Visual-servo NLP-based 6DOF Manipulator Grasp System

1st Yuxiao Hua
School of System Design and Intelligent Manufacturing
Southern University of Science and Technology
Shenzhen, China
1628280289@qq.com

2nd Hongjing Tang
School of System Design and Intelligent Manufacturing
Southern University of Science and Technology
Shenzhen, China
12011827@mail.sustech.edu.cn

3rd Xizhe Hao
School of System Design and Intelligent Manufacturing
Southern University of Science and Technology
Shenzhen, China
12012123@mail.sustech.edu.cn

***Abstract*—Traditional industrial robots work mostly by means of demonstration or offline programming, following a prescribed path, with a single function, and cannot work effectively once the scene changes. The vision robot uses vision technology to obtain information about the target and its surrounding environment, and can guide the industrial robot to complete the path planning and target grasping through decision-making. [2] This project aims to build a system called "Visual-servo NLP-based 6DOF Manipulator Grasp System", which realises the recognition of target objects and the grasp of the robotic arm in ROS, and introduces a natural language processing (NLP) system and a voice control system. Through these studies, we have obtained some important results and conclusions. The realisation of this system is of great significance. It can improve the accuracy and intelligence of robotic arm grip, and provide new solutions for industrial automation and intelligent manufacturing.**

***Index Terms*—Visual-servo, NLP, 6-DOF manipulator**

I. Introduction

In the field of visual servo grasping for robotic arms, many important research results have been achieved. Researchers have proposed various novel methods and algorithms to improve the accuracy and speed of target detection and identification, increase the accuracy and robustness of position estimation and tracking, optimize the control algorithm and feedback strategy, and achieve efficient and flexible grasping actions.

These research results have been widely used in industrial automation. Vision servo gripping of robotic arms can be used for tasks such as part gripping and assembly on automated assembly lines, cargo handling and stacking in warehouse logistics, and surgical assistance and patient gripping in the medical field. In these application scenarios, robotic arms can automatically identify, track and grasp objects according to the characteristics and requirements of the target object, greatly improving work efficiency and accuracy.

In our studies, we aim to construct a system in ROS where the targeted object recognized by a depth camera can be grasped by a robotic arm and placed in a designated location. Moreover, an NLP system is added to recognize the target object from a simple sentence input, such as recognizing "apple" from "I need an apple". Finally, the system is expanded to introduce a voice-controlled system.

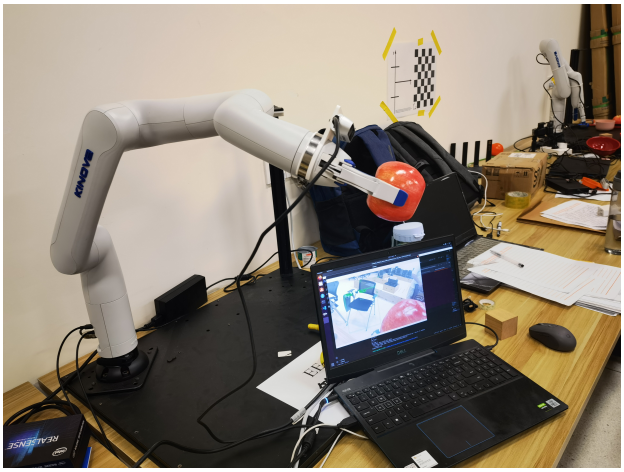

Fig. 1. The Outlook of the System

II. Review

In recent years, robotics had been widely used in industrial tasks in virtue of their high efficiency and accuracy, especially in the assembly industry. However, most of the robots are controlled by teaching device in the market, which limits their working path and makes robot lost their flexibility. While with the development of vision sensor, the robot equipped vision can accomplish more dangerous and accurate positioning tasks, such as underwater exploration [3]. Therefore, the purpose of this paper is to use the image tracking algorithm and visual servoing closed-loop controller to control an uncalibrated 6-DOF robot to complete the positioning and grasping tasks in real world.

III. METHOD

A. Object Recognition

We introduce opencv to read image captured from the camera and display it on the screen of computer in real time. In order to improve the quality of the images and decrease the transmission delay, we apply two methods, selecting an appropriate resolution and setting a buffer in order to release the pressure of continuous image input. Moreover, since the process of image processing and the operation of the manipulator must be parallel, multi-thread is used for the system to work properly, in case that the picture is out of sync with the actual situation.

In order to implement object recognition, we apply MediaPipe. It provides a Box Tracking module that can be used for object detection and tracking in videos or live camera feeds. The module is based on a combination of deep learning-based object detection models and traditional computer vision algorithms for tracking and motion estimation.

