

Research on Automatic Grasping System of Industrial Robot Based on Computer Vision

Yiqun Deng^{1,a,*}

¹*Shenzhen Middle School, Shenzhen, China*

a. peterdyq@ldy.edu.rs

**corresponding author*

Abstract: The first paper focuses on YOLOv5 model optimization for advanced image processing, cu. The second paper explores low-cost machine vision cameras, achieving 97% accuracy in grasp verification. The third paper implements SVM algorithm with ArUco markers for hand-eye calibration. The adjustments made to the connection of sensing and motion system significantly expanded the scope of the detection while maintain the accuracy of grasping.. The fourth and fifth experiments improved the existing models by adding coordinate systems and wrist-mounted cameras separately. By doing so, it is possible to catch clearer and more precise pictures of the target objects. The final two research focus on promoting the automatic aspect of the system. Utilizing two networks and corresponding datasets, for one; employing multiple innovative vision-based techniques, for another. These studies collectively demonstrate the advancement in machine vision-based robotic grasping through various technical approaches, including model optimization, hardware evaluation, and algorithm implementation for improved accuracy and efficiency.

Keywords: Automation, Robotic grasping system, Computer vision.

1. Introduction

Since the birth of robots, imitative functions of humans have always been an important research topic for scientists. Among the various functions of robotics, grasping takes an unreplaceable place, for it represents one of the most basic motions of human and exhibits its practicability in numerous fields. As the automation of mechanics spread over industries, a rapid-increasing number of computer-vision-based automated grasping system has attracted extensive interests of researchers and engineers. During the past few years, with the emergence of innovative technologies, researchers have continuously enhanced grasping systems by incorporating cutting-edge developments in the field. For now, the key technologies of grasping systems are divided into the following 3 parts. Sensors like force sensors monitor grip strength to protect fragile objects, while distance sensors help accurately position the arm. Cameras and depth sensors enable the identification of object shapes and locations, facilitating smarter grasping decisions. Control algorithms include PID methods for smooth motion and adaptive algorithms that adjust parameters based on environmental changes. Reinforcement learning further optimizes grasping strategies through trial and error. Vision systems employ computer vision and deep learning techniques for object recognition, thereby enhancing accuracy in dynamic settings. Three dimensional reconstruction provides detailed models, improving grasping precision. In summary, these technologies collectively advance automated grasping capabilities,

driving progress in industrial automation and everyday applications. Following these technological developments, this paper presents research that aims to calibrate accuracy and enhance both the applicability and practicality of machine-vision-based models through various techniques, such as redesigning cameras' working logic, and introducing more advanced mathematics method. On the other hand, however, the included research undeniably has some limitations and defections. For instance, in the first research just as the researchers admitted, they lack dataset gathered from more realistic environment and thus still leave a small blank between the proposed model and a robust one. Another significant example is the fifth experiment. Although not addressed by the researchers, they overlooked several crucial physical properties of the target objects, particularly material density and compliance. Suppose the object is nearly as soft as a fluid, the force that required to use and direction as well as speed will definitely differ from that the model calculated. However, it is undeniable that the improvements done by the researchers are already a great leap from the past [1].

2. Industrial robot automatic grasping system

Design of and research on the robot arm recovery grasping system based on machine vision focuses on the design and implementation of a robot arm recovery grasping system that utilizes machine vision to enhance precision and efficiency in object retrieval [2]. The innovative aspect lies in the integration of advanced image processing techniques to enable the robot arm to identify, locate, and grasp objects in dynamic environments. The methodology involves using a combination of machine learning algorithms, such as the channel pruning technique for improving the perplexity of the YOLOv5 model for object recognition, promoting the number of parameters from more than 7 million to less than 3 million and cutting down the FLOPS value and the model's volume at a cost of only 0.2% of the original accuracy; automatic Gamma correction algorithm that can shift the image contrast and median filtering method for noise reduction are utilized to address to difference in light intensity of the environment; and real-time vision feedback is also included to adjust the grasping strategy based on environmental changes. With those algorithms combined, the researchers ultimately succeeded in improving the success rate of the robot arm's operations and presenting a relatively mature prototype of the robot. However, the researchers acknowledge that further investigation into the structures and dynamics of the robot's hardware is necessary under more realistic circumstances.

