Teaching Robots to Do Object Assembly using Multi-modal 3D Vision

Weiwei Wan\textsuperscript{a}, Feng Lu\textsuperscript{b}, Zepei Wu\textsuperscript{a}, Kensuke Harada\textsuperscript{a}

\textsuperscript{a}National Institute of Advanced Industrial Science and Engineering, Japan
\textsuperscript{b}Beihang University, China

Abstract

The motivation of this paper is to develop an intelligent robot assembly system using multi-modal vision for next-generation industrial assembly. The system includes two phases where in the first phase human beings demonstrate assembly to robots and in the second phase robots detect objects, plan grasps, and assemble objects following human demonstration using AI searching. A notorious difficulty to implement such a system is the bad precision of 3D visual detection. This paper presents multi-modal approaches to overcome the difficulty: It uses AR markers in the teaching phase to detect human operation, and uses point clouds and geometric constraints in the robot execution phase to avoid unexpected occlusion and noises. The paper presents several experiments to examine the precision and correctness of the approaches. It demonstrates the applicability of the approaches by integrating them with graph model-based motion planning, and by executing the results on industrial robots in real-world scenarios.

Keywords:
3D Visual Detection, Robot Manipulation, Motion Planning

1. Introduction

The motivation of this paper is to develop an intelligent robot assembly system using multi-modal vision for next-generation industrial assembly: (1) A human worker assembles objects in front of a vision system. (2) The system detects the position and orientation of the objects and learns how to do assembly following the human worker’s demonstration. (3) An industrial robot performs assembly tasks using the data learned from human demonstration. It detects objects in its workspace, picks up them, and does assembly using assembly planning and motion planning.

The difficulty in developing such a system is precise visual detection. Two problems exist where the first one is in the human teaching phase, namely how to precisely detect the position and orientation of the objects in human hands during manual operation; The second one is in the robot execution phase, namely how to precisely detect the position and orientation of objects in the workspace and perform assembly.

This paper develops a novel multi-modal approach to solve the two problems. First, we attach AR markers to the objects for assembly and track them by detecting and transforming the marker positions during human demonstration. We don’t need to worry about occlusions in this process since the teaching phase is manual and is performed by human beings who are smart enough to actively avoid occlusions and ensure good exposure to vision systems. The modal employed in this phase is the markers (RGB image) and the geometric relation between the markers and the object models. The tag “AR(RGB)” is used to denote this modal. Second, during robot execution, we roughly detect the pose of the objects by matching the object model to the point clouds obtained from depth camera and use the geometric constraints of horizontal table surface to refine the detected results. The robot execution phase is automatic and is not as flexible as human teaching. Therefore we use markerless approaches to avoid occlusions. We fuse the point clouds and geometric constraints to make up the partial loss and noises, and improve the detected results. The geometric constraints are based on an assumption that when an object is placed on the surface of a table, it stabilizes at a limited number of poses. These poses freeze some Degree of Freedom and improve precision. The modal employed in this phase is the cloud point data and the geometric constraints of a horizontal surface. The tag “Depth+Geom” is used to denote it. Moreover, we propose an improved graph model based on our previous work to perform integrated assembly planning and motion planning.

Our contribution is we use different modals according to the requirements and limitations of different phases. In the human teaching phase, AR markers are used since human beings could control the operation and actively avoid occlusion. In the robot execution phase, point clouds are used to find a rough pose since occlusion happens frequently. Geometric constraints are used to improve the rough results. Experiments are performed to examine the precision and correctness of our approaches. We quantitatively show the advantages of “AR(RGB)” and “Depth+Geom” in next-generation industrial assembly and concretely demonstrate the process of searching and planning using the improve graph model. The developed approaches are integrated with Kawada Nextage Robots to show the applicability in real-world scenarios.

2. Related Work

This paper is highly related to studies in 3D object detection for robotic manipulation and assembly and the literature review...
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات