Robot Arms in Action: Manipulation, Interaction, and Perception

Alap Kshirsagar
Technische Universität Darmstadt
Email: alap@robot-learning.de

Humans are the most intelligent of animals because they have hands - Anaxagoras (Greek Philosopher, 500-428 BC). Humans use their arms to manipulate objects, interact with each other, and perceive through touch. Robots with similar capabilities can automate repetitive and hazardous tasks, assist and collaborate with humans, and contribute to enhanced productivity across diverse industries. Thus, I am interested in the overarching research question: How can robots effectively utilize their arms for **manipulation**, **interaction**, and **perception**? In this research statement, I describe my previous, ongoing and future research work aimed at these three desired functionalities of robotic arms.

I. Research To Date

A. Manipulation and Interaction

Human-robot collaborative manipulations (HRCM), for example, object handovers and handshakes, lie at the intersection of **manipulation** and **interaction** functions of robotic arms. In my research, I have explored techniques from optimal control, formal methods, imitation learning, and reinforcement learning for HRCM.

Timing-specified controllers for handovers: One of the most common HRCM is an object handover. The importance of human-robot handovers has resulted in many handover controllers proposed in the literature [1]. However, the existing controllers do not provide guarantees on the timing of a handover. Such timing guarantees may be crucial in productivity-oriented industrial tasks and fast-paced life-critical scenarios like surgery. To address this issue, we proposed a controller [2] for human-robot handovers automatically synthesized from formal specifications in Signal Temporal Logic. Further, we developed receding-horizon controllers that allow users to specify timing parameters for a handover and provide feedback if the robot cannot satisfy those constraints [3]. These features make such controllers useful for settings where end-users need to tune the robot's behavior without programming knowledge.

Learning HRCM from human demonstrations: Learning HRCM from human demonstrations allows non-expert users to teach robots and produce humanlike behaviors desirable for humanoid robots. Several approaches have been proposed for generating humanlike robot behavior from human demonstrations of unimanual handovers [4–12]. However, transferring large, deformable, or delicate objects requires bimanual handovers. There is very little work on bimanual human-robot handovers [13–16], and these works do not address humanlike motion generation. To facilitate research on humanlike bimanual handovers, we built the first public datasets of bimanual human-to-human handovers [17] and multiple sequential human-to-human handovers [18] involving unimanual, bimanual, and self-handovers. We proposed a framework [19] consisting of a Hidden Markov Model (HMM) and convex optimization to generate humanlike robot reaching motions while adhering to constraints such as maintaining a constant grip width in robot-to-human bimanual handovers.

Similar to object handovers, many collaborative manipulations, such as handshakes, fistbumps, object transport, and collaborative assembly, can naturally be broken down into underlying segments sequenced to achieve suitable behavior. We demonstrated [20] that incorporating a separate HMM over observations at the transition states – the states at the boundaries of two segments of an interaction – enhances the segmentation abilities of HMMs. However, this approach and several other approaches for learning HRCM from human demonstrations [21–24] require training a separate model for each type of interaction. To learn multiple HRCM from human demonstrations with a single model, we proposed a framework [25] consisting of a Recurrent Mixture Density Network (RMDN) and a Variational Autoencoder (VAE). The RMDN encodes the latent trajectory of the human partner and regularizes the robot embeddings in the VAE. The RMDN+VAE model was then used to generate reactive robot motion, achieving better results than methods that use either modular latent space dynamics using HMM [24] or LSTM-based latent dynamics models [26], on a variety of interactions – handwaves, handshakes, fistbumps, and handovers.

Learning HRCM with reinforcement learning: Reinforcement learning (RL) has been successfully demonstrated for autonomous manipulation and locomotion tasks [27, 28]. However, there is limited work on learning HRCM with RL due to the sample-efficiency and safety requirements of HRCM. We evaluated a sample-efficient RL algorithm, "Guided Policy Search" (GPS) [29], for generating a robot's reaching motions in handovers [30, 31]. The key findings were that the learned policy had good generalizability over changes in the object's mass but poor generalizability over changes in the handover location. We also reviewed [32] existing safe robot reinforcement learning (SRRL) methods and argued that interactive behaviors need more attention from the SRRL community to achieve SRRL in human-centered environments.

Non-verbal communication in HRCM: Nonverbal cues

such as gaze behaviors are used to support shared workspace collaboration [33]. These gestures are important for coordinating tasks and establishing common ground relative to task states and the internal states of the collaborators. In a series of studies [34, 35], we investigated human-inspired robot receivers' gaze behaviors in human-to-robot handovers. Our results revealed that, for both observers and participants in a handover, when the robot exhibited a *Face-Hand-Face* gaze (gazing at the giver's face and then at the giver's hand during the reach phase and back at the giver's face during the retreat phase), participants considered the handover to be more likable, anthropomorphic, and communicative of timing. We did not find evidence of any effect of the object's size or fragility or the giver's posture on the gaze preference. These findings could help the design of non-verbal cues in HRCM.

B. Perception

Besides manipulation and interaction, robots can utilize their arms for **perception**, i.e., to understand and explore their environment through touch. My research has focused on utilizing vision-based tactile sensors (VBTS) like GelSight Mini [36] to perceive object properties. VBTS provide a cost-effective and high-resolution alternative to traditional tactile sensors, capitalizing on advancements in camera technology and computer vision.

Active texture recognition: Various applications, such as laundry separation, waste sorting, and material handling, can benefit from the rapid classification of textures. Classification of fabrics has been tackled with different types of sensors using both supervised and active methods [37–41]. Our study [42] focused on achieving sample-efficient texture recognition using VBTS. We compared two information-theoretic active exploration strategies aimed at minimizing predictive entropy or variance of probabilistic classifiers. Additionally, we assessed neural network architectures, uncertainty representations, data augmentation impact, and dataset variability for tactile texture recognition. We found that the active exploration strategy had a minor influence on texture recognition accuracy, with data augmentation and dropout rate playing more significant roles. In a comparative study, our best approach achieved a 90.0% accuracy in under five touches, surpassing the 66.9% accuracy achieved by humans. This finding underscores the effectiveness of VBTS in texture recognition for practical applications.

Object hardness classification: The ability to detect an object's hardness can aid robots in performing tasks like waste sorting, automated harvesting, and handling delicate objects in household tasks. Prior works on tactile hardness estimation [43, 44] used large datasets with ground-truth Shore hardness values to train a deep neural network to estimate the Shore hardness of objects. Humans, however, perceive hardness not as an absolute value but in comparison with other objects. Inspired by how humans perceive hardness, we investigated whether a robotic manipulator equipped with VBTS can learn to classify objects of different hardnesses without having access to actual Shore hardness values. We found that the CNN-LSTM architecture used in [43] trained with a categorical cross-entropy loss achieved $76 \pm 4\%$ accuracy in discerning the hardness similarity of untrained hardness classes to the trained classes, with 20 touches per object. Human participants achieved $80 \pm 9\%$ accuracy on the same task in our study [45], requiring 1 to 20 touches. Our ongoing work seeks to improve the VBTS' hardness similarity detection using active exploration.

In-hand object pose estimation: Accurate in-hand object pose estimation is crucial for robotic manipulation and assembly tasks. In this ongoing work, we combine two VBTS and an RGB-D camera to estimate the 6D pose of the object grasped by the robotic arm. VBTS offer resilience to visual occlusion inherent in in-hand manipulation, while the camera provides a broader field of view. Unlike existing deep-neural-network-based in-hand pose estimation methods [46–48], which require prior training datasets, we are exploring an approach that requires only the 3D model of the object, using a weighted Iterative Closest Point (ICP) algorithm, dynamically evaluating and adapting contributions from each sensor modality.