Performance Evaluation of Low-Cost Machine Vision Cameras for Image-Based Grasp Verification introduces a cost-effective machine vision-based system for grasp verification in autonomous robots, as shown in Figure 1 [3].

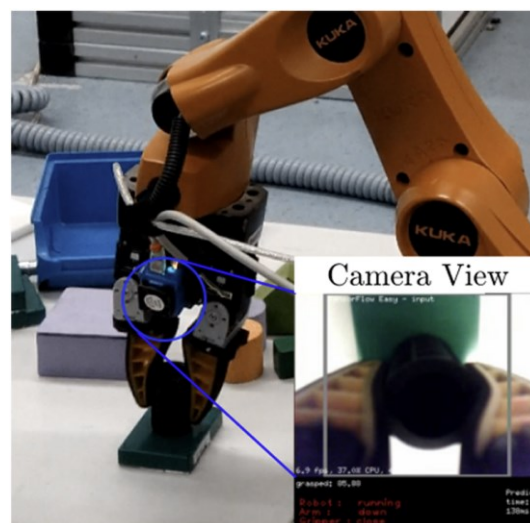


Figure 1: YouBot gripper with machine vision camera JeVoisA33 along with the camera view [3]

Grasp verification is essential for ensuring the successful completion of manipulation tasks by robots, yet it is contingent upon the selection of appropriate sensors. The researchers address this challenge by proposing a system that employs machine vision cameras capable of performing deep learning inference on-board. The research focuses on two low-cost machine vision cameras, the JeVois A33 and the Sipeed Maix Bit, and evaluates their suitability for real-time grasp verification. These cameras facilitate local data processing, thereby reducing the dependency on a centralized server and enhancing the system's responsiveness and reliability. To systematically assess the performance of the cameras, the authors present a novel parameterized model generator. This generator "produces end-to-end CNN models to benchmark the latency and throughput of the selected cameras" [3]. The experimental results indicate that the JeVois A33, despite higher power consumption, can execute a broader range of deep learning models with performance comparable to that of the Sipeed Maix Bit, which features a dedicated CNN accelerator. The paper's primary contribution lies in the comprehensive performance evaluation of the low-cost machine vision cameras for grasp verification tasks. The integration of the JeVois camera with the KUKA youBot gripper demonstrates a reliable grasp verification system, achieving a per-frame accuracy of 97%. This research provides a valuable contribution to the field of robotics by offering a cost-effective solution for grasp verification, potentially expanding the application of machine vision cameras in robotic systems.

Robotic grasping based on machine vision and SVM utilizes the Support Vector Machine (SVM) algorithm to specify the grasping rectangle of objects and calibrate the hand-eye system [4]. This calibration method can determine the relationship between the robot coordinate system and the camera coordinate system, so as to improve the accuracy of the robot end effector reaching the object. In order to promote the precision of the model, the researchers also include the grasping data from the Cornell University to train the model from a deeper degree. The experimenters first calibrate the eye-to-hand system, then collect the images and detect the number of objects. Next, they use morphological processing and determine the grasping angle based on SVM. Finally, they run the robot and test the motion control of it based on the calculated grasping angle. Eventually, the researchers succeeded in expanding the scope of identification while maintaining high grasping accuracy, creating the model shown in the Figure 2 [4].

