Autonomous recognition grasping robot based on speech input

盛李杰: 12011127 | 程耀宇: 12010922

包辰博: 12012309 | 张中堂: 12012330

张子尚: 12012305

2023/3/6

AncoraSIR.com

SUSTech
Southern University
of Science and Technology

Autonomous recognition grasping robot based on speech input

- In the collaborative robotics category, it is extremely important to be able to establish the relationship between human and robotic arm in a reasonable way, and with our group's point of view, language is a good means to do so. Language can convey very specific information and is well used by the human operator during the interaction process. Controlling the robotic arm by voice command can reduce the time and labor intensity of manual operation and improve productivity and efficiency. And in certain environments, such as abnormal temperature, high altitude or other dangerous environments, voice controlled robotic arms can replace manual operation, reducing the risk of injury to workers.
- For the reading to be examined, we have chosen “Learning 6-DoF Object Poses to Grasp Category-level Objects by Language Instructions”. This paper studies the task of any objects grasping from the known categories by free-form language instructions. This task demands the technique in computer vision, natural language processing, and robotics. Critically, the key challenge lies in inferring the category of objects from linguistic instructions and accurately estimating the 6-DoF information of unseen objects from the known classes. In contrast, previous works focus on inferring the pose of object candidates at the instance level. This significantly limits its applications in real-world scenarios.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Autonomous recognition grasping robot based on speech input

- Data usage in this project can be split into two sections, namely the speech recognition section and the image identification section. Voice message and images with features and labels will be used in this project. Concerning the voice recognition, we will probably use the dataset provided by iFLYTEK. With the image detector, we will use the RefCOCO dataset, which provides textual descriptions and images from MSCOCO dataset.
- We are going to use speech recognition to obtain correct and mobilized input and output. This is achieved by using relevant datasets and speech recognition packages. Computer vision is also needed to identify the size and position of the objects. Finding the forward and inverse kinematics solution and path planning will be carried out with MATLAB.
- After completing the project, we expect that the robotic arm could get the target object we say. We can analyze the results through three indexes: (1) the number of robotic arm targets that the speech recognition can convey; (2) the accuracy of the robot arm in identifying the target; (3) the autonomous planning ability of the manipulator to grasp the target. Through collecting the data of the evaluation indexes, we can assess how well the robotic arm performs. After all the training part, we expect that (1)speech recognition can convey all the statements in standard form into the target of the robotic arm; (2) the accuracy of the robot arm in identifying the target will reach 80% and it can handle the situation of non-object recognized; (3) the arm could grasp the target autonomously.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

What is the problem that you will be investigating?

Why is it interesting?

- In the collaborative robotics category, it is extremely important to be able to establish the relationship between human and robotic arm in a reasonable way, and with our group's point of view, language is a good means to do so. Language can convey very specific information and is well used by the human operator during the interaction process.
- Imagine you are in the middle of an assembly and you need to get a hex wrench while your hands are holding the assembly in place, at which point you say something like, "A hex wrench, please", at which point the robotic arm recognizes your request and hands you the hex wrench you need.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

What is the problem that you will be investigating?

Why is it interesting?

Improve efficiency: Controlling the robotic arm by voice command can reduce the time and labor intensity of manual operation and improve productivity and efficiency.

Enhance safety: In certain environments, such as high temperature, low temperature, high altitude and other dangerous environments, voice controlled robotic arms can replace manual operation, reducing the risk of injury to workers.

AncoraSIR.com

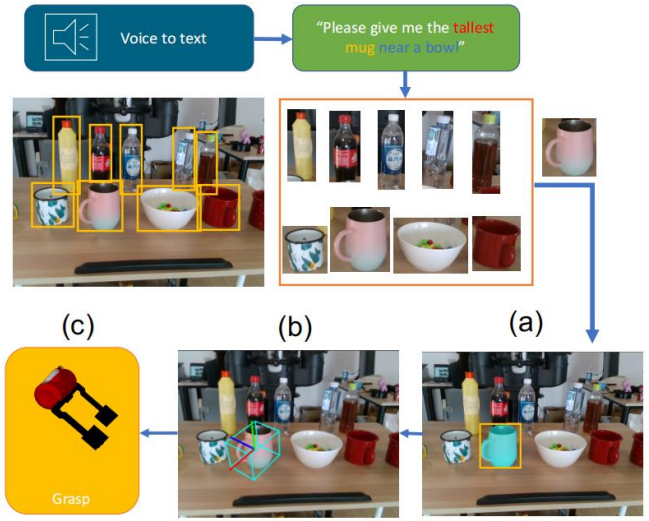
What reading will you examine?

