Learning peripersonal space in a humanoid robot and its application to safe human-robot interaction

Phuong D.H. Nguyen[1], Matej Hoffmann[2], Ugo Pattacini[1], Giorgio Metta[1]

Abstract— **The paper presents research to develop the peripersonal space (PPS) representation in robots through a self-supervised learning procedure, which is motivated by the development of perception and motor skills in humans. This representation is constructed by the integration of multisensory data from robots' sensors (stereo cameras, artificial skin and proprioception), and serves as spatial perception of the space surrounding the robot body. A novel approach is proposed to develop this representation through the design of specific motor activities that will make use of, *e.g.* motor babbling and reaching-with-avoidance. We will also show how this representation aims to help the robot accomplish motor tasks in complex situations, such as Human-robot Interaction (HRI). Finally, we will describe the accomplishments and future steps to complete the proposed plan.**

I. INTRODUCTION

The abilities to adapt and act autonomously in an unstructured and human-oriented environment are necessarily vital for the next generation of robots, which aim to safely cooperate with humans. While this adaptability is natural and feasible for humans, it is still very complex and challenging for robots.

Many neuroscientific findings show that there are multisensory integration processes occurring in humans to represent the space close to the body that is termed peripersonal space (PPS) [1]. The PPS serve as a "safety margin" to facilitate objects manipulation [2], [3] and to ease a variety of human actions such as reaching and locomotion with obstacle avoidance [2], [4]. Notably, this is not the case for the far space away from the human body [5]. Moreover, this spatial representation is incrementally trained and adapted (i.e. expanded, shrunk, enhanced, etc.) through motor activities, as reported in [1], [4], [6], and more.

Those results suggest that by exploiting motor activities in exploratory tasks, agents can on the one hand develop their perception of the space around their bodies, and on the other hand use the spatial representation they have built to improve the quality of their motor skills.

The goal of this research is to construct a PPS representation for the upper body of a humanoid robot by leveraging on the repertoire of its motor actions, and then to use such enhanced spatial perception to finally refine the motor capabilities of the robot, especially in cluttered and dynamic environments. Specifically, the proposed research aims to contribute to the understanding and propose models and implementations pertaining to the following points:

- Mechanisms of development and learning of PPS representation from visual, tactile, and proprioceptive information;
- The interaction of motor skills (such as reaching capabilities) and multimodal perception;
- The utility of new adaptive PPS representations in control settings – in particular planning and reaching with simultaneous obstacle avoidance.

The developed models and algorithms will then be validated on the iCub humanoid robot for human-robot interaction in a cluttered environment.

II. RELATED WORKS

Computational models: Serino *et al.* [6] and Maggoso *et al.* [7], [8] analyzed two neural networks to deal with audiotactile and visuotactile stimuli, respectively. They both suggest bio-inspired networks for the PPS representation, and then assign the connection weights that model the neuronal plasticity. The models were only tested without a body and only in a simple static scenario, assuming body parts to be still. Moreover, they have not designed a training procedure, except for the tool-use case presented in [8].

Robotics models: Roncone *et al.* [9] proposed a model to investigate an integrated representation of visual and tactile sensors. The outcome is a visual collision predictor of objects being close to a robot's body, which is constructed by visuo-tactile contingency. This model can be used for a simple reaching/avoidance controller. However, they rely on a well-structured visual tracker for data collection and *a priori* knowledge of a robot kinematic model for frame transformation (between different sensory sources) rather than via autonomous learning.

Antonelli *et al.* [10] and Chinellato *et al.* [11] adopted radial basis function networks to construct the mapping (forward and inverse transformations) between stereo visual data and proprioceptive data by performing gazing and reaching activities. Their mapping requires markers for feature extraction with known disparity, and is apparently beneficial only for multi-sensory transformation and not as a spatial perception of the body's surroundings.

