Progress And Comprehensive Analysis of Intelligent Grasping Technology for Robotic Arms Based on Computer Vision

Junjie Shen *

College of Mechanical and Energy Engineering, Zhejiang University of Science and Technology, Hangzhou, Zhejiang Province, 310015, China

* Corresponding Author Email: 1210114032.zust.edu.cn

Abstract. With the continuous progress of technology and the increasing demand, industrial robotic arms have gradually developed into an important production tool. The current research goal for the development and application of industrial robotic arms is to improve production efficiency, reduce costs, improve quality and safety, in order to achieve industrial automation and intelligence. The application of computer vision technology can improve the autonomy, flexibility, and intelligence level of industrial robotic arms, enabling them to play a better role in industrial production. This article first introduces several different visual systems, including Eye-in-hand and Eye-to-hand (monocular, binocular vision systems, heterogeneous perception systems) for robotic arms. Then, three algorithms for feature extraction of robotic arm grasping targets were introduced and explained: SIFT, SURF, and ORB. Finally, summarize the analysis of workpiece target feature extraction and matching techniques applied to industrial robotic arms. This article summarizes several technical research directions for improving the performance of industrial robotic arms. It has reference significance for improving the production efficiency and quality of robotic arms, achieving automation and intelligence, adapting to complex environments, and expanding the application fields of robotic arms. In the future, intelligent recognition and grasping control of robotic arms may be further integrated with other technologies such as artificial intelligence, machine learning, sensor technology, etc., to enhance the intelligence level of the system. The robotic arm can adapt better to different tasks and environments through continuous learning and adaptive adjustment, improving its versatility and flexibility.

Keywords: Industrial robotic arm; object detection; grasping prediction; SURF algorithm.

1. Introduction

With the development of technology, industrial robot technology is gradually becoming stronger. Today's industrial robotic arms can quickly and accurately complete repetitive tasks, and can work continuously 24 hours a day, greatly improving production efficiency and shortening production cycles. They are widely used in various industries, reducing human labor for various tasks and greatly improving work efficiency and stability [1]. The operation accuracy of the robotic arm is high, which can reduce the impact of human factors on product quality, improve product consistency and stability. Firstly, the use of robotic arms can replace some manual operations, especially in some dangerous and harsh working environments, which can reduce labor costs and worker workload. Secondly, the robotic arm can quickly adapt to different production needs and product changes, achieving flexibility and customization in production. Industrial robotic arms are an important component of building automated production lines and intelligent factories. Combining them with other automation equipment can help improve the efficiency and intelligence level of the entire production system. Furthermore, the widespread application of advanced technologies such as robotic arms can help enhance the competitiveness of enterprises, promote industrial upgrading, and adapt to market development needs.

Traditional robotic arms can only operate through a single or several simple modes and cannot meet more complex and variable working conditions. For example, if the size, shape, and position of the object that needs to be grasped are different, the robotic arm cannot accurately determine the grasping or lose the judgment target. Generally speaking, for simple target recognition and positioning tasks, the accuracy of a robotic arm can reach over 90%. But in some complex industrial

scenarios, the recognition and positioning accuracy of the robotic arm may be slightly lower, reaching a range of 80% to 90%. So currently, for high demand situations, the grasping accuracy of robotic arms is relatively low, and new computer vision technologies need to be improved. The use of high-precision sensors and advanced algorithms can improve the recognition and positioning accuracy of robotic arms. For example, using computer vision technology can provide more accurate information on the position and posture of target objects. Make the robotic arm more intelligent, efficient, and stable in operation.

This article reviews several technical research directions for improving the performance of industrial robotic arms. Introduced several different visual systems applied to industrial robotic arms, computer algorithms for workpiece feature extraction, and their respective advantages and disadvantages. The research involves the integration of multiple fields of technology, such as computer vision, robotics, control theory, etc., which helps to promote innovation and development of related technologies. Research is of great significance for improving production efficiency, ensuring work safety, promoting technological innovation, and industrial upgrading. It will provide strong support for the development of automation and intelligence in various fields.

This article first introduces the functions and advantages and disadvantages of the Eye in hand system and Eye to hand system (monocular, binocular vision system, heterogeneous perception system) applied to robotic arms. Then, the feature extraction and matching algorithms for robotic arm grasping targets, including SIFT, SURF, ORB, and improved SIFT algorithms, were introduced and explained. Finally, the analysis of workpiece target feature extraction and matching techniques applied to industrial robotic arms was summarized.