II. Future Work

Human-Robot Collaborative Manipulation: In the future, I plan to develop robot controllers for dynamic HRCM skills like human-robot partner juggling. Further, to facilitate the development and evaluation of algorithms and systems for HRCM, I propose to create a simulation framework and benchmarks for several physical human-robot interaction tasks in industrial and domestic environments. I also plan to investigate non-verbal communication in HRCM, specifically in social interactions such as handshakes and fistbumps.

Tactile Perception: Humans use specific "exploratory procedures (EPs)" [49] to identify object properties, for example, lateral motion to detect texture and pressure to detect deformability. I plan to investigate whether these exploratory procedures are optimal for robotic arms equipped with tactile sensors or whether robots can learn – with reinforcement learning – exploratory procedures better suited for their specific type of tactile sensors.

Bi-manual Visuo-tactile Manipulation: Robotic grasping and pouring are two fundamental visuo-tactile manipulation skills and have been an area of active research for several decades. However, bi-manual grasping and pouring with different types of robotic grippers remain open challenges. I seek to investigate methods for learning gripper-aware policies for bi-manual robotic grasping and pouring with visuo-tactile input.

Co-optimization of Hand and Action: Finally, going back to Anaxagoras' quote at the beginning of this research statement, there is a strong interdependence between hand design and control policy for manipulation. Considering this interdependence, I aim to develop methods for co-optimizing robot hand design and manipulation policy. This goal can be achieved by leveraging the recent advances in massive parallel physics simulations and numerical optimization by reinforcement learning.

References

[1] Valerio Ortenzi, Akansel Cosgun, Tommaso Pardi, Wesley P Chan, Elizabeth Croft, and Dana Kulić. Object handovers: a review for robotics. *IEEE Transactions on Robotics*, 37(6):1855–1873, 2021.

[2] Alap Kshirsagar, Hadas Kress-Gazit, and Guy Hoffman. Specifying and synthesizing human-robot handovers. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5930–5936, 2019.

[3] Alap Kshirsagar, Rahul Kumar Ravi, Hadas Kress-Gazit, and Guy Hoffman. Timing-specified controllers with feedback for human-robot handovers. In *IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1313–1320, 2022.

[4] Miguel Prada, Anthony Remazeilles, Ansgar Koene, and Satoshi Endo. Implementation and experimental validation of Dynamic Movement Primitives for object handover. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014.

[5] Guilherme Maeda, Marco Ewerton, Rudolf Lioutikov, Heni Ben Amor, Jan Peters, and Gerhard Neumann. Learning interaction for collaborative tasks with probabilistic movement primitives. In *IEEE-RAS International Conference on Humanoid Robots*, November 2014.

[6] Guilherme J. Maeda, Gerhard Neumann, Marco Ewerton, Rudolf Lioutikov, Oliver Kroemer, and Jan Peters. Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks. *Autonomous Robots*, 41 (3):593–612, March 2016.

[7] Andras Kupcsik, David Hsu, and Wee Sun Lee. Learning dynamic robot-to-human object handover from human feedback. *Robotics Research*, 1:161, 2017.

[8] José R Medina, Felix Duvallet, Murali Karnam, and Aude Billard. A human-inspired controller for fluid human-robot handovers. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 324–331, 2016.

[9] Antonis Sidiropoulos, Efi Psomopoulou, and Zoe Doulgeri. A human inspired handover policy using gaussian mixture models and haptic cues. *Autonomous Robots*, 43 (6):1327–1342, 2019.

[10] Katsu Yamane, Marcel Revfi, and Tamim Asfour. Synthesizing object receiving motions of humanoid robots with human motion database. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1629–1636, 2013.

[11] Xuan Zhao, Sakmongkon Chumkamon, Shuanda Duan, Juan Rojas, and Jia Pan. Collaborative human-robot motion generation using LSTM-RNN. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, November 2018.

[12] Wei Yang, Chris Paxton, Maya Cakmak, and Dieter Fox. Human grasp classification for reactive human-to-robot handovers. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[13] Aaron M. Bestick, Samuel A. Burden, Giorgia Willits, Nikhil Naikal, S. Shankar Sastry, and Ruzena Bajcsy. Personalized kinematics for human-robot collaborative manipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 2015.