Once the objects have been detected, the Box Tracking module applies a set of filtering and tracking algorithms to estimate the position and movement of each object across frames. These algorithms use information from the detected bounding boxes, as well as information from previous frames, to track the objects and maintain their identity across frames.

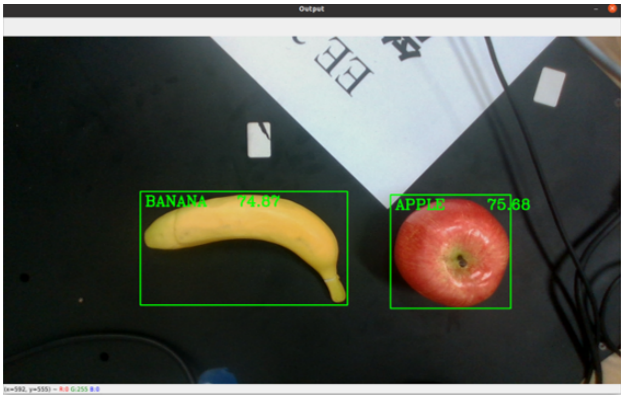

Fig. 2. Google Mediapipe Box Tracking

B. Visual-servo Control

Visual-servo is widely used in robotic arm gripping. It is the process of using computer vision technology and control systems to enable the robotic arm to accurately perceive and track target objects and achieve precise grasping. This is mainly achieved through the following steps:

1) **Position estimation**: Once the target object is detected, the vision servo system can calculate the position and attitude information of the target object relative to the robotic arm. This can be achieved through a two-dimensional visual approach.
2) **Tracking and feedback**: The robotic arm uses the feedback information of the visual servo system to adjust its position and attitude through the control algorithm, so that the robotic arm can accurately track the movement of the target object. This real-time feedback ensures that the robotic arm's distance and attitude from the target object remain within the required range.
3) **Gripping execution**: Once the position and attitude of the target object are accurately estimated and tracked, the robotic arm can perform the gripping action. This can be direct gripping, clamping or other gripping methods suitable for a specific target object.

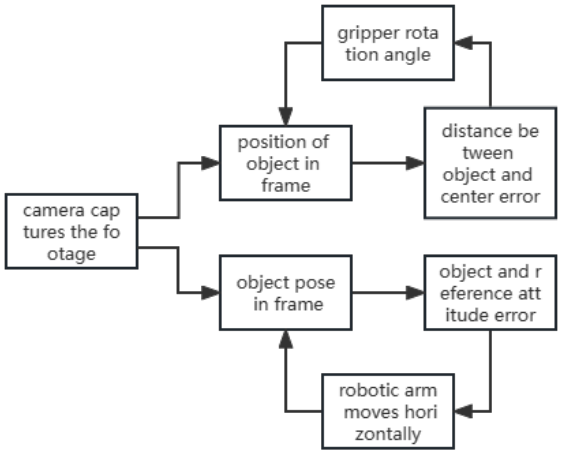

Fig. 3. Visual-servo Control Block Diagram

In our project, after obtaining the position of the object in the frame of camera, we will transform it into the coordinate of the manipulator. Since the camera is fixed on the gripper and the direction of them are on the same axis, we need to ensure the center of the object coincidences with the center of the frame of the camera. With the errors of both x-direction and y-direction calculated, a command of the next relative position of the gripper will be sent to the manipulator. In order to speed up the process of correcting errors, the distance to the next position is proportional to the error, which means the larger the error, the larger the moving distance. When the gripper finishes the movement, the camera will calculate the error again and sent the next moving command, until the error reduces to within a given threshold in case of over-fitting.

C. Natural Language Processing

To make the computer understand the keyword "apple" in the sentence "I need an apple", natural language processing (NLP) [1] is required. The following is a simple processing flow:

1) **Tokenization**: The sentence is divided into individual words or tokens, such as "I", "need", "an", and "apple".
2) **Part-of-speech tagging**: Each word is labeled with its part of speech to determine its grammatical role in the sentence. For example, "apple" is a noun.
3) **Semantic analysis**: The meaning of the words is analyzed based on the context and grammar rules to determine their

meaning in the sentence. For example, ”apple” refers to a type of fruit in this sentence.

4) **Named entity recognition**: The text is analyzed to identify named entities, such as ”apple” being recognized as a fruit entity.
5) **Contextual analysis**: Further analysis of the context and common sense knowledge is used to determine that ”apple” is the main topic and object needed in the sentence.

