Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning

and Large-Scale Data Collection

深度学习和大规模数据采集让机器人抓握学会手眼协调

We describe a learning-based approach to hand-

eye coordination for robotic grasping from

monocular images.

过程

To learn hand-eye coordi-

nation for grasping, we trained a large convo-

lutional neural network to predict the probabil-

ity that task-space motion of the gripper will re-

sult in successful grasps, using only monocular

camera images and independently of camera cal-

ibration or the current robot pose. This requires

the network to observe the spatial relationship

between the gripper and objects in the scene,

thus learning hand-eye coordination. We then

use this network to servo the gripper in real time

to achieve successful grasps.

数据

To train our net-

work, we collected over 800,000 grasp attempts

over the course of two months, using between 6

and 14 robotic manipulators at any given time,

with differences in camera placement and hard-

ware.

结果

Our experimental evaluation demonstrates

that our method achieves effective real-time con-

trol, can successfully grasp novel objects, and

corrects mistakes by continuous servoing.

Our method consists of two components:

a grasp success predictor, which uses a deep convolutional neural network

(CNN) to determine how likely a given motion is to pro-

duce a successful grasp,

a continuous servoing mechanism that uses the CNN to continuously update the robot’s

motor commands. By continuously choosing the best pre-

dicted path to a successful grasp, the servoing mechanism

provides the robot with fast feedback to perturbations and

objectmotion, as well as robustness to inaccurate actuation

Visual servoing, also known as vision-based robot control and abbreviated VS, is a technique which uses feedback information extracted from a vision sensor (visual feedback[1]) to control the motion of a robot.

There are two fundamental configurations of the robot end-effector (hand) and the camera:[4]

Eye-in-hand, or end-point closed-loop control, where the camera is attached to the moving hand and observing the relative position of the target.

Eye-to-hand, or end-point open-loop control, where the camera is fixed in the world and observing the target and the motion of the hand.