

# Intuitive Tasks Planning Using Visuo-Tactile Perception for Human Robot Cooperation

Sunny Katyara<sup>1,2</sup>, Fanny Ficuciello<sup>2</sup>, *Senior Member, IEEE*, Tao Teng<sup>3</sup>, Fei Chen<sup>4</sup>, *Senior Member, IEEE*, Bruno Siciliano<sup>2</sup>, *Fellow, IEEE*, Darwin G. Caldwell<sup>1</sup>, *Senior Member, IEEE*,

**Abstract**—Designing robotic tasks for co-manipulation necessitates to exploit not only proprioceptive but also exteroceptive information for improved safety and autonomy. Following such instinct, this research proposes to formulate intuitive robotic tasks following human viewpoint by incorporating visuo-tactile perception. The visual data using depth cameras surveils and determines the object dimensions and human intentions while the tactile sensing ensures to maintain the desired contact to avoid slippage. Experiment performed on robot platform with human assistance under industrial settings validates the performance and applicability of proposed intuitive task formulation.

## I. INTRODUCTION

In last two decades, the robots have gained sufficient social trust and are extensively participating with humans in performing certain tasks that require cognitive abilities of humans to be combined with precision and strength of robots [1][2]. However, such a social trust on robots requires to endow them with multi-sensory information especially the visuo-tactile feedback to make instant decisions, detect obstacles, recognize human intervention and adapt to varying environment proactively [3][4]. The visuo-tactile data thus enables the robot to understand the non-verbal cues to collaborate more intuitively from human perspective.

Exploiting visuo-tactile information in cluttered environment for human-robot joint carrying task, a framework is proposed in [5]. Under this approach, the robot tasks are defined using standard stack-of-task (SoT) formulation but without taking into account the human intuition and ergonomics. In the same way, a modified technique is presented in [6] for industrial assembly tasks. Wherein, the adaptive gains and homotopy are introduced in conjunction with visuo-tactile data for switching human-robot roles smoothly according to task requirements but however it does not consider progressive mutations in the agents' (i.e, robot and human) behavior and environment. Hence, to the extent of our knowledge, there is no any intuitive task formulation framework available that considers the human ergonomics and task progress in

This research is supported by projects "LEARN-REAL" funded by EU H2020 ERA-Net Chist-Era and "HARMONY" funded by EU H2020 R&I under grant agreement 101017008. (Corresponding author: Sunny Katyara)

<sup>1</sup> Department of Advanced Robotics, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genova, Italy (e-mail: name.surname@iit.it).

<sup>2</sup> Department of Information Technology and Electrical Engineering and PRISMA Lab, University of Naples Federico II, Naples 80125, Italy (e-mail: name.surname@unina.it).

<sup>3</sup> Università Cattolica del Sacro Cuore, Via E.Parmense 84, Piacenza, 29122, Italy (e-mail: name.surname@unina.it).

<sup>4</sup> Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: f.chen@ieee.org).

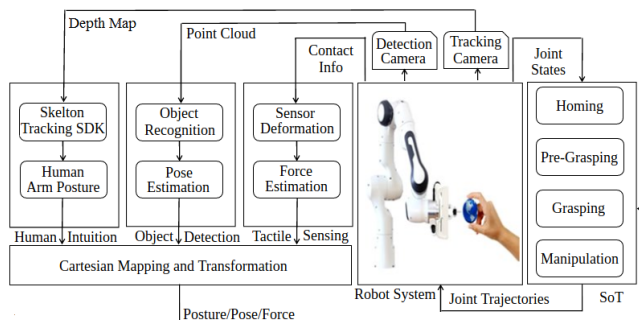


Fig. 1: Proposed research methodology on formulating intuitive robot actions using an idea of SoT for co-manipulation using visual data for human intuition estimation and object detection and the tactile sensing for flexible object interaction. Human gestures are determined using a standard skeleton tracking algorithm. For object detection, RANSAC algorithm together with SVM classifier is used while the tactile sensing ensure to modulate the gripper's force profile to avoid object slippage based on the friction cone criteria.

planning hierarchical robot actions, thereby exploiting visuo-tactile perception for fine co-manipulation tasks. The results presented here are the part of our recent works in [7][8].

## II. RESEARCH METHODOLOGY

The robot tasks are defined in two groups i.e, the Cartesian tasks accounting for position and orientation of end-effector and the force tasks ensuring flexible and adaptive interaction with the human and objects in the environment. All the primary and secondary tasks are defined in a stack with hard and soft priorities being assigned to each at different levels and are executed sequentially, following a standard hierarchical control formulation called SoT. However, the tasks in SoT framework are Quadratic Programming (QP) problems and are formulated according to [9][10].

The defined tasks are subsequently augmented with visual and tactile information for intuitive decision making. Hence, a standard skeleton tracking algorithm (i.e, deep ConvNets) [11] is used for tracing the gestures of active human arm using an aligned depth map captured with RGBD tracking camera in Fig. 2 (a), which is registering the poses of 18 human joints in the local camera frame. Next, a modified RANSAC algorithm together with Support Vector Machine (SVM) classifier is used for object semantic segmentation and recognition respectively and the poses of candidate objects are enumerated by computing the centroid of their processed point cloud captured with RGBD detection camera in Fig. 2 (b). Moreover, for interaction tasks i.e, grasping and manipulation defined in SoT, the tactile feedback is explicitly

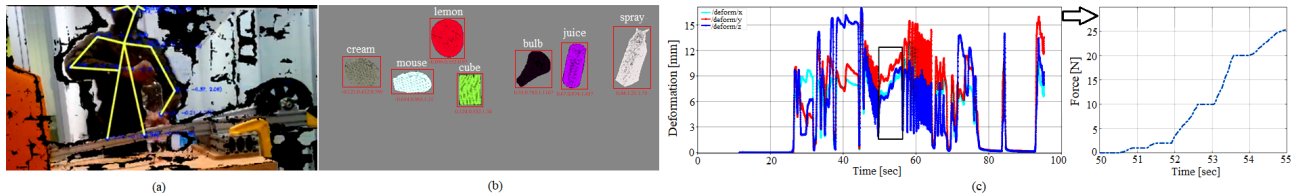


Fig. 2: Visuo-tactile perception in formulating intuitive robotic tasks, (a) represents the aligned depth map on skeleton tracking of human subject performing desired actions on object in the scene, (b) is the point cloud of the scene with objects being recognized and their poses being enumerated locally, (c) is the 3D deformation output of sponge based tactile sensors attached to gripper and the mapped force profile of gripper using shallow neural network for interaction modulation.

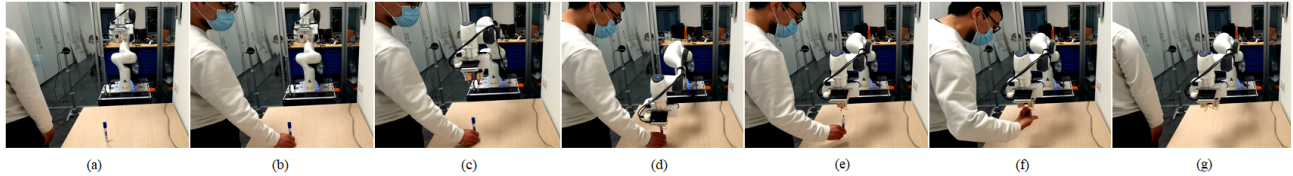


Fig. 3: Robot collaborating with human on detaching marker from its cap, (a) represents the initial configuration of all the actors (human, robot and object), (b) illustrates the human gripping marker from its bottom, in (c) the robot arm assumes pre-grasping posture based on the pose of active human arm wrist, (d) shows the robot arm grasping cap from its centroid posture, in (e) human pulls down the marker while the gripper holds the cap, in (f) the gripper releases cap on detecting open human palm, in (g) the robot arm returns to homing position following its human partner. (Video: <https://www.youtube.com/watch?v=j41rgnEavx4>).

being used. The installed tactile sensors [12] provide 3D deformation output and it is mapped to gripper’s force profile using a shallow neural network (with 5 hidden neurons), in Fig. 2 (c). All the sensory observations are primarily in local reference frames and transformed into robot base using suitable transformations for homogeneous computations.