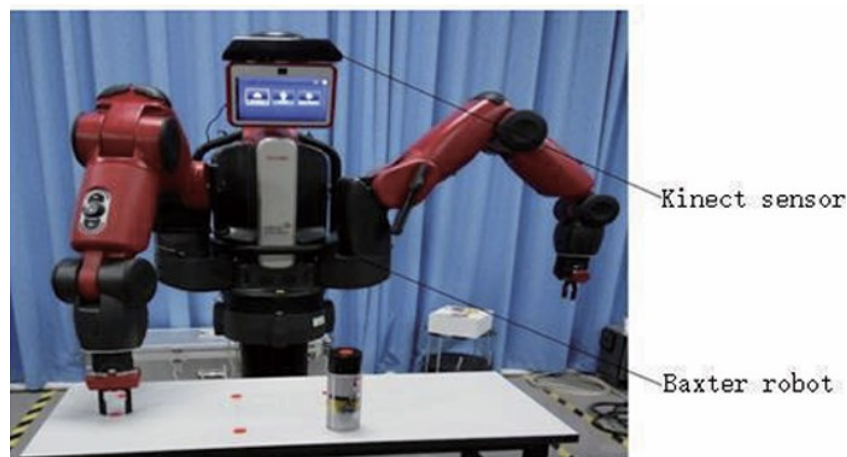


Figure 2: Robotic grasping system [4]

Several other related studies also incorporating SVM methods have been conducted to improve the applicability of grasping robotic arms.

Lei Yang et al. aims to improve the accuracy of mechanical manipulators [5]. This study constructs a system model that integrates a camera and a robotic arm for detecting and grasping objects. In the

experiment, multiple coordinate systems are established for the grasping system, including the robot base, end-effector, camera, image plane, and object coordinate systems. The transformation relationships between these systems are detailed using matrix models. The DLM is also used to calibrate the relationship between the image coordinates and the robot coordinates by using a series of known points. In the final test for the results, the outcome proves to be eligible: the grasping error rate is acceptable and meets the conventional requirement. The researchers conclude that this improved mechanical arm system has significant potential for various engineering applications.

Automated Modeling and Robotic Grasping of Unknown Three-Dimensional Objects, another innovative research, reviewed the historic progresses of vision based grasping system and provides its own innovative contribution on promoted system and disposing logic and methods [6]. To be specific, the researchers use single wrist-mounted video camera to take silhouettes of the object, then utilize the laser beams to combine with the former results. In the paper, previous experiments and attempts to perfect former models are reviewed and commented. Based on them, the authors reflected the drawbacks of those prototypes--to list a few, predetermined objects' shapes, limited grasping mechanical hardware and restricted dimensions applicable--and propose a novel approach that elevates the simplicity and robustness of the model while balancing the strength of the jaws and speed of reaction. Overall, the alternated model exhibited in Figure 3 can be applied in unknown circumstances with its outstanding ability to accustom to its targets [7].

Ruohuai Sun et al discusses an approach to enhance collaborative robotic recognition and grasping system. This approach primarily utilizes deep learning networks to expand the capabilities of robotic grasping systems. Unlike current systems installed in factories, this new model consists of two separate network components. The YOLOv3 network, trained with the COCO dataset, and the GG-CNN network, trained with the Cornell grasping dataset, process inputs and outputs respectively. The former network "identifies the object category and position", while the latter one "predicts the grasping pose and scale". With this new model, as shown in Figure 3 [7].

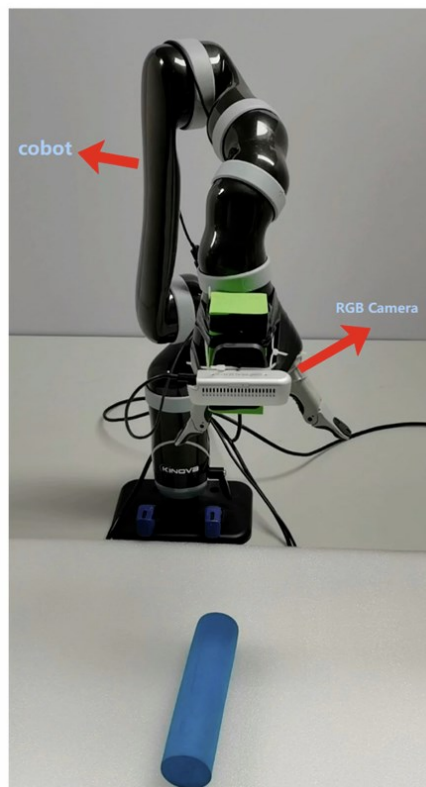


Figure 3: The relative location of the robot and the RGB camera in the workspace [7]

The detection speed increased by more than one-seventh, and the accuracy improved by 4% compared to using YOLOv3 alone. In brief, the parallel network structure validates its effectiveness through actual grasping tests.

