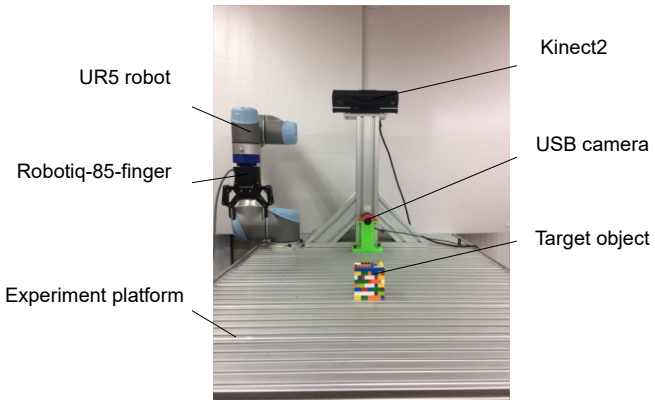
Visual identification and location algorithm for robot based on the multimodal information

Wei Yufeng, Liang Dongtai*, Liang Dan, Xing Shumin

School of Mechanical Engineering and Mechanics, Ningbo University, Ningbo, Zhejiang 315000, China



The experiment platform

Overview: In recent years, various kinds of robots have got studied and used with the development of the robotics technology. In industrial production, workers often need to complete a large number of handling and assembly work. Using robots instead of the workers to complete these tasks would help to increase the efficiency of factories and reduce the labor intensity. Sensing the position of objects is one of the key problems for the robots to complete the task of picking objects. Machine vision is an effective methods to solve this problem, and it is also one of the research hotspots in the robot field at present. Traditional methods are based on two-dimensional images. Because two-dimensional images can not meet the demand, researchers extend their vision to three-dimensional images. Recent years, neural networks have been applied comprehensively, and machine vision has developed rapidly. However, for the traditional research methods, most of the sources of image information are single and basically come from one type of image. As the scene is always rather complicated, the traditional methods are usually faced with the problems such as incomplete image information, large recognition error and low accuracy.

To overcome these problems, a visual identification and location algorithm based on multi-modal information is proposed, and the fusion process is performed by extracting the multimodal information of the two-dimensional image and the point cloud image to realize object recognition and positioning. Firstly, the target 2D image information is obtained by RGB camera. The contour is recognized through the contour detection and matching process. Then, the image SIFT feature is extracted for location tracking and the position of object is obtained. Meanwhile, in order to identify the point cloud image, a point cloud image is captured by RGB-D camera and the best model can be sorted through pre-processing, Euclidean cluster segmentation, computing VFH feature and KD-tree searching. Thus, the orientation can be obtained by registering the point clouds. Finally, use the two-dimensional images and point cloud image to process object information, and to complete the identification and positioning of the target. The effect of the method is verified by the robotic gripping experiment. The results show that the multi-modal information of two-dimensional image and point cloud images can be used to identify and locate different target objects. Compared with the processing method using only two-dimensional or point cloud single-mode image information, the positioning error can be reduced to 50%, and the robustness and accuracy can be improved effectively.

Citation: Wei Y F, Liang D T, Liang D, *et al*. Visual identification and location algorithm for robot based on the multimodal information[J]. *Opto-Electronic Engineering*, 2018, **45**(2): 170650

Supported by National Natural Science Foundation of China (51305214), the Public Technology Application Project of Zhejiang (2017C31094) and the Natural Science Foundation of Ningbo (2017A610124)

* E-mail: liangdongtai@nbu.edu.cn