### III. RESULTS AND DISCUSSION

To better evaluate the reliability and robustness of proposed intuitive task formulation, an industrial test scenario i.e, removing marker from the cap is considered, as shown in Fig. 3. At first, all the characters i.e, robot, human and environment are in their initial configurations in Fig. 3 (a) and then the human tries to grip the marker from bottom in Fig. 3 (b) which is determined by tracking camera and thus the robot arm assumes the pre-grasping posture (i.e, 40 cm above the active human arm wrist) following human arm gesture in Fig. 3 (c). In this configuration, the detection camera recognizes the cap-marker pair in the scene and estimates their poses, which are used by the robot system to grasp the cap in Fig. 3 (d). Once the contact is established with cap-marker, the robot arm lifts it up (20 cm) in Fig. 3 (e) to provide a sufficient space to human partner to perform required action (pulling down) on the marker in Fig. 3 (f). After completing required task, the robot arm returns to its homing position following the human gesture in Fig. 3 (g).

### IV. CONCLUSIONS

This research proposed to exploit visuo-tactile information in formulating robotic tasks in accordance with human intuitions and task progress. Firstly, the visual feedback from tracking camera using an aligned depth map estimated the gesture of active human arm to guide the robot arm to cooperate accordingly and later the detection camera enumerated the object pose from the filtered point cloud, which was sent to gripper for respective grasping action. With the object

being grasped, as detected by tactile sensors, the human executed designated task while the tactile sensors modulated the gripper’s force profile to maintain the continuous desired contact with the object consistently.

### REFERENCES

- [1] O. Khatib, K. Yokoi, O. Brock, K. Chang, and A. Casal, “Robots in human environments: Basic autonomous capabilities,” *The International Journal of Robotics Research*, vol. 18, no. 7, pp. 684–696, 1999.
- [2] A. De Santis, B. Siciliano, A. De Luca, and A. Bicchi, “An atlas of physical human–robot interaction,” *Mechanism and Machine Theory*, vol. 43, no. 3, pp. 253–270, 2008.
- [3] I. Mourtua, A. Iburguren, J. Kildal, L. Susperregi, and B. Sierra, “Human–robot collaboration in industrial applications: Safety, interaction and trust,” *International Journal of Advanced Robotic Systems*, vol. 14, no. 4, p. 1729881417716010, 2017.
- [4] A. Cherubini and D. Navarro-Alarcon, “Sensor-based control for collaborative robots: Fundamentals, challenges, and opportunities,” *Frontiers in Neurobotics*, vol. 14, p. 113, 2021.
- [5] D. J. Agravante, A. Cherubini, A. Bussy, P. Gergondet, and A. Kheddar, “Collaborative human-humanoid carrying using vision and haptic sensing,” in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 607–612.
- [6] A. Cherubini, R. Passama, P. Fraisse, and A. Crosnier, “A unified multimodal control framework for human–robot interaction,” *Robotics and Autonomous Systems*, vol. 70, pp. 106–115, 2015.
- [7] S. Katyara, F. Ficuciello, F. Chen, B. Siciliano, and D. G. Caldwell, “Vision based adaptation to kernelized synergies for human inspired robotic manipulation,” *arXiv preprint arXiv:2012.07046*, 2020.
- [8] S. Katyara, F. Ficuciello, T. Teng, F. Chen, B. Siciliano, and D. G. Caldwell, “Formulating Intuitive Stack-of-Tasks with Visuo-Tactile Perception for Collaborative Human-Robot Fine Manipulation,” *arXiv preprint arXiv:2103.05676*, 2021.
- [9] A. Rocchi, E. M. Hoffman, E. Farnioli, and N. G. Tsagarakis, “A whole-body stack-of-tasks compliant control for the humanoid robot coman,” in *IEEE/RSJ international conference on intelligent robots and systems (IROS 2015)*, Hamburg, Germany, vol. 28, 2015.
- [10] S. Katyara, F. Ficuciello, D. G. Caldwell, F. Chen, and B. Siciliano, “Reproducible Pruning System on Dynamic Natural Plants for Field Agricultural Robots,” *arXiv preprint arXiv:2008.11613*, 2020.
- [11] “<https://www.cubemos.com/skeleton-tracking-sdk>,” Mehlbeerstrasse 2, 82024 Taufkirchen (Munich) Germany, 2021.
- [12] “<http://www.touchence.jp/en/products/cube.html>,” 2-21-10 Kitaueno, Taito-ku, Tokyo 110-0014, 2021.