Autonomous Robotic Manipulation: Real-Time, Deep-Learning Approach for Grasping of Unknown Objects is another related study that focuses on developing a real-time, data-driven deep-learning approach for robotic grasping [8]. The experimenters aim to reduce propagation errors and abrogate the requirement of numerous other complex traditional methods. To be specific, instead of tracking the widespread method of manually setting up kinematics functions, the new model combined the empirical methods that “provide an increased cognitive and adaptive capacity to the robots” with multiple vision-based deep-learning techniques in order to generalize the model’s applicable range. The main contribution points are as follows: First, they integrate multiple simple and effective techniques for grasp generation to enhance the computer vision. Second, real-time control is proposed for autonomous manipulation, elevating the flexibility and intuition for the user. Thirdly, they employ a flexible multi-view grasp generation and break the constraints of CNN [8].

3. Conclusion

Overall, the research presented above demonstrates fundamental approaches to improving computer vision-based robotic grasping systems. In general, they employed machine vision along with their own innovative technologies and methods: some researchers utilized multiple datasets to train their models while integrating two interconnected operating systems. Some used improved and integrated algorithms to raise the accuracy of the model; some utilized the wrist-mounted camera and laser beams to elevate the grasping strength and calibrate the grasping angle. These improvements, whether focused on cost reduction or accuracy enhancement through algorithmic or hardware modifications, significantly contribute to the system's practical applications. There are, indeed, some issues that required further research and fixing. In general, there are still many shortcomings researchers have to overcome in the future. To list a few: Lack of realistic circumstances, insufficient training data, and vague discussion about the material--both the grasping object and the grasping arms. Therefore, future research should address not only algorithmic and mechanical improvements but also focus on enhancing model robustness. In realistic circumstances, a variety of issues need to be considered for the robotic system to function normally. In brief, in order to adapt to future requirements of robotic systems in industries, scholars and engineers have to tackle the completeness, adjustability and broad applicability of their practicality.

References

- [1] Murali, A., Li, Y., Gandhi, D., & Gupta, A. (2018). *Learning to grasp without seeing*. In *Proceedings of the International Symposium on Experimental Robotics (ISER)* (pp. 375-386). https://doi.org/10.1007/978-3-030-33950-0_33
- [2] Chiu, Y. J., Yuan, Y. Y., & Jian, S. R. (2022). *Design of and research on the robot arm recovery grasping system based on machine vision*. *International Journal of Advanced Robotic Systems*, 19(1), 1-12. <https://doi.org/10.1016/j.jksuci.2024.102014>
- [3] Nair, D., Pakdaman, A., & Plöger, P. G. (2019). *Performance Evaluation of Low-Cost Machine Vision Cameras for Image-Based Grasp Verification*. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (pp. 4675-4681).
- [4] Huang, Y., Zhang, J., & Zhang, X. (2019). *Robotic grasping based on machine vision and SVM*. *International Journal of Advanced Manufacturing Technology*, 102(1), 1027-1037. <https://doi.org/10.1109/CYBER46603.2019.9066460>
- [5] Yang, L., Wu, S., Lv, Z., & Lu, F. (2020). *Research on manipulator grasping method based on vision*. *Journal of Physics: Conference Series*, 1570(1), 012051.

- [6] Bone, G. M., Lambert, A., & Edwards, M. (2008). *Automated Modeling and Robotic Grasping of Unknown Three-Dimensional Objects*. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 292-298).
- [7] Du, G., Wang, K., Lian, S., & Zhao, K. (2021). *Vision based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review*. *Artificial Intelligence Review*, 54(3), 1677-1734. <https://doi.org/10.1007/s10462-020-09888-5>
- [8] Sun, R., Wu, C., Zhao, X., Zhao, B., & Jiang, Y. (2021). *Object Recognition and Grasping for Collaborative Robots Based on Vision*. *Sensors* 2024, 24(1), 195, <https://doi.org/10.3390/s24010195>