2. Visual System for Industrial Robotic Arms

Object recognition and grasping technology based on computer vision is a relatively complex technology in industrial robotic arm technology. This technology mainly uses image processing and machine learning technologies to recognize target objects and assist robots in target grasping. At present, research on robotic arm grasping based on computer vision mainly includes two categories: Eye-in-hand and Eye-to-hand [2].

2.1. Eye-in-Hand

Eye in Hand is a combination of the camera and robotic arm of a visual robotic arm. The principle is to install the camera at the end of the robotic arm and determine the position and posture of the camera relative to the robotic arm through hand-eye calibration, in order to achieve precise positioning and grasping of the target object. Hand-eye calibration is one of the key technologies for achieving eye-in-hand applications, which is of great significance for improving the autonomy, flexibility, and accuracy of robots. The accuracy and reliability of calibration directly affect the performance of the system and the quality of task completion, which can help robots or other automation systems accurately understand the relative position and posture relationship between hands and visual sensors. Through hand-eye calibration, the system can accurately match, and associate images or data obtained from visual sensors with hand movements. This enables robots to perceive and process tasks related to hand operations more accurately. Therefore, in practical applications, a careful calibration process and parameter adjustment are usually required to ensure the accuracy and effectiveness of hand-eye calibration [3].

2.2. Eye-to-Hand

Eye-to-hand can achieve remote control of the robotic arm, eliminating the need for operators to directly contact the arm, thus avoiding some dangers and limitations. It can provide more direct visual feedback, making the operation of the robotic arm more precise and flexible. However, there may also be some limitations, such as the interference of robotic arm movements on the eyes outside the hand, or the obstruction of vision in complex environments. In addition, the coordination between the

eyes and the robotic arm also requires precise control systems and algorithm support. So in different application scenarios, having hands outside the eyes is a better choice.

2.2.1 Monocular vision system

There are several different object recognition technologies on the market today. There is a monocular vision based robotic arm grasping control method. The combination of monocular vision system and robotic arm can improve the autonomy, flexibility, and accuracy of the robotic arm, enabling it to better adapt to complex work tasks and environments. However, in practical applications, factors such as the accuracy, frame rate, and image processing capability of the visual system need to be considered to ensure effective integration and collaborative work with the robotic arm.

The adaptability and flexibility of monocular vision systems. It can enable the robotic arm to adapt to different work tasks and scenarios. It can recognize objects of different shapes, sizes, and positions, and can flexibly grasp and operate according to needs, improving the versatility and adaptability of the robotic arm. In human-machine collaboration scenarios, monocular vision systems can help robotic arms understand the intentions and actions of human operators, achieving safer and more efficient collaboration. The robotic arm can make corresponding responses based on visual information and cooperate with human operators to complete tasks. The real-time visual feedback of the monocular vision system enables the robotic arm to quickly respond to environmental changes and make corresponding adjustments. For example, when the position of an object changes, the robotic arm can adjust the grasping strategy or action in a timely manner through visual detection, improving work efficiency and success rate [4].

The field of view of a monocular vision system is usually limited and may not cover the entire area operated by the robotic arm. It can only obtain image information through a single camera and cannot directly measure the depth of an object. This may lead to inaccurate estimation of object distance and spatial position in certain situations, especially without the assistance of other depth sensing technologies. This means that there may be visual blind spots at certain angles or positions, making it impossible to detect certain objects or actions. In complex backgrounds or situations with object occlusion, monocular vision systems may be disturbed, making it difficult to accurately recognize and track target objects, which may lead to misjudgment or loss of targets. Monocular image processing and visual algorithms require a certain amount of computation time, which may cause some delay in real-time control of the robotic arm. In high-speed motion or applications with strict time requirements, faster visual processing capabilities may be required. The accurate operation of a monocular vision system usually requires initialization and calibration to ensure the correct position, posture, and parameters of the camera. Inaccurate calibration or environmental changes may affect the performance of the system. So, to overcome these limitations, it can be combined with other sensor technologies [5].

2.2.2 Multi camera vision system

In the literature on multi vision technology, it is mentioned that in actual visual positioning and grasping experiments, the binocular vision camera obtains the coordinate position of the target and uses the transformation matrix set by hand eye calibration to convert the coordinates of the object in the camera to the coordinate system of the robotic arm. Then, based on the spatial coordinate system of the robotic arm, path planning is carried out to obtain the motion situation of each joint, enabling the control robotic arm to complete the grasping action. Hand eye calibration is the process of determining the coordinate transformation relationship between pixel coordinate systems and spatial robotic arm coordinate systems [6]. The accuracy of calibration seriously affects the accuracy of robotic arm grasping operations. Visual grasping experiments were conducted in different scenarios, especially when the target is too close to the camera. It is necessary to rotate the robotic arm and change the camera's recognition posture to better achieve target recognition. In other scenarios with obstacles, target recognition tasks can also be well completed at a moderate distance. The feasibility of this technology has been verified through a large number of experiments. Compared to monocular

