A framework for safe and intuitive human-robot interaction for assistant robotics

A framework for safe and intuitive human-robot interaction for assistant robotics

Pangcheng David Cen Cheng, Fiorella Sibona, Marina Indri
Dipartimento di Elettronica e Telecomunicazioni
Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
{pangcheng.cencheng, fiorella.sibona, marina.indri}@polito.it

***Abstract*—The brand new paradigm of Industry 5.0 envisages an increased leading role of the human operator in the production lines of the next future. Human-centric oriented solutions are going to be developed based on proactive human-robot collaborations, able to better exploit the skills and capabilities of both humans and cobots, mainly thanks to artificial intelligence. Several functionalities must be assured to reach such a goal, guaranteeing safety and flexibility, from human action prediction to object recognition and affordance. This paper offers an overview of the existing solutions for the various, separate issues, proposing a general framework for mobile manipulators assisting human workers, in a context of mass customization.**

***Index Terms*—Mobile manipulator, assistant robotics, human-robot interaction**

I. Introduction

Since its very first definition, the concept of Industry 4.0 has evolved, alongside the technologies it involved and the market pull behind it. The quest for new solutions and novel research directions suggests that a new industrial revolution is taking place to comply with these new requirements: Industry 5.0. Its most distinctive feature is the role of human workforce in the industrial automated smart processes. The human operator will increasingly have a leading role in mass customization, given the cognitive skills that intelligent machines still lack, e.g., creativity and critical thinking [1].

These new human-centric oriented solutions are envisioned to present autonomous cobots as human assistants, able to improve their supporting role thanks to Artificial Intelligence (AI). In this context, the cobot takes care of the operations requiring less cognitive skills along the supply chain management, and covers repetitive and routine monitoring tasks. By executing background but fundamental work, the cobots allow humans to focus on tasks requiring complex reasoning and decision making skills. In such a way, the human together with the AI enabled machine become a symbiotic system allowing for intelligence augmentation [2]. In order to support the human activity, the robot needs to be able to correctly perceive its surroundings, recognizing humans and understanding the environment. With this aim, Industry 4.0 solutions already featured complex and heterogeneous sensor systems [3]; on top of that, the current cobot evolution - Industry 5.0 cobot - has the responsibility to (i) lighten the human workload taking up repetitive and demanding work, and (ii) enhance its own perception and collaborative capabilities to enable a proactive Human–Robot Collaboration (HRC) paradigm.

Proactive HRC can be achieved by implementing bi-directional empathy and holistic understanding during the collaborative execution, spatio-temporal cooperation prediction, estimating the interaction among all elements involved (human, robot and workpiece - if any), and teamwork self-organization learning as a result of converging knowledge [4]. Sensor data are used for information extrapolation for a complete interpretation of the cobot surroundings, as a foundation for implementing prediction of human operator intentions and bi-directional multimodal communication to improve the interaction. To do so, the authors of [5] suggest that much of the relevant information can be retrieved from vision sources, relying on computer vision-based cognition of object, human, environment, and exploiting visual reasoning to bridge the gap between scene understanding and proactive decision-making. However, they point out how vision data might not be enough, given the complexity of understanding the object affordance properties, i.e., the interactive properties of objects for manipulation, and the need for an unambiguous interpretation of gestures for proper interaction. Thus, to achieve an effective interaction between humans and cobots in collaborative industrial applications, ambiguities must be avoided, possibly integrating further relevant pieces of information, such as natural language during bi-directional communication.

To implement a robotic assistant for manufacturing applications, deploying object/tool recognition and grasping capabilities is key. In [6], a mobile manipulator able to autonomously navigate while detecting humans and objects to automate small and medium-sized enterprises production is proposed. A pointvoxel-region based convolutional neural network is used to detect objects in a robust way, allowing to get rid of 3D point cloud data uncertainty issues, while a 2D camera is employed to calibrate the cobot relative position to workstations. In [7], an event-based robotic grasping framework for known and unknown objects in a cluttered scene is presented. In particular, the model-based solution consists of 3D scene reconstruction, object clustering according to Euclidean distance, position-based visual servoing, and grasp planning of objects whose shape is known a-priori. The mentioned works are part of a substantial body of literature that exploits AI to implement object recognition and manipulation.

