Understanding Human Behavior for an Efficient Human-Robot Collaboration

Matteo Terreran, Leonardo Barcellona, Davide Allegro, Emanuele Menegatti, Stefano Ghidoni
Department of Information Engineering
University of Padua, Padua, Italy

***Abstract*—When considering close human-robot collaboration, perception is crucial to guarantee a safe and effective interaction. This contribution presents a perception system designed to be integrated in a robotic workcell that can effectively observe and analyze the production process involving both a robot and a human operator.**

***Index Terms*—Human-robot cooperation, cooperative production, behavior recognition, body pose estimation.**

I. Introduction

Close human-robot collaboration is a key element for a new production paradigm that is, at the same time, efficient, flexible, customizable, and careful towards the human operator. Such efficient and natural collaboration can be achieved only through a deep understanding of the production process, which includes 1) objects involved in the procedure, and 2) what the human operator is doing. While the first element can be addressed by means of object detection, that can be addressed by means of brilliant solutions already proposed in the scientific literature, the second element is still an open problem.

II. Human Perception in Human-Robot Collaborative Scenarios

Perceiving humans is a challenging task, that results even more complicated in a robotic workcell where collaborative tasks take place: they often require complex manipulations that lead to a number of occlusions. To tackle such events, we propose a system based on a network of RGB-D cameras positioned around the robot workcell, providing information from multiple points of view to be robust to occlusions. All the cameras are calibrated both intrinsically and extrinsically, in order to express the information from each camera in a common coordinate reference frame. Each camera is attached to a processing device (a PC or an edge device with enough processing power) which analyzes the RGB-D data stream by means of AI-driven perception modules, providing mid-level information about people in the scene, including body posture and the segmentation of the body parts, as shown in Figure 1. The human perception system is composed of several AI-based modules developed to provide a holistic understanding of the human worker, considering different types of information such as human pose and human volume [1].

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101006732, Project Drapebot.

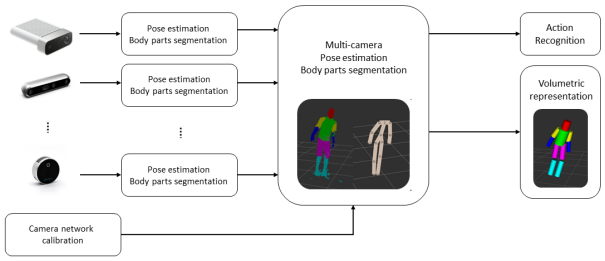

Fig. 1. Overview of the human perception system. A dedicated processing node analyzes the RGB-D stream of each camera, computing human pose estimation and body part segmentation. All information is combined together by a central PC exploiting the camera network calibration, allowing to compute a volumetric human representation and to recognize human actions.

One of the most important elements of the system is the human action recognition module, that is based on a Graph Convolutional Network (GCN) taking as input sequences of human body configurations (or human skeletons) computed by the previous modules [2]. Sequences of skeletons provide a robust representation of the human movements, free from any disturbances like external objects, lighting, and aesthetic differences of people. The vision system recognizes the person's actions at various levels: a general classification of the type of action taking place (e.g., pick, place, request, hand to), and a finer recognition of the main direction of the movement and its intensity (e.g., small, medium, high intensity) useful for better characterize particular actions such as *pulling, pushing* or *pointing*.

III. Conclusions

This paper outlined a complete perception system for handling human perception at mid and high level, enabling complex collaborative processes. The sensor infrastructure and the main IA modules were cited. The research group is currently actively working on the high level part of the system, dealing with action recognition.

References

[1] Leonardo Barcellona, Matteo Terreran, Daniele Evangelista, and Stefano Ghidoni, "Clustering-based refinement for 3d human body parts segmentation", in 2022 17th International Conference on Intelligent Autonomous Systems (IAS-17), 2022. Accepted, in press.

[2] Matteo Terreran, Margherita Lazzaretto, and Stefano Ghidoni, "Skeleton-based action and gesture recognition for human-robot collaboration", in 2022 17th International Conference on Intelligent Autonomous Systems (IAS-17), 2022. Accepted, in press.