vision systems, monocular vision systems have better depth perception. Through the perspective differences of multiple cameras and the principle of triangulation, a multi camera vision system can more accurately calculate the depth information of objects. This is crucial for the operation, obstacle avoidance, and spatial positioning of robotic arms. At the same time, the multi camera vision system has a more comprehensive field of view, and multiple cameras can provide a wider field of view, reducing blind spots in the field of vision. This helps the robotic arm to better perceive the surrounding environment, detect more objects and actions [7]. Of course, multi camera vision can improve accuracy and reliability. Through the fusion of information from multiple perspectives, the accuracy of object detection, localization, and measurement can be improved. The redundancy of multiple cameras can also increase the reliability of the system and reduce the impact of a single camera failure on the overall system. When an object partially obstructs the field of view, other unobstructed cameras can still provide useful information. Multi camera vision systems can overcome occlusion problems through information fusion and more accurately track and recognize targets. Finally, the multi camera vision system can be flexibly configured and laid out, and the positions and angles of multiple cameras can be flexibly arranged according to specific application requirements to adapt to different robotic arm structures and working environments.

However, multi vision systems also face some challenges and complexity. For example, issues such as camera calibration, image synchronization, and data fusion. Meanwhile, increasing the number of cameras will also bring more data processing and computational burden, so better algorithm support is needed.

2.2.3 RGB-D visual system

RGB-D is a 3D image format where each pixel has four attributes: red, green, blue, and depth. The combination of embedded technology and sensor technology can effectively reduce the amount of deep computing and achieve high-precision positioning, overcoming some of the drawbacks of multi vision technology. The robotic arm can obtain color and depth information of objects through RGB-D sensors, thereby recognizing the shape, position, and posture of objects and achieving precise grasping operations. However, RGB-D sensors also have some drawbacks, such as being relatively expensive and increasing the cost of robotic arm systems. In certain special environments, such as strong light, low light, or surfaces with strong reflections, the accuracy of RGB-D technology may be affected. Processing RGB-D data requires high computational power and complex algorithms, with high requirements for the system. The measurement range of RGB-D sensors is usually limited and may not be able to cover large working areas [8].

3. Workpiece Feature Extraction and Matching Algorithm

In the computer vision-based workpiece position perception task applied to robotic arms, feature extraction of workpiece targets in images and matching at different frames/perspectives is a key technical component. This chapter focuses on summarizing SURF based feature extraction and matching algorithms.

3.1. SURF Algorithm

Scale Invariant Feature Transform (SIFT) is the most classic feature extraction and matching algorithm. It is a feature extraction algorithm used for computer vision and image processing. It is mainly used to detect and describe key points (feature points) in images, and extract feature descriptors with scale and rotation invariance. The SIFT algorithm has been widely applied in the field of computer vision, such as image matching, object recognition, image retrieval, etc. It has good stability and robustness, and can handle image scaling, rotation, lighting changes, and other situations. However, the SIFT algorithm has a high computational complexity and may not be suitable for applications that require high real-time performance. The Speeded Up Robust Features (SURF) algorithm is a feature extraction algorithm based on scale space, which is an improvement of the SIFT algorithm. Therefore, most existing robot recognition technologies use the SURF algorithm.

The main principle of SURF algorithm is to first construct the scale space, using a method called Hessian matrix to detect scale invariant feature points in images. By applying Gaussian blur to images at different scales, a scale space pyramid can be constructed. Next is key point detection, which uses the Hessian matrix to detect extreme points in the image in each scale space, which are considered stable key points. The eigenvalues of the Hessian matrix can be used to determine the scale and direction of key points. Next is direction calculation, using a method called integral image to calculate the main direction for each key point. By calculating the horizontal and vertical gradients of pixels in an image and accumulating the gradients in a fixed size area, a histogram representing the direction of the gradient can be obtained. The peak in the histogram represents the main direction. Another aspect is descriptor calculation, which uses a method called accelerated integral image to calculate the descriptors of key points. A descriptor is a vector that contains feature information about the area around the key point. The SURF algorithm uses the Haar wavelet response of images to calculate descriptors, which can improve computational efficiency while maintaining the accuracy of features. Finally, feature matching can be performed by comparing the descriptors of key points. The SURF algorithm uses a method called fast nearest neighbor search to find the most matching feature points [9].