During proactive collaboration, the assistant cobot and human operator may interact in several ways. One research direction for human-robot interfaces for collaborative assembly exploits Augmented Reality (AR) to provide the user with information coming from the robotic assistant. If correctly accepted by the user, AR can improve assembly cycle time performances [8]. In [9], a virtual representation of the cobot is rendered over the real robot allowing for a direct visual feedback of the set of waypoints to be executed, previously instructed through gaze and speech. The operator then can let the robot execute the set path, through a speech command. Natural language and gesture interpretation is being featured in a large part of proposed solutions. In [10], the operator is supported by a natural language-enabled virtual assistant, which translates the operator's requests, informs him/her about the robot status, and supports new operators during the training process.

The involved collaborators can have variable decision-making power, depending on the executed operation and context. For example, the work presented in [11] uses a mobile manipulator as a flexible solution for tending operations on computer-numerical-controlled and additive manufacturing machines. Given the cost of such equipment, they prefer a semi-automated execution modality, where low-level decisions and grasping planning are automatically computed by the robot, but high level decisions and pre-grasping poses are supervised by the human operator. This way, the operator can perform risk-informed decisions to accept, refine, or decline system-generated plans.

It is worth noting that most of the work provided by recent research directions for human-robot collaborative tasks, offer approaches involving AI-enabled mobile manipulators, since they embody the main capabilities demanded to a robotic assistant: flexibility, autonomous mobility and improved dexterity. Even though some solutions opt for supervised or semi-supervised collaboration, decision making data provided by the human operator during this kind of applications could be collected with the aim of providing a dataset to the AI-enabled cobot. In such a way, by training on human-generated past collected information, a cobot would be able to perform high level decisions on its own.

In fact, the cobot can learn from monitoring data, either to learn and emulate the human operators' motion, or to acquire it for later use, with the aim of action recognition and prediction. In [12], a dynamic movement primitives model is employed to simultaneously learn the motions of a human operator's body and hand, to be fed as reference trajectories for the mobile manipulator base and end-effector, respectively. Then an unscented model predictive control strategy is used for solving the trajectory tracking control problem for a mobile manipulator with uncertain parameters and disturbances.

To the best of the authors knowledge, current solutions solve separately human action prediction, object recognition, object affordance manipulation, safe motion control, however, there is still a gap for what concerns the safe human-robot interaction for assistive applications that involve object affordance. This paper has the aim to propose a framework for mobile manipulators assisting human workers. In the context of mass customization, the human participates actively in the production line since some unique human skills are required, such as creativity, problem solver mindset, promptness at analysing, and decision making. The role of the robotic system in this case is to give a sort of assistance to the human, which can enhance the performance of the overall production time.

The paper is organized as follows: Section II unfolds the proposed framework within a human-centered manufacturing scenario, while Section III describes the robot features divided in main functional blocks. Finally, Section IV draws some conclusions and sketches the future work.

II. Framework proposal

The proposed framework will be focused on mobile manipulators, since they provide more flexibility, with respect to fixed base manipulators, that could come in handy for material/tool handling/delivering from/to the human operator. In order to develop this framework, the robotic system should have at least the following features:

- An AI-based perception system for recognition and learning of human actions, tools/instruments, context, working spaces, whose data come from the sensors mounted on the robot and the ones used to monitor the workstation (Figure 1). The idea is that the robot should observe and learn from the human, while he/she is performing some tasks, e.g., assembly or inspection tasks. A sequence of tasks performed by the human can be predicted depending on the context, e.g., in an assembly task, the human worker may use a type of screwdriver and then use another screwdriver or wrench. Once the human actions are classified, the robot should be able to predict the next tool the human would like to use (so as to deliver the right tool as soon as the human completes the current task) and then take the tool that the human is no longer using.