[14] Seyed Sina Mirrazavi Salehian, Nadia Figueroa, and Aude Billard. Coordinated multi-arm motion planning: Reaching for moving objects in the face of uncertainty. In *Proceedings of Robotics: Science and Systems*, 2016.

[15] Seyed Sina Mirrazavi Salehian, Nadia Figueroa, and Aude Billard. A unified framework for coordinated multi-arm motion planning. *The International Journal of Robotics Research*, 37(10):1205–1232, April 2018.

[16] Wei He, Jiashu Li, Zichen Yan, and Fei Chen. Bidirectional human-robot bimanual handover of big planar object with vertical posture. *IEEE Transactions on Automation Science and Engineering*, 2021.

[17] Alap Kshirsagar, Raphael Fortuna, Zhiming Xie, and Guy Hoffman. Dataset of bimanual human-to-human object handovers. *Data in Brief*, 48:109277, 2023.

[18] Alap Kshirsagar, Raphael Fortuna, Zhiming Xie, and Guy Hoffman. Multi-sensor dataset of multiple sequential human-to-human object handovers in shelving and un-shelving tasks, 2023. URL https://doi.org/10.5281/zenodo.7895500.

[19] Yasemin Göksu, Antonio De Almeida Correia, Vignesh Prasad, Alap Kshirsagar, Dorothea Koert, Jan Peters, and Georgia Chalvatzaki. Kinematically constrained human-like bimanual robot-to-human handovers. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 497–501, 2024.

[20] Fabian Hahne, Vignesh Prasad, Alap Kshirsagar, Dorothea Koert, Ruth Maria Stock-Homburg, Jan Peters, and Georgia Chalvatzaki. Transition state clustering for interaction segmentation and learning. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 512–516, 2024.

[21] Sylvain Calinon, Paul Evrard, Elena Gribovskaya, Aude Billard, and Abderrahmane Kheddar. Learning collaborative manipulation tasks by demonstration using a haptic interface. In *International Conference on Advanced Robotics*, pages 1–6, 2009.

[22] Heni Ben Amor, Gerhard Neumann, Sanket Kamthe, Oliver Kroemer, and Jan Peters. Interaction primitives for human-robot cooperation tasks. In *IEEE international conference on robotics and automation (ICRA)*, pages 2831–2837, 2014.

[23] Joseph Campbell and Heni Ben Amor. Bayesian interaction primitives: A slam approach to human-robot interaction. In *Conference on Robot Learning*, pages 379–387, 2017.

[24] Vignesh Prasad, Dorothea Koert, Ruth Stock-Homburg, Jan Peters, and Georgia Chalvatzaki. Mild: Multimodal interactive latent dynamics for learning human-robot interaction. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 472–479, 2022.

[25] Vignesh Prasad, Alap Kshirsagar, Dorothea Koert Ruth Stock-Homburg, Jan Peters, and Georgia Chalvatzaki. Moveint: Mixture of variational experts for learning human-robot interactions from demonstrations. *IEEE Robotics and Automation Letters*, 2024.

[26] Judith Bütepage, Ali Ghadirzadeh, Özge Öztimur Karadağ, Mårten Björkman, and Danica Kragic. Imitating by generating: Deep generative models for imitation of interactive tasks. *Frontiers in Robotics and AI*, 7:47, 2020.

[27] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013.

[28] Bharat Singh, Rajesh Kumar, and Vinay Pratap Singh. Reinforcement learning in robotic applications: a comprehensive survey. *Artificial Intelligence Review*, pages 1–46, 2022.

[29] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.

[30] Alap Kshirsagar, Guy Hoffman, and Armin Biess. Evaluating guided policy search for human-robot handovers. *IEEE Robotics and Automation Letters*, 6(2):3933–3940, 2021.

[31] Alap Kshirsagar, Tair Faibish, Guy Hoffman, and Armin Biess. Lessons learned from utilizing guided policy search for human-robot handovers with a collaborative robot. In *International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*, pages 52–57, 2022.

