APPRENTICE: Towards a Cobot Helper in Assembly Lines

Yunus Terzioğlu[1], Özgür Aslan[2], Burak Bolat[2], Batuhan Bal[2], Tuğba Tümer[2], Fatih Can Kurnaz[2], Sinan Kalkan[2], Erol Şahin[2]

***Abstract*— In this paper, we describe a vision of human-robot collaboration on assembly lines, where a collaborative robotic manipulator, a.k.a. cobot, operates as a work-mate for the worker. Specifically, the cobot, referred to as the "apprentice" since it lacks the ability to replace the worker, can only aid the worker by observing his status during the assembly process, and handing over (and stowing away when done) the tools and the parts. Towards this end, we first describe the vision, and outline the challenges involved in developing an "apprentice" cobot and share the current state of the work done.**

I. Introduction

Collaborative robotic manipulators (a.k.a. cobots), designed to work safely alongside humans, are envisioned to take industrial automation to the next level, to increase the efficiency of a human worker. With the projected worldwide cobot market size to grow 5 times between 2020 and 2025 [1], these cobots are expected to take part in frequently changing tasks, mainly in medium and small businesses. This brings in many challenges pertaining to (i) how easily such robots can be used and how much programming and design they require, (ii) how user-friendly and helpful they are, and (iii) how appealing they are as co-workers.

II. The Apprentice Vision for Human-Robot Collaboration

The ÇIRAK and its successor KALFA (Apprentice and Journeymen in Turkish) projects are based on the observation that human workers are superior to cobots in the tasks involving manipulation. Carrying out the versatile manipulation tasks of an assembly-line worker at the same speed and finesse is likely to remain beyond the reach of cobots in the near future. In our projects, we propose to develop technologies towards an "apprentice cobot", an under-skilled robotic helper that can track the state of the worker, the assembly process and hand over (and stow away when done) the necessary tools and the parts to the worker for the current state of the process. Such a system would only require limited manipulative capabilities given that it is coupled with a cognitively intelligent interaction with the human worker.

The "apprentice cobot" vision can be seen in its trailer video (Figure 1) where the human assembles an IKEA chair with a cobot as his apprentice. In the snapshots of this video apprentice cobot is; (a) waiting attentively with breathing animation (can be seen in video and is not captured in the snapshot) for the start of the assembly, (b) handing over the screwdriver to the worker, (c) attentively following the worker's assembly through mutual gaze, (d) leaning back and waving (can be seen in the video and is not captured in the snapshot) to refuse to stow away the screwdriver since the worker still has work to do with the screwdriver, (e) reaching out to take the screwdriver once current job involving a screwdriver is completed, (f) stowing away the screwdriver.

The realization of such an apprentice cobot requires;

- improved human-robot interaction skills through the use of non-verbal behaviors,
- perception abilities to track (1) human body pose and gaze direction, (2) tools and parts in the workspace,
- awareness of the status of the assembly process,
- ability to discover assembly sequences,
- real-time motion planning in free space, and
- guarantees on safety.

In the rest of the paper, we will discuss the challenges towards implementing these capabilities and report briefly our results.

III. Human-Robot Interaction using non-verbal behaviors

Drawing on the character animation principles Appeal, Arcing, and Secondary Action, we designed a set of social cues for a commercially popular cobot platform, a UR5 robot arm (Universal Robots, Odense, Denmark) equipped with a 2F-140 two-finger gripper (Robotiq, Lévis, Canada) (see Figure 1) that included giving it a head-on-neck look by augmenting its appearance and implementing gaze and posture cues (Appeal), generating smooth motion trajectories for the arm (Arcing), and introducing breathing motions to the robot during its idle operation (Secondary Action).

In the ÇIRAK project, it was shown that applying some of Disney's animation principles to a Cobot improves the quality of human-robot interaction (HRI) [2]. The KALFA project will advance these proof-of-concept works in three directions to develop a full non-verbal communication ability in cobots: (1) After evaluating how all of Disney's animation principles can be applied to improve the quality of HRI, these principles will be formally defined and integrated into cobots as parameterized "HRI filters". (2) Methods for detecting non-verbal communication cues of the workers in the assembly scenarios will be developed. (3) Non-verbal communication cues from workers will be associated with HRI-filters in order to increase the harmony between cobot

[1] Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA terzioglu.y@northeastern.edu

[2] KOVAN Research Lab., ROMER (Center for Robotics and AI Technologies) and Dept. of Computer Eng., Middle East Technical University, Turkey {aslan.ozgur, burak.bolat, balbatuhan, tugba.tumer, fatih.kurnaz, skalkan, erol}@metu.edu.tr

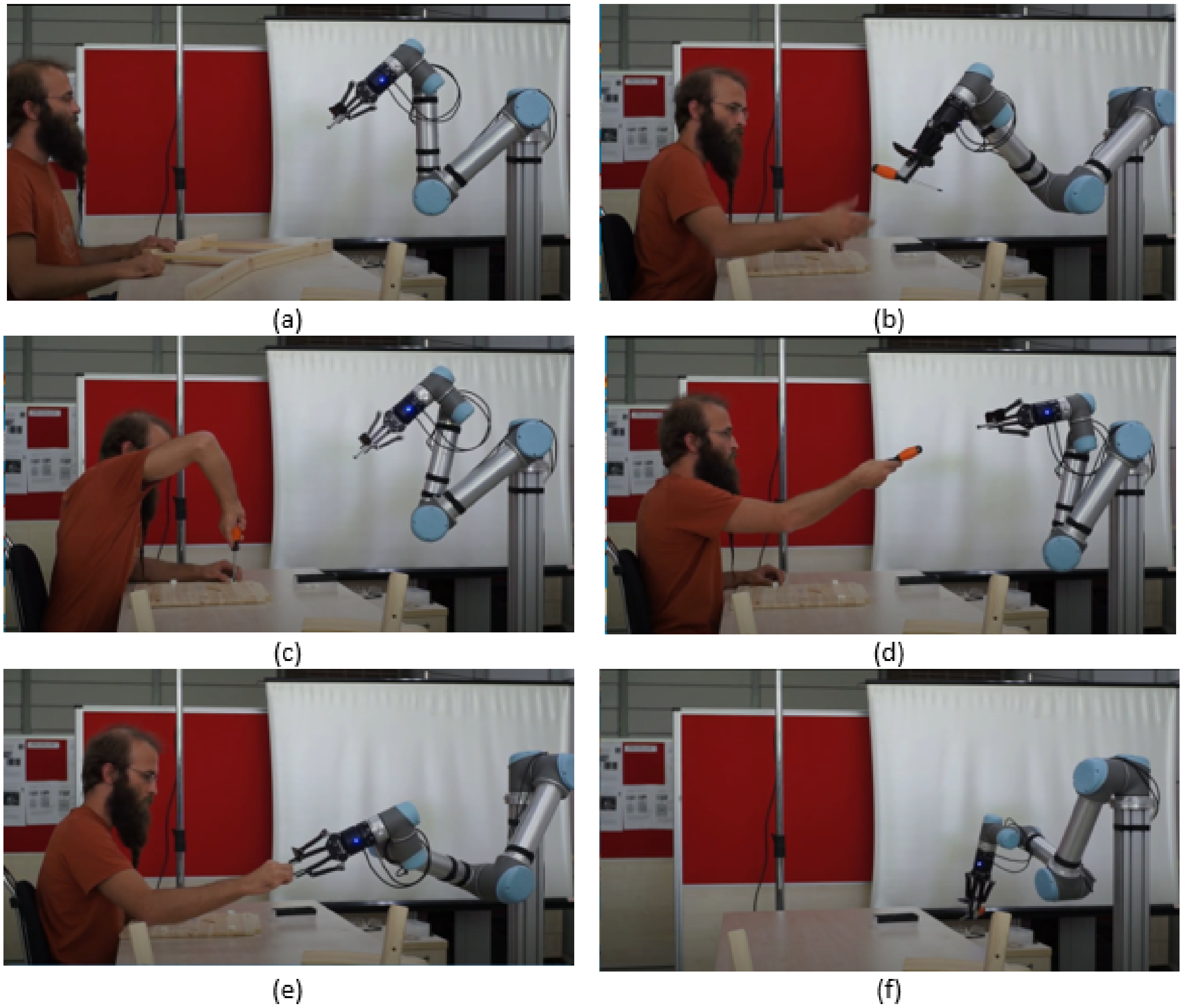

Fig. 1. Snapshots from the apprentice trailer video (at `https://youtu.be/CsV363jeuJs`). See the text for description.

and the worker. The effects of these methods on the quality of HRI will be measured by human-robot experiments.

A. *Related work*

Nonverbal communication, allowing transfer of information via social clues via facial expressions, gestures, body language etc., is very essential for human-robot interaction [3]. Its importance has been identified in many studies. For example, Salem et al. stated that regardless of the gesture congruence, arm gestures lead to a more sympathetic, lively, active and engaged interpretation of a robot [4]. Moreover, extroverted and abrupt gestures increase engagement and human awareness [5], [6]. Arm gestures are also investigated under synchronization congruence in human-robot interactions. Shen et al. reported adaptive robot velocity with respect to interacting participants increases gesture recognition, task performance and social interaction [7]. In other studies, it is found that participants synchronize their frequency of gestures to the robot's gesture frequency, unlike their phase difference [8], [9]. Gaze has a major effect in Kinesics. Stanton and Stevens revealed the impact of gaze to be task-dependent [10]. Proxemics studies the usage of the interaction space. Interactions with closer distance than the human-human cases were observed [11], [12], [13]. Time perception and manipulation of perceived time studied by Komatsu and Yamada [14]. Song and Yamada analyzed color, sound, vibration and their combinations [15] in social robotics.

IV. Perception

A robotic co-worker needs to be able to perceive humans, their gaze (and intention), objects parts, tools, and other utilities in the environment. This requires locating such objects in the camera frames in general by placing bounding boxes around them using deep object detectors. However, this is often insufficient as robotic control takes place in 3D. Therefore, perception of humans, objects, parts, tools, etc. needs to finally provide 3D information in the robot's 3D coordinate frame.

Perception incurs many challenges: (i) Obtained 3D information needs to be very precise since, otherwise, the assembly will fail. (ii) Perception should be robust to clutter,

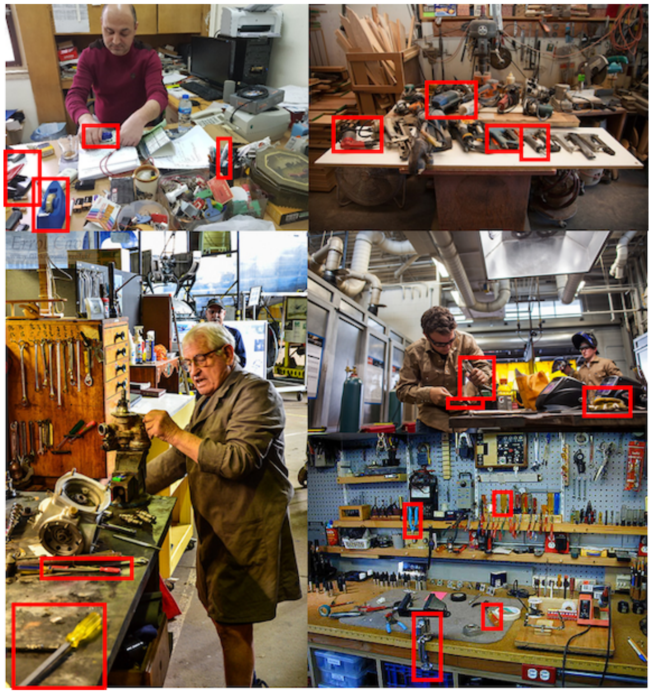

Fig. 2. An important problem in robotic assembly is perceiving the environment. However, common workplaces are unstructured and pose challenges to existing perception methods. [16])

disorganized environment, illumination conditions (e.g. the challenging conditions in Figure 2). (iii) Perception should be aware of ambiguous cases and should be explainable.