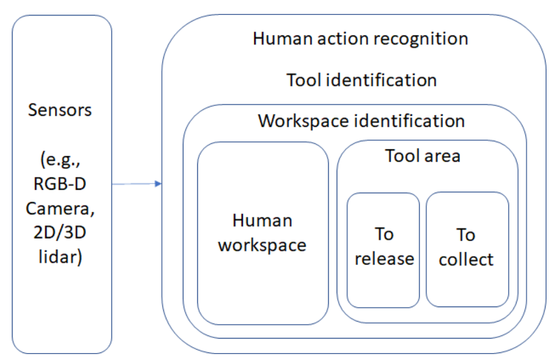

Fig. 1: AI-based system for human action recognition.

- A control system involving three blocks: (i) a safe trajectory generator, which depends on the tool that the robot is holding, (ii) a safe trajectory tracking system, and (iii) a safe trajectory replanner/corrector for those cases in which the robot needs to perform a trajectory correction or replanning its trajectory while avoiding in a safe manner the human operator, in order to avoid unintentional harm to him/her. The scheme is shown in Figure 2.

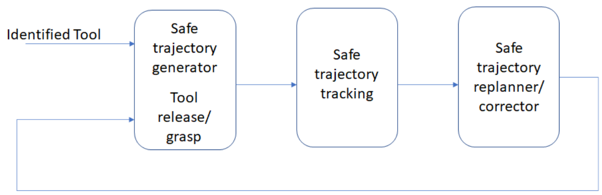

Fig. 2: Control block diagram for tool manipulation.

III. Framework feasibility analysis

In this section, the functional blocks that are considered fundamental for the development of the framework are described, while reviewing available current solutions. In this preliminary stage of the framework development, some assumptions are made for the sake of simplicity and to show correctly its workflow. The tasks performed by the human are limited to a sequence of short actions, and they are assumed to require a small number of tools.

A. Human action recognition

Usually, a human tends to repeat some actions while performing assembly tasks. This kind of actions can be used in the training process in order to recognize and predict the human actions. For instance, in an assembly task in which the human opens a device/product, inspects/modifies it, and then seals it up, it is likely that the human needs different tools in order to complete the task. The sequence in which the human will use the next tool can be predicted by the robot.

Hidden Markov Model (HMM) algorithms are popular for human motion prediction, since the hidden state transitions can be used when there are uncertainties due to weak motion recognition. In [13], HMM is used to predict a sequence of human actions by generating motion and observation probability matrices. Moreover, the principle of object affordance is included in the HMM model proposed by [14] in order to give some context to the human actions.

B. Tool identification

The tools can be classified a-priori by tool type, based on their utility and also on how they are usually grasped/held/used by the human operator. Identifying the tool may allow the robot to grasp it correctly and deliver it in a safe manner. Moreover, being aware of which tools are used allows the robot to put them away where all the other tools are kept. The tool identification process should also involve an affordance detection [15], that localizes, classifies and labels the affordance of the detected objects. Through this feature, it is possible to identify the graspable and non-graspable parts of a tool, e.g., the handle and the blade of a knife. There are several algorithms that are able to identify and detect the affordance of different objects. For instance, the authors in [16] presented *AffContext*, an object-agnostic affordance recognition neural network that analyses and predicts affordances of object parts in an image. In particular, *AffContext* is able to recognize the object elements even for novel ones that were not present in the training dataset.

C. Workspace organization

In a complex assembly task process, the human worker may need a set of different tools, however, it is better to have the working space as clear as possible, while keeping the tools that might be used repeatedly close and those that are dangerous/heavy/seldom used in a dedicated tool box/space, which can be accessed in case of necessity.

In order to define clear workspaces, which can be easily classified and comprehended by the robotic AI system, the working space (Figure 3) can be divided in the following areas:

- Human workspace. It is the area where the human can perform the tasks normally.
- Tool area. The tool area can be divided into two subregions:
  - An area in which the tools are released by the robot. The tools released in this part of the workspace are predicted by the AI system depending on the task performed by the human operator.
  - An area in which the robot collects the tools that the human is not using. The tools are retrieved by the robot and put back in place in the tool deposit.

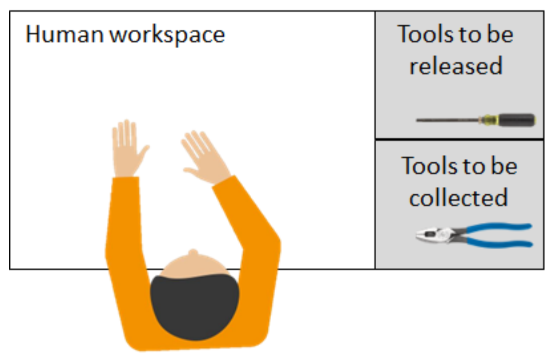

Fig. 3: Envisioned working space organization.