On the other hand, Contla [12] focused on the plastic nature of PPS to account for the modification the body undergoes, and on the impact of this plasticity on the confidence levels of reaching activities. The hypothesis is validated only in a simulated environment. Contla's work

[1]Phuong D.H. Nguyen, Ugo Pattacini, and Giorgio Metta are with iCub Facility, Istituto Italiano di Tecnologia, Genova 16163, Italy {phuong.nguyen, ugo.pattacini, giorgio.metta}@iit.it

[2]Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic {matej.hoffmann@fel.cvut.cz

is mainly concerned with the reachable space of the robot, whereas we focus on the PPS as "margin of safety" instead (see Section I).

The above review makes evident that the current research is very little regarded with building a model of the PPS through self-supervised learning as well as its exploitation to enhance the robot motor capabilities.

III. GENERAL APPROACH

To tackle the research questions, we propose a general approach for the project as follows:

- We evaluate and extend the PPS model of Roncone *et al.* [9] for the Human-robot Interaction (HRI) scenario, where the learned PPS representation serves as a collision predictor against the visually detected obstacles and as an aggregation of physically detected collisions (via tactile sensors). This guarantees the safety for robot's interaction with environments. Also, a robot control system for interaction scenarios needs designing with a master motion planner and a controller;
- We introduce a modulation of the PPS representation for adaptive robot behavior. The modulation can result in expanding or shrinking the "safety margin" depending for example on the properties of the relevant objects in the scene (e.g. fragile, threatening) or on the social context of the interaction. As a result, the robot will be able to interact with human partners in a shared workspace according to different internal states (e.g., relaxed vs. stressed);
- We finally propose a novel PPS model utilizing a neural network to integrate multi-sensory information from the stereo-vision, distributed skin and proprioception, which aims to seamlessly substitute the model of Roncone *et al.* in HRI architecture. The main purpose of the alternative model is to overcome the limitations of the available one (*i.e.* based on *a priori* robot's kinematic model, using visual tracker), and to enable the autonomous action-based learning procedure.

IV. EXPERIMENTS & RESULTS

In this section, we briefly describe our accomplishments in realizing the final aim of the project.

A. Motion planning algorithm for robotic manipulators in dynamic environment

In [13], we present a fast heuristic motion planning algorithm designed for a humanoid robot that employs the sampling-based RRT* algorithm directly in the Cartesian space and in a hierarchical fashion: (i) a collision-free path is planned for the end-effector; (ii) corresponding collision-free points for every via-point are searched for the robot elbow. The method is then validated in diverse scenarios, comprising *batch run-time measurements*, *tests for asymptotic optimality* and *benchmarks against state-of-the-art*.

The results demonstrate that our solution delivers real-time performance (generates path plans in a fraction of second on a standard PC) in the vast majority of cases in a significantly cluttered environment. Second, the results suggest that asymptotic optimality of the plans is preserved even for the additional control points. Third, a comparison with state-of-the-art algorithms on the same scenario shows that solutions cannot be found in reasonable time (less than 10s) when using other algorithms.

This method was applied to the iCub in real settings in the frame of the EU Project WYSIWYD [1] where our method guaranteed collision-free for robots' motion in a table top scenario.

B. Compact real-time avoidance of a humanoid robot for Human-robot Interaction

Taking inspiration from PPS representations in humans, we present a framework on the iCub humanoid that dynamically maintains such a protective safety zone, composed of the following main elements: (i) a visual human 2D key-points estimation pipeline employing a deep learning based algorithm, extended into 3D using disparity; (ii) a distributed adaptive PPS representation around the robot's body parts, augmented from [9]; (iii) a visually reactive controller that incorporates all obstacles entering the robot's safety zone on the fly into the task (see [14]). The proposed solution is flexible and versatile since the safety zone around individual robot and human body parts can be selectively modulated (*e.g* stronger avoidance of the human head compared to rest of the body). Our system works in real time and is self-contained, with no external sensory equipment and use of onboard cameras only.

Pilot experiments in physical HRI scenario, *i.e.* reaching static target or following a trajectory with human experimenter interfering, demonstrate that an effective safety margin between the robot's and the human's body parts is kept during interaction.