In the project, we constructed a perception pipeline for detecting tools, objects, humans, and their gazes with the help of deep learning applied on RGB-D video streams. Although existing object detectors, pose estimation, and human detection models, gaze estimation networks are very capable, they are not as robust as advertised on benchmarks and therefore, they often require tuning or combination with disambiguating sources of information, e.g. the task knowledge, any other type of contextual information, temporal consistency etc. In addition, the learning-based methods which have been trained on datasets that do not follow the characteristics of working environments that we consider, and therefore, the collection of labeled datasets is often necessary. For instance, in our project, we curated a tool detection dataset [16] specifically purposed for detecting tools in human-robot collaboration settings

V. Assembly

Obtaining a precise assembly plan is a labor intensive, tedious task. Despite efforts in the literature for learning how to combine parts to assemble the final object in 3D [17], [18], [19], these efforts are very limited to toy settings (in terms of objects and environments) and there remain crucial open issues: (i) It is still an open issue to assemble an object by looking at e.g. the IKEA assembly manual. (ii) Moreover, learning to do the assembly with a human co-worker and/or other robots has not been addressed. (iii) In addition, detecting whether there has been an error in the assembly, predicting the error isolating its source, rectifying the error are highly necessary for widespread applicability and acceptance of such robots.

In our project, we are focusing our efforts along two directions:

Automating assembly plan creation using Deep Reinforcement Learning: In the ÇIRAK project, a precise assembly plan was manually prepared for step-by-step execution of actions. This plan included the parts, the tools and the details like which tools should be used on what parts at which step of the assembly sequence. In the KALFA project we proposed to learn the assembly plan using Deep Reinforcement Learning by interacting with the parts and the tools within the cobot's simulation environment, thus facilitating the use of cobot in assembly scenarios by people with little technical skills.

Determining the sources of errors using a causality model: In the ÇIRAK project and similar studies can detect an error during the installation by comparing the current state of the assembly with the previously defined steps in a plan. However, they cannot detect the source of the errors or determine the steps that should be taken in order not to repeat the errors. The KALFA Project proposes to learn a causal model in the simulation environment from interplay between parts, tools, factors and assembly stages, and to use this causal model to determine the sources of errors when an anomaly is detected.

A. Related work

An important problem in robotic assembly is the precise generation of the assembly plan, which necessitates 3D perception of object parts and how those parts should be manipulated to perform the assembly. Advances in machine learning have paved the way for addressing this tedious task by directly learning the assembly from the 3D models of the parts and the target object, e.g. for furniture assembly problem [17], [18], [19]. For example, Li et al. [17] proposed two network modules to extract information from the image of the assembled furniture and part point clouds. Moreover, Huang et al. [18] introduced a dynamic graph learning framework to make predictions only from the part point clouds. In a similar line of approach, in Li et al.'s work [19], learned relationships are used to predict the position and the scale of parts instead of their poses (position and orientation).

For an assembly task, another important challenge is precise manipulation, which is hard to achieve using standard controllers. To this end, the trending approach is using Reinforcement Learning (RL) to learn adaptive control strategies. Due to the complexity of the assembly task, RL algorithms can get stuck at local minima and yield sub-optimal controllers. To overcome this issue, recent studies [20], [21] propose guiding RL with additional information. For example, Thomas et al. [21] use CAD data to extract a geometric motion plan and a reward function that tracks the motion plan. Luo et al. [20] utilizes the force information obtained by the robot's interaction with the environment.

VI. Safety and Motion Planning

The cobot needs to perceive the occupied space (which is varying due to the movement of the worker and the parts and tools) and should plan and adjust its motions in real-time. This challenge was not addressed within our projects.

In order to ensure the safety of human worker, we have implemented a ROS watchdog node that is placed between the motion control node and the UR5, which checked the desired pose and velocity commands against predefined constraints, to ensure the operation of the cobot remain within a desired volume excluding the worker's body (but not the arms) and under predefined velocity thresholds.

VII. Conclusion

In this paper, we presented the "apprentice cobot" vision for human-robot collaboration and described the challenges involved in accomplishing it, shared our current results and future research plans.