D. Mobile manipulator control system

Since a collaborative application is performed in a shared working space, safety is a relevant issue and the generated motions must take into account not only what is being executed but also how, to avoid hazardous situations. Inspiration can derive from tool-path smoothing methods, which aim at reducing rough tool motions that directly influence the dynamic performance of its motion control. For example, the tool path smoothing and interpolation algorithm proposed in [17] based on the finite impulse response filter, demonstrated to generate a path that has better tracking performance in the motion control process. This type of smoothing approaches could be applied for improving tracking of generated paths according to recognized human actions, to ensure a reliable behaviour.

Once the tool is identified, either after predicting the next tool the human operator may need, or when the unneeded tool must be put away, in order to ensure a safe interaction between the robot and the human, a safe trajectory generator should be used for each tool type, e.g., screwdrivers, hammers, pliers. This is because each of these tools can cause harm if they are not handled correctly, in particular while being held by a moving robot. The control system needs to consider the object affordance information as well as the motion of the human.

As the action is recognized, the cobot should be able to act according to this new information. To do so, control and manipulation approaches can be borrowed from non-collaborative frameworks. For example, in [18], the authors propose a methodology to perform sensor-based manipulation planning to automatically compute collision-free feasible trajectories. Then, a controller is associated to each trajectory segment generated by the manipulation planning algorithm. Given a list of tasks and relative sensor information, each controller computes a control variable corresponding to a sequence of tasks, which minimizes tasks' errors taking into account tasks' priorities. Moreover, [19] provides a high-accuracy method for estimating a workpiece pose for workpiece exchange, which consists in compensating the potentially poor positioning of the mobile base with a marker-based alignment of the manipulator. Exploiting end-effector camera images of fixed tags applied on workstations, the manipulator iteratively adjusts the camera to a previously recorded camera pose, so as to correctly compute the workpiece pose.

In this preliminary stage, the trajectory tracking will consider only the kinematic model of the mobile manipulator, while the safety issue will be dealt with using the onboard sensors, so as to avoid any contact or unintended harm to the human. In this case, the human can be initially modelled as a virtual obstacle with an enlarged inflation radius [20]. Future developments should consider the dynamic model of the robot in case of force/torque interactions; for instance, an impedance control architecture for collaborative material handling could be compliant with the safety measures in this framework [21].

IV. Conclusions and Next Steps

A framework that enables a safe and intuitive human-robot interaction for assistive tasks is sketched in this paper. In particular, several functional blocks were identified and investigated in order to evaluate their feasibility and the current state-of-the-art. Current solutions work separately on their own, however, to the best of the authors knowledge, there is not yet an existing solution that combines all the functional blocks in a human-centered manufacturing environment, particularly for a trajectory planner that considers both the object affordance and safety towards the human operator.

The next steps will involve the development of each functional block, taking into account the data complexity and compatibility coming from each block to be fed to the control system, in particular for the safe trajectory tracking and tool grasping.

References

[1] P. K. R. Maddikunta, Q.-V. Pham, B. Prabadevi, N. Deepa, K. Dev, T. R. Gadekallu, R. Ruby, and M. Liyanage, "Industry 5.0: A survey on enabling technologies and potential applications," *Journal of Industrial Information Integration*, vol. 26, p. 100257, 2022.

[2] L. Zhou, S. Paul, H. Demirkan, L. Yuan, J. Spohrer, M. Zhou, and J. Basu, "Intelligence augmentation: Towards building human-machine symbiotic relationship," *AIS Transactions on Human-Computer Interaction*, vol. 13, no. 2, pp. 243–264, 2021.

[3] A. Bonci, P. D. Cen Cheng, M. Indri, G. Nabissi, and F. Sibona, "Human-robot perception in industrial environments: A survey," *Sensors*, vol. 21, no. 5, p. 1571, 2021.