[32] Shangding Gu*, Alap Kshirsagar*, Yali Du*, Guang Chen, Jan Peters, and Alois Knoll. A human-centered safe robot reinforcement learning framework with interactive behaviors. *Frontiers in Neurorobotics*, 17, 2023.

[33] B Hayes and B Scassellati. Challenges in shared-environment human-robot collaboration. acm. In *IEEE International Conference on Human-Robot Interaction*, 2013.

[34] Alap Kshirsagar, Melanie Lim, Shemar Christian, and Guy Hoffman. Robot gaze behaviors in human-to-robot handovers. *IEEE Robotics and Automation Letters*, 5(4): 6552–6558, 2020.

[35] Tair Faibish*, Alap Kshirsagar*, Guy Hoffman, and Yael Edan. Human preferences for robot eye gaze in human-to-robot handovers. *International Journal of Social Robotics*, 14(4):995–1012, 2022.

[36] GelSight. GelSight Mini System, April 2023. URL https://www.gelsight.com/product/gelsight-mini-system. [Online; accessed 1. Feb. 2023].

[37] Tasbolat Taunyazov, Yansong Chua, Ruihan Gao, Harold Soh, and Yan Wu. Fast texture classification using tactile neural coding and spiking neural network. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9890–9895, 2020.

[38] Ruihan Gao, Tian Tian, Zhiping Lin, and Yan Wu. On explainability and sensor-adaptability of a robot tactile texture representation using a two-stage recurrent networks. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1296–1303, 2021.

[39] Shiyao Huang and Hao Wu. Texture recognition based on perception data from a bionic tactile sensor. *Sensors*, 21(15):5224, 2021.

[40] Rui Li and Edward H Adelson. Sensing and recognizing surface textures using a gelsight sensor. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1241–1247, 2013.

[41] Guanqun Cao, Yi Zhou, Danushka Bollegala, and Shan Luo. Spatio-temporal attention model for tactile texture recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9896–9902, 2020.

[42] Alina Böhm, Tim Schneider, Boris Belousov, Alap Kshirsagar, Lisa Lin, Katja Doerschner, Knut Drewing, Constantin A Rothkopf, and Jan Peters. What matters for active texture recognition with vision-based tactile sensors. In *IEEE International Conference on Robotics and Automation*, 2024.

[43] Wenzhen Yuan, Chenzhuo Zhu, Andrew Owens, Mandayam A Srinivasan, and Edward H Adelson. Shape-independent hardness estimation using deep learning and a gelsight tactile sensor. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 951–958, 2017.

[44] Yaohui Chen, Jiahao Lin, Xuan Du, Bin Fang, Fuchun Sun, and Shanjun Li. Non-destructive fruit firmness evaluation using vision-based tactile information. In *International Conference on Robotics and Automation (ICRA)*, pages 2303–2309, 2022.

[45] L. Lin, A. Boehm, B. Belousov, A. Kshirsagar, T. Schneider, K. Peters, J. Doerschner, and K. Drewing. Task-adapted single-finger explorations of complex objects. *Eurohaptics Conference*, 2024.

[46] Tomoki Anzai and Kuniyuki Takahashi. Deep gated multi-modal learning: In-hand object pose changes estimation using tactile and image data. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9361–9368, 2020.

[47] Snehal Dikhale, Karankumar Patel, Daksh Dhingra, Itoshi Naramura, Akinobu Hayashi, Soshi Iba, and Nawid Jamali. Visuotactile 6d pose estimation of an in-hand object using vision and tactile sensor data. *IEEE Robotics and Automation Letters*, 7(2):2148–2155, 2022.

[48] Yuan Gao, Shogo Matsuoka, Weiwei Wan, Takuya Kiyokawa, Keisuke Koyama, and Kensuke Harada. In-hand pose estimation using hand-mounted rgb cameras and visuotactile sensors. *IEEE Access*, 11:17218–17232, 2023.

[49] Susan J Lederman and Roberta L Klatzky. Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, 71(7):1439–1459, 2009.