C. Merging physical and social interaction for effective human-robot collaboration

We extended the work in [14] by designing a complete system in [15] (shown in Fig. 1) that merges elements of physical and social HRI, namely:

- A compact human-centered visual perception system for humanoid robots, which can detect human pose, and also recognize and track humans' manipulating objects;
- A simple symbolic "storage" of humans, objects, tools information to support social interaction, which contains the knowledge representations converted from perceived sensory representations of an environment;
- A visuo-tactile reactive controller that exploits the stereo-vision and the artificial skin of the iCub to allow the robot to safely react in both *pre*- and *post*-collision phases corresponding to visual and tactile stimuli respectively.

Through two interaction experiments (*i.e.* human-robot and robot-human object hand-over), we show that the complete system works in real-time controlling the robot's activities while guaranteeing safety for the human experimenter.

[1] wysiwyd.upf.edu

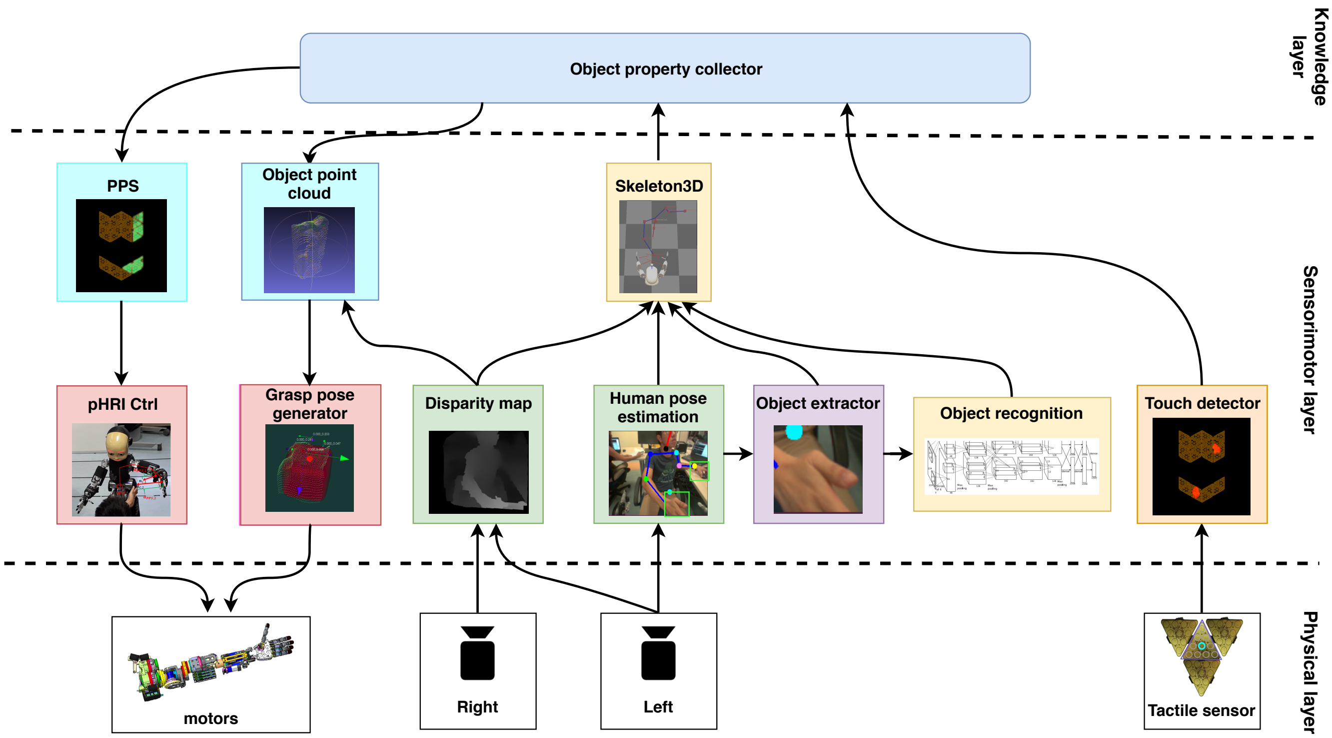

Fig. 1. Overview of the overall system comprising perception (right side) and action (left side) pathways.

The proposed visual perception system was also utilized to replace the wearable sensory suit for human tracking task in an ergonomic and reconfigurable Human-robot Collaboration [16]. The comparison results of tracking experiment (between our vision system and the wearable suit) prove the effectiveness and feasibility of our replacement for industrial application.