[4] S. Li, R. Wang, P. Zheng, and L. Wang, "Towards proactive human–robot collaboration: A foreseeable cognitive manufacturing paradigm," *Journal of Manufacturing Systems*, vol. 60, pp. 547–552, 2021.

[5] J. Fan, P. Zheng, and S. Li, "Vision-based holistic scene understanding towards proactive human–robot collaboration," *Robotics and Computer-Integrated Manufacturing*, vol. 75, p. 102304, 2022.

[6] Z. Zhou, L. Li, A. Fürsterling, H. J. Durocher, J. Mouridsen, and X. Zhang, "Learning-based object detection and localization for a mobile robot manipulator in sme production," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102229, 2022.

[7] X. Huang, M. Halwani, R. Muthusamy, A. Ayyad, D. Swart, L. Seneviratne, D. Gan, and Y. Zweiri, "Real-time grasping strategies using event camera," *Journal of Intelligent Manufacturing*, pp. 1–23, 2022.

[8] F. Schuster, U. Sponholz, B. Engelmann, and J. Schmitt, "A user study on AR-assisted industrial assembly," in *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2020, pp. 135–140.

[9] W. P. Chan, G. Hanks, M. Sakr, T. Zuo, H. M. Van der Loos, and E. Croft, "An augmented reality human-robot physical collaboration interface design for shared, large-scale, labour-intensive manufacturing tasks," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 308–11 313.

[10] C. Li, A. K. Hansen, D. Chrysostomou, S. Bøgh, and O. Madsen, "Bringing a natural language-enabled virtual assistant to industrial mobile robots for learning, training and assistance of manufacturing tasks," in *2022 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2022, pp. 238–243.

[11] S. Al-Hussaini, S. Thakar, H. Kim, P. Rajendran, B. C. Shah, J. A. Marvel, and S. K. Gupta, "Human-supervised semi-autonomous mobile manipulators for safely and efficiently executing machine tending tasks," *arXiv preprint arXiv:2010.04899*, 2020.

[12] D. Qin, A. Liu, J. Xu, W.-A. Zhang, and L. Yu, "Learning from human demonstrations for wheel mobile manipulator: An unscented model predictive control approach," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[13] H. Liu and L. Wang, "Human motion prediction for human-robot collaboration," *Journal of Manufacturing Systems*, vol. 44, pp. 287–294, 2017.

[14] M. Cramer, J. Cramer, K. Kellens, and E. Demeester, "Towards robust intention estimation based on object affordance enabling natural human-robot collaboration in assembly tasks," *Procedia CIRP*, vol. 78, pp. 255–260, 2018.

[15] M. Hassanin, S. Khan, and M. Tahtali, "Visual affordance and function understanding: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–35, 2021.

[16] F.-J. Chu, R. Xu, C. Tang, and P. A. Vela, "Recognizing object affordances to support scene reasoning for manipulation tasks," *arXiv preprint arXiv:1909.05770*, 2019.

[17] H. Sun, J. Yang, D. Li, and H. Ding, "An on-line tool path smoothing algorithm for 6R robot manipulator with geometric and dynamic constraints," *Science China Technological Sciences*, vol. 64, no. 9, pp. 1907–1919, 2021.

[18] J. Mirabel, F. Lamiraux, T. L. Ha, A. Nicolin, O. Stasse, and S. Boria, "Performing manufacturing tasks with a mobile manipulator: from motion planning to sensor based motion control," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2021, pp. 159–164.

[19] Y. Oba, K. Weaver, A. Parwal, H. Nagasue, and M. Fujishima, "High-accuracy pose estimation method for workpiece exchange automation by a mobile manipulator," *CIRP Annals*, vol. 70, no. 1, pp. 357–360, 2021.

[20] M. Indri, F. Sibona, and P. D. Cen Cheng, "Sensor data fusion for smart AMRs in human-shared industrial workspaces," in *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, vol. 1. IEEE, 2019, pp. 738–743.

[21] M. Logothetis, C. P. Bechlioulis, and K. J. Kyriakopoulos, "Decentralized impedance control of mobile robotic manipulators for collaborative object handling with a human operator," in *2021 29th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2021, pp. 741–746.