Acknowledgment

This work is partially supported by TÜBİTAK through projects "ÇIRAK: Compliant Robot Manipulator Support for Assembly Workers in Factories" (117E002) and "KALFA: New Methods for Assembly Scenarios with Collaborative Robots" (120E269).

References

[1] Statista.com. Size of the global market for collaborative robots from 2017 to 2025. [Online]. Available: https://www.statista.com/statistics/748234/global-market-size-collaborative-robots/

[2] Y. Terzioğlu, B. Mutlu, and E. Şahin, "Designing social cues for collaborative robots: the role of gaze and breathing in human-robot collaboration," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 343–357.

[3] S. Saunderson and G. Nejat, "How robots influence humans: A survey of nonverbal communication in social human–robot interaction," *International Journal of Social Robotics*, vol. 11, no. 4, pp. 575–608, 2019.

[4] M. Salem, K. Rohlfing, S. Kopp, and F. Joublin, "A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction," in *2011 Ro-Man*. IEEE, 2011, pp. 247–252.

[5] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, "To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 313–323, 2013.

[6] L. D. Riek, T.-C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson, "Cooperative gestures: Effective signaling for humanoid robots," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 61–68.

[7] Q. Shen, K. Dautenhahn, J. Saunders, and H. Kose, "Can real-time, adaptive human–robot motor coordination improve humans' overall perception of a robot?" *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 1, pp. 52–64, 2015.

[8] T. Lorenz, A. Mörtl, and S. Hirche, "Movement synchronization fails during non-adaptive human-robot interaction," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2013, pp. 189–190.

[9] E. Ansermin, G. Mostafaoui, X. Sargentini, and P. Gaussier, "Unintentional entrainment effect in a context of human robot interaction: An experimental study," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 1108–1114.

[10] K. L. Koay, D. S. Syrdal, M. Ashgari-Oskoei, M. L. Walters, and K. Dautenhahn, "Social roles and baseline proxemic preferences for a domestic service robot," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 469–488, 2014.

[11] M. Walters, K. Dautenhahn, R. Boekhorst, K. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," vol. 2005, 09 2005, pp. 347 – 352.

[12] D. Shi, E. G. Collins Jr, B. Goldiez, A. Donate, X. Liu, and D. Dunlap, "Human-aware robot motion planning with velocity constraints," in *2008 International Symposium on Collaborative Technologies and Systems*. IEEE, 2008, pp. 490–497.

[13] K. L. Koay, D. S. Syrdal, M. Ashgari-Oskoei, M. L. Walters, and K. Dautenhahn, "Social roles and baseline proxemic preferences for a domestic service robot," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 469–488, 2014.

[14] T. Komatsu and S. Yamada, "Exploring auditory information to change users' perception of time passing as shorter," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: https://doi.org/10.1145/3313831.3376157

[15] S. Song and S. Yamada, "Expressing emotions through color, sound, and vibration with an appearance-constrained social robot," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 2–11. [Online]. Available: https://doi.org/10.1145/2909824.3020239

[16] F. C. Kurnaz, B. Hocaoglu, M. K. Yılmaz, İ. Sülo, and S. Kalkan, "Alet (automated labeling of equipment and tools): A dataset for tool detection and human worker safety detection," in *European Conference on Computer Vision*. Springer, 2020, pp. 371–386.

[17] Y. Li, K. Mo, L. Shao, M. Sung, and L. Guibas, "Learning 3d part assembly from a single image," 2020.

[18] J. Huang, G. Zhan, Q. Fan, K. Mo, L. Shao, B. Chen, L. Guibas, and H. Dong, "Generative 3d part assembly via dynamic graph learning," 2020.

[19] J. Li, C. Niu, and K. Xu, "Learning part generation and assembly for structure-aware shape synthesis," 2020.

[20] J. Luo, E. Solowjow, C. Wen, J. A. Ojea, A. M. Agogino, A. Tamar, and P. Abbeel, "Reinforcement learning on variable impedance controller for high-precision robotic assembly," 2019.

[21] G. Thomas, M. Chien, A. Tamar, J. A. Ojea, and P. Abbeel, "Learning robotic assembly from cad," 2018.