To provide context and background

"Learning 6-DoF Object Poses to Grasp Category-level Objects by Language Instructions"

1、This paper studies the task of any objects grasping from the known categories by free-form language instructions. This task demands the technique in computer vision, natural language processing, and robotics.

2、 Critically, the key challenge lies in inferring the category of objects from linguistic instructions and accurately estimating the 6-DoF information of unseen objects from the known classes. In contrast, previous works focus on inferring the pose of object candidates at the instance level. This significantly limits its applications in real-world scenarios.



AncoraSIR.com

What data will you use?

- Data usage in this project can be split into two sections, namely the speech recognition section and the image identification section. The former transforms voice message to words, while the latter recognizes features from images and relates them to labels. Voice message and images with features and labels will be used in this project.
- Concerning the voice recognition, we will probably use the dataset provided by iFLYTEK. With the image detector, we will use the RefCOCO dataset, which provides textual descriptions and images from MSCOCO dataset.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

What method or algorithm are you proposing?

- **Speech recognition package**

----Correct and easily mobilized input and output

（iFLY tek open platform & Wechat offline Voice Package）

- A clear **interactive interface** for speech recognition

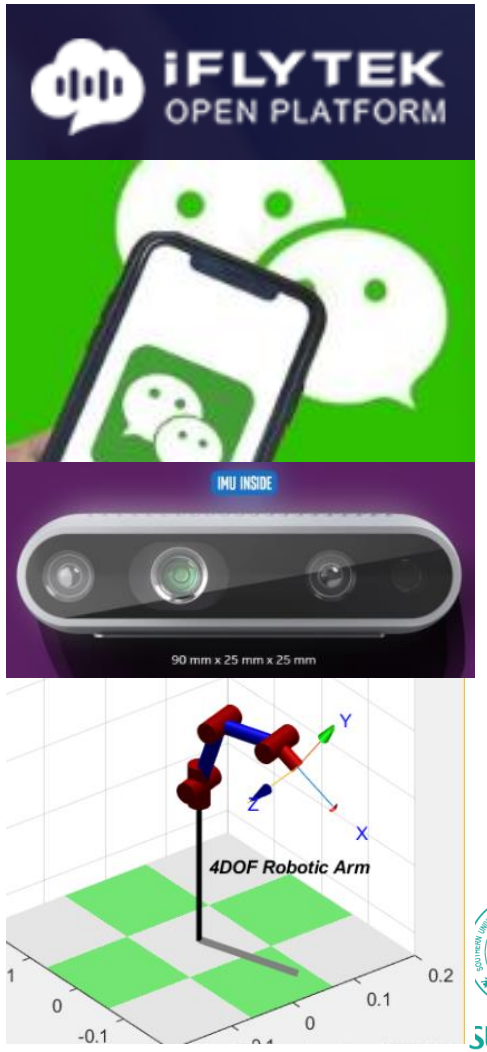
- **Computer vision**

---- Depth camera identifies the size and position of objects

（(RGBD Camera Recognition Color Filtering/Edge Recognition)）

- Forward and inverse **kinematics solution and path planning**

MATLAB toolbox simulation and kinematics solution



AncoraSIR.com

How will you evaluate your results?

Qualitatively, what kind of results do you expect (e.g., plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g., what performance metrics or statistical tests)?

- Evaluation index:
- (1) The number of robotic arm targets that the speech recognition can convey
- (2) The accuracy of the robot arm in identifying the target
- (3) The autonomous planning ability of the manipulator to grasp the target

- Expected result:
- (1) speech recognition can convey all the statements in standard form into the target of the robotic arm.
- (2) The accuracy of the robot arm in identifying the target reach 80% and it can handle the situation of non-object recognized
- (3) The arm could grasp the target autonomously

- Ways of the evaluation analyzed
- Through collecting the data of the evaluation index mentioned before, we can get how well the arm performs.

AncoraSIR.com

SUSTech
Southern University of Science and Technology