Overall, the SURF algorithm achieves the detection and description of scale invariant feature points in images by using techniques such as Gaussian blur, Hessian matrix, and integrated image. It has high computational efficiency and robustness, making it suitable for large-scale image processing and computer vision applications.

3.2. Improved SURF Algorithm

In practical applications, the difficulty of recognizing and grasping objects of different shapes and sizes varies. Therefore, it is necessary to design appropriate algorithms and system architectures to achieve precise grasping of target objects. Pan et al. Proposed an improved SURF algorithm for robotic arm recognition, positioning, and grasping.

The scale space constructed by the SURF algorithm can extract feature points with scales, and the r-BRIEF descriptor can generate binary encoded feature descriptors that facilitate high-speed feature matching. Compared with traditional BRIEF descriptors, it also has the property of rotation invariance. Therefore, the traditional SURF algorithm combined with the r-BRIEF descriptor can achieve high-speed matching and accurate recognition of image features. The specific process of the improved SURF algorithm is to use the traditional SURF algorithm to extract target image feature points, determine the main direction of the extracted feature points, and then use r-BRIEF descriptors and low variance filtering to describe and reduce the dimensionality of the extracted features. After dimensionality reduction, only 3-dimensional descriptors are retained for each feature point. By performing 2 Nearest Neighbor feature point cross matching based on Hamming distance to screen out mismatched point pairs, and then cooperating with the RASAC algorithm to further eliminate the remaining mismatched point pairs in the image [10]. Finally, output the feature matching image of the target, and compare the process of the traditional SURF algorithm and the improved SURF algorithm as shown in Fig. 1.

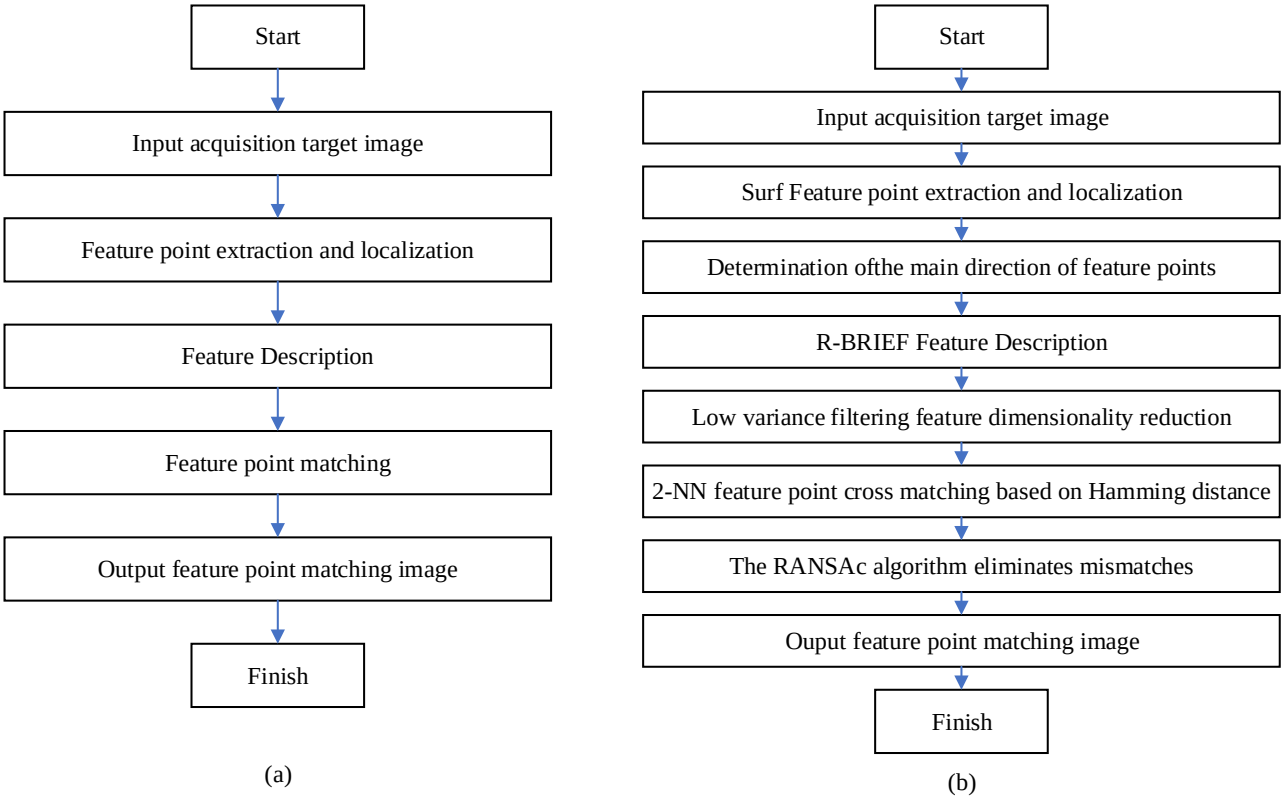


Fig. 1 SURF algorithm flow. (a) is before improvement; (b) is after improvement.