D. Learning visuomotor mapping in simulation and transferring to real world for robotics manipulation tasks

Recently, we design a framework to learn the visuomotor mapping in a single step [17] rather than considering the two problems (*i.e.* robot's kinematic modeling and visual-based pose estimation) independently and finding an offset mapping subsequently as in classical approach [18]. More specifically, we suggest to learn the mapping from an imprecise model in simulation using two components (as shown in Fig. 2): (i) A deep neural network (DNN) estimates the arm's joint configuration given images captured with the two eyes of the simulated robot and the corresponding head configuration. (ii) An image-to-image translation method bridges the domain gap to allow application of the DNN in the real world, since the image statistics between simulation and real world differ significantly.

In various experiments, we first show that the visuomotor predictor provides accurate joint estimates of the iCub's hand in simulation, and also can be used to obtain the systematic error of the robot's joint measurements on the physical iCub robot. We demonstrate that a calibrator can be designed to automatically compensate this error, and then validate that this enables accurate reaching of objects while circumventing manual fine-calibration of the robot.

V. CONCLUSIONS & FUTURE WORKS

In this paper, we have proposed a bio-inspired approach (i.e. learning via motor activities) to integrate the multi-sensory information (i.e. visual, tactile and proprioceptive) forming the spatial perception of surroundings for humanoid robots–peripersonal space representation, and to develop the sensorimotor competences from that enhanced perception. In addition, we have presented our achievements that consists in the design and realization of a *Multiple Cartesian point motion planning algorithm, Visuo-tactile control system for HRI* and *Visuomotor learning framework*, which were all successfully published ([13], [14], [17], [19], [20]) or submitted ([15], [16]).

The successive step will be concerned with extending the visuomotor mapping model [17] to additionally incorporate the tactile input. An action based learning process will also rely on our proposed *motor babbling* method [17], extended such that it can deal with a cluttered environment with randomly allocated obstacles. The simulation environment will be mainly exploited for data collection due to the safety, while domain adaptation methods like domain randomization, image-to-image translation will be used for bridging the reality gap. The resulting model will be used to estimate the spatial and temporal information of possible collisions of the robot's arm with visually detected objects, such that robot's collision-free motion planning can be generated. An advantage of the proposed method is that visual stimuli can be mapped directly into joint space in real-time, where well-established motion planning techniques such as Rapidly exploring Random Trees (RRT*) and Probabilistic RoadMap (PRM*) [21] can be applied.

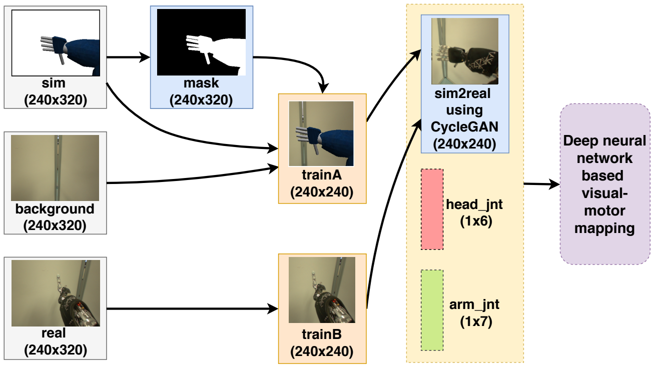

Fig. 2. Overview of the overall learning framework. Images obtained using a simulator are first being implanted with real background, and then CycleGAN [22] is used to synthesize realistically looking "sim2real" images. These are used as inputs to a deep neural network along with the head joints obtained from the simulator. The aim of the deep network is to estimate the arm joint configuration.

ACKNOWLEDGMENT

Phuong D.H. Nguyen was supported by a Marie Curie Early Stage Researcher Fellowship (H2020-MSCA-ITA, SECURE 642667). Matej Hoffmann was supported by the Czech Science Foundation under Project GA17-15697Y.

REFERENCES

[1] J. Cléry *et al.*, "Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: Knowns and unknowns," *Neuropsychologia*, vol. 70, pp. 313–326, Apr. 2015.