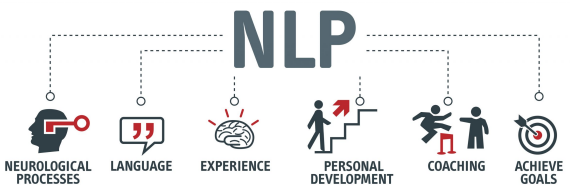

Fig. 4. Natural Language Processing

The above processing flow is a basic step in natural language processing to help the computer understand the keyword ”apple” in the sentence. In practical applications, further adjustments and optimizations are often necessary, such as using deep learning techniques to improve processing accuracy and efficiency.

IV. Results

Our system can work properly for most of the cases and the response speed is relatively fast. After finishing the setting up procedure, which includes the initial process of the camera and the manipulator, the program will ask you to input the command sentence into the console. Based on the sentence, it will automatically analyse whether you need something and if there exists the specific object you really need. Then it will adjust its position on horizontal plane to the directly above of the object with visual-servo control method. Then the gripper will gradually approach to the object on vertical direction. Once it moves to the appropriate position, the gripper will close to a certain extent, depending upon the type of object. Finally, the manipulator will send the grasped object to the hand of the operator. After restoring the system to the initial state, we can conduct another mission with a new command sentence.

V. Discussion

As to the object recognition function, the system responses perfectly for apple and orange, probably because their circle shape. However, for banana and other anomalous objects, the performace of method of edged rectangular box is not as expected. This is because if the banana is placed tilt to the axis, the center of the box fails to coincidence with the center of the object, especially for the curved ones. We are trying to solve the problem with another method for determining the position, or to be precise, the center of the object. The box should be placed with a certain angle to ensure the maximum occupation of the object in the area of the box. Another idea is to attach edge detection method to the original one. After finding the object with a rough range, use edge detection to determine a more precise range with the box. This will improve the performance of the system in case of anomalous objects.

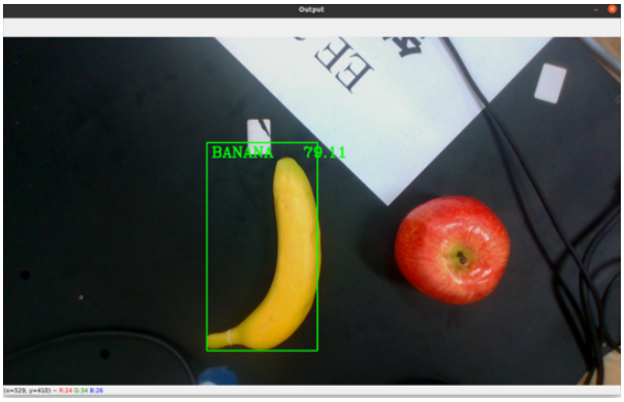

Fig. 5. Case of Anomalous Objects

VI. Conclusion

In this study, a 6-DOF robotic arm handling system based on visual-servo and natural language processing is introduced. Traditional industrial robots face the limitations of rigid path planning and fixed functions during operation, and cannot adapt to the changes of complex scenes. To overcome these problems, we propose a novel system that combines vision technology and natural language processing. Through vision technology, we are able to obtain information about the target object and its surroundings in real time, and servo control the robotic arm according to this information to achieve accurate grasping. At the same time, we introduced a natural language processing system and a voice control system to enable operators to interact directly with the robotic arm through voice commands. We implemented the system on the ROS platform and conducted a series of experiments to verify its performance. Experimental results show that our system can achieve accurate identification and stable grasping of target objects, and has a fast response speed. However, object recognition still has certain challenges for some unconventionally shaped objects and needs to be further improved. We propose the idea of using edge detection and more accurate positioning methods to improve recognition performance. In general, the system has important practical application value, which can improve the accuracy and intelligence of robotic arm grasping, and provide new solutions for industrial automation and intelligent manufacturing.

References

[1] Guida, G., Mauri, G. (1986). Evaluation of natural language processing systems: Issues and approaches. Proceedings of the IEEE, 74, 1026-1035.
[2] LIN Yizhong, CHEN Xu. Research progress of robot positioning and grasping based on machine vision[J].Automation and Instrumentation,2021(3):9-12.
[3] Y. Lee, J. Choi, J. Jung, T. Kim and H. Choi, ”Underwater robot exploration and identification using dual imaging sonar: Basin test”, 2017 IEEE Underwater Technology (UT), pp. 1-4, 2017.

VII. Appdendix I——Contribution

Hongjing Tang: visual-servo control and camera configuration

Yuxiao Hua: multi-thread, NLP and final report

Xizhe Hao: voice control and PPT presentation

VIII. Appdendix II——Github link

https://github.com/HuaYuXiao/Visual-servo-NLP-based-6DOF-Manipulator-Grasp-System