Conducted in-depth research on the improved SURF algorithm combined with binocular stereo vision three reconstruction for the recognition, positioning, and grasping experimental system of the robotic arm, ultimately guiding the robotic arm to carry out real-time grasping experiments on the target. The experimental results show that the proposed improved SURF algorithm can achieve a recognition accuracy of 93.1%, and the overall matching time is shortened by 51% compared to the traditional SURF algorithm. It can significantly improve the matching speed compared to the traditional SURF algorithm while ensuring high matching accuracy. It has high practical significance and application value for fast and accurate recognition of targets in the working area of the robotic arm, as well as real-time and accurate grasping of the robotic arm [11].

4. Conclusion

Eye-in-hand technology involves the installation of a camera at the end of a robotic arm. It utilizes hand-eye calibration to determine the relative position and posture between the camera and the robotic arm. This enables precise positioning and grasping of target objects. The significance of hand-eye calibration lies in its ability to enhance robot autonomy, flexibility, and precision. The accuracy and reliability of this calibration directly impact system performance and task completion quality. To ensure effectiveness, practical applications require meticulous calibration processes and parameter adjustments. This article summarizes several technical research directions for improving the performance of industrial robotic arms. It has profound reference significance for improving the production efficiency and quality of robotic arms, achieving automation and intelligence, adapting to complex environments, and expanding the application fields of robotic arms. In the future, intelligent recognition and grasping control of robotic arms may be further integrated with other technologies such as artificial intelligence, machine learning, sensor technology, etc., to enhance the intelligence level of the system. The robotic arm can adapt better to different tasks and environments through continuous learning and adaptive adjustment, improving its versatility and flexibility.

Based on current visual technology and existing algorithms, significant progress has been made in the recognition and grasping of robotic arms. However, algorithms still need to be improved and

cannot achieve 100% accuracy. It is expected that in the coming years, with further research, algorithms will gradually be improved. With the advancement of visual technology, the recognition and grasping accuracy of robotic arms can also be improved, not only in industry, it can also be widely applied in people's industries and daily lives.

References

[1] Dou Xiao, Huo Guoqing. Three modes of robot development and their policy value. Think Tank Theory and Practice, 2024, 9(01): 64-70.

[2] ZHENG Zhenyu, GAO Jian, ZHENG Zhuoqian, et al. High Precision Hand-eye Calibration Method Based on 3D Visual Point Cloud Registration. Journal of Machine Design, 2023, 40(S2):51-56.

[3] Tai Chengyue, Chen Yeqing, Li Haoran, et al. Research on hand eye calibration method for visual guided robot system based on M-shaped standard blocks. Computer Measurement and Control, 2024, 32 (1): 133-141.

[4] Hu Henan. Research on 3D object detection technology based on monocular vision. University of Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences), 2024.

[5] LI Chen, XU Zunyi, YAN Chunxiang et al. Design and Implementation of Intrusion Detection System Based on Monocular Vision and YOLOv5 Algorithm. Software Guide, 2024: 1-6.

[6] Chen Hanbin. Research on intelligent grasping control of robotic arms based on binocular vision positioning. Mechanical and Electrical Engineering Technology, 2023,52 (12): 115-11.

[7] Wang Zhaoqing, Niu Chaoyi, She Wei et al. Research progress on 3D reconstruction technology based on passive vision. Journal of Zhengzhou University (Science Edition): 2024: 1-7.

[8] Mai C, Chen H, Zeng L , et al.A Smart Cane Based on 2D LiDAR and RGB-D Camera Sensor-Realizing Navigation and Obstacle Recognition.Sensors (Basel, Switzerland),2024,24(3).

[9] Jia Yixin, Deng Weiyong, Yin Qiang, et al. Enhanced Reality Image Matching Method Based on Improved SURF. Computer Technology and Development, 2024,34 (01): 59-64.

[10] Li Xinyan, Zhou Min, Zhang Meimei, et al. Research on Identification, Positioning, and Grasping of Mechanical Arms Based on Improved SURF Algorithm. Combination Machine Tool and Automation Processing Technology, 2024, (01): 47-52.

[11] Song Xiang, Xu Sixiang, Yang Lifa, etc. Binocular vision measurement method based on nonlinear diffusion and high-dimensional M-SURF descriptors. Photoelectron laser, 2024: 413.