[2] N. P. Holmes *et al.*, "The body schema and multisensory representation(s) of peripersonal space," *Cognitive Processing*, vol. 5, no. 2, pp. 94–105, Jun. 2004.

[3] C. Goerick *et al.*, "Peripersonal space and object recognition for humanoids," in *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, IEEE, 2005, pp. 387–392.

[4] E. Làdavas *et al.*, "Action-dependent plasticity in peripersonal space representations," *Cognitive Neuropsychology*, vol. 25, no. 7-8, pp. 1099–1113, Dec. 2008.

[5] A. Farne *et al.*, "Neuropsychological evidence of modular organization of the near peripersonal space," *Neurology*, vol. 65, no. 11, pp. 1754–1758, 2005.

[6] A. Serino *et al.*, "Extending peripersonal space representation without tool-use: Evidence from a combined behavioral-computational approach," *Frontiers in Behavioral Neuroscience*, vol. 9, Feb. 2015.

[7] E. Magosso *et al.*, "Visuotactile representation of peripersonal space: A neural network study," *Neural computation*, vol. 22, no. 1, pp. 190–243, 2010.

[8] E. Magosso *et al.*, "Neural bases of peri-hand space plasticity through tool-use: Insights from a combined computational–experimental approach," *Neuropsychologia*, vol. 48, no. 3, pp. 812–830, 2010.

[9] A. Roncone *et al.*, "Peripersonal Space and Margin of Safety around the Body: Learning Visuo-Tactile Associations in a Humanoid Robot with Artificial Skin," *PLOS ONE*, vol. 11, no. 10, e0163713, 2016.

[10] M. Antonelli *et al.*, "On-line learning of the visuomotor transformations on a humanoid robot," in *Intelligent Autonomous Systems 12*, Springer, 2013, pp. 853–861.

[11] E. Chinellato *et al.*, "Implicit Sensorimotor Mapping of the Peripersonal Space by Gazing and Reaching," *EEE Trans. Auton. Ment. Dev.*, vol. 3, no. 1, pp. 43–53, Mar. 2011.

[12] S. Ramírez Contla, "Peripersonal Space in the Humanoid Robot iCub," 2014.

[13] P. D. Nguyen *et al.*, "A fast heuristic Cartesian space motion planning algorithm for many-DoF robotic manipulators in dynamic environments," in *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*, IEEE, 2016, pp. 884–891.

[14] D. H. P. Nguyen *et al.*, "Compact Real-time Avoidance on a Humanoid Robot for Human-robot Interaction," in *The 2018 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 2018, pp. 416–424.

[15] P. D. Nguyen *et al.*, "Merging physical and social interaction for effective human-robot collaboration," in *Humanoid Robots (Humanoids), 2018 IEEE-RAS 18th International Conference on*, (submitted), 2018.

[16] W. Kim *et al.*, "A Reconfigurable and Adaptive Human-Robot Collaboration Framework for Improving Worker Ergonomics and Productivity," *IEEE Rob. Autom. Mag.*, 2018, (under review).

[17] P. D. H. Nguyen *et al.*, "Transferring Visuomotor Learning from Simulation to the Real World for Robotics Manipulation Tasks," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2018.

[18] J. Hollerbach *et al.*, "Model Identification," in *Springer Handbook of Robotics*, B. Siciliano *et al.*, Eds., 2008, pp. 321–344.

[19] T. Fischer *et al.*, "iCub-HRI: A Software Framework for Complex Human–Robot Interaction Scenarios on the iCub Humanoid Robot," *Frontiers in Robotics and AI*, vol. 5, p. 22, 2018.

[20] C. Moulin-Frier *et al.*, "DAC-h3: A Proactive Robot Cognitive Architecture to Acquire and Express Knowledge About the World and the Self," *IEEE Transactions on Cognitive and Developmental Systems*, vol. PP, no. 99, 2017.

[21] S. Karaman *et al.*, "Sampling-based algorithms for optimal motion planning," *International Journal of Robotics Research*, vol. 30, no. 7, pp. 846–894, 2011.

[22] J.-Y. Zhu *et al.*, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," in *IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.