisms* (pp. 284-295). Springer International Publishing.
- [3] Cong, V.D., & Hanh, L.D. (2023). A review and performance comparison of visual servoing controls. *International Journal of Intelligent Robotics and Applications*, 7, 65-90.
- [4] Zhang, H., Li, M., Ma, S., Jiang, H., & Wang, H. (2021). Recent advances on robot visual servo control methods. *Recent Patents on Mechanical Engineering*, 14(3), 298-312.
- [5] Grimbale, M. J., & Majecki, P. (2020). *Nonlinear Industrial Control Systems*. Springer London.
- [6] Borase, R. P., Maghade, D. K., Sondkar, S. Y., & Pawar, S. N. (2021). A review of PID control, tuning methods and applications. *International Journal of Dynamics and Control*, 9, 818-827.
- [7] Sun, R., Lei, T., Chen, Q., Wang, Z., Du, X., Zhao, W., & Nandi, A. K. (2022). Survey of image edge detection. *Frontiers in Signal Processing*, 2, 826967.
- [8] Zhang, X. (2022). SMC for nonlinear systems with mismatched uncertainty using Lyapunov-function integral sliding mode. *International Journal of Control*, 95(10), 2710-2725.
- [9] Gambhire, S. J., Kishore, D. R., Londhe, P. S., & Pawar, S. N. (2021). Review of sliding mode based control techniques for control system applications. *International Journal of dynamics and control*, 9(1), 363-378.
- [10] Utkin, V., Poznyak, A., Orlov, Y. V., & Polyakov, A. (2020). *Road map for sliding mode control design*. Berlin/Heidelberg, Germany: Springer International Publishing.
- [11] Parsapour, M., RayatDoost, S., & Taghirad, H. D. (2013, February). Position-based sliding mode control for visual servoing system. In *2013 First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)* (pp. 337-342). IEEE.
- [12] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12), 6999-7019.

- [13] Ramirez, A. G., Lara, C., Betev, L., Bilanovic, D., & Kebschull, U. (2018). Arhuaco: Deep learning and isolation-based security for distributed high-throughput computing. arXiv preprint arXiv:1801.04179.