Computer Vision and Human–Robot Collaboration Supported Design-to-Robotic-Assembly

Bier, Henriette; Khademi, Seyran; van Engelenburg, Casper; Prendergast, J. Micah; Peternel, Luka

DOI
10.1007/s41693-022-00084-1

Publication date
2022

Document Version
Final published version

Published in
Construction Robotics

Construction Robotics
https://doi.org/10.1007/s41693-022-00084-1

ORIGINAL PAPER

Computer Vision and Human–Robot Collaboration Supported Design-to-Robotic-Assembly

Henriette Bier[1] · Seyran Khademi[1] · Casper van Engelenburg[1] · J. Micah Prendergast[1] · Luka Peternel[1]

Received: 19 July 2022 / Accepted: 24 November 2022

Abstract
While half of all construction tasks can be fully automated the other half relies to a certain degree on human support. This paper presents a *Computer Vision* (CV) and *Human–Robot Interaction/Collaboration* (HRI/C) supported *Design-to-Robotic-Assembly* (D2RA) approach that links computational design with robotic assembly. This multidisciplinary approach has been tested on a case study focusing on urban furniture and involving experts from respective disciplines and students.

Keywords Architecture · Computational design · Robotic assembly · Computer vision · Huma-robot interaction

1 Introduction

Industrial robots have been used in a wide range of production processes since the 70 s but more recently, researchers and engineers started to explore their potential in architecture and building construction (inter al. Bier 2018) and meanwhile more than 100 institutions and start-ups employ today industrial robots.[1]Considering that automation could raise productivity growth by 0.8–1.4 percent annually and almost half the activities in the global economy have the potential to be automated but less than 5 percent of all occupations can be automated entirely,[2]it is clear that people will continue working alongside machines.

While automation has been successfully implemented for over 5 decades in other industries, the building construction industry has suffered from slow technology adoption, skilled labour shortage, and productivity has lagged behind remaining one of the most dangerous activities involving more fatalities than any other sector in the EU.[3] These critical problems are addressed in the presented research by developing new *Design-to-Robotic-Assembly* (D2RA) methods that involve *Computer Vision* (CV) and *Human–Robot Interaction/ Collaboration* (HRI/C) approaches.[4] Such modes of design to construction involve agency of both humans and non-humans. Thus, agency is not located in one or another but in the heterogeneous associations between them (inter al. Bier 2014, 2018).

1.1 State-of-the-art

Architectural applications of HRI explored at ICD/ITKE and ETH are at the very beginning of development (inter al. Vasey et al. 2016; Mayer et al. 2017). They qualify as human-assisted robotic processes rather than HRI as they do not involve adaptive collaboration (inter al. Amor et al. 2014; Agravante et al. 2019). For example, in ETH's project Hanging Gardens[5] using multiple standard industrial robotic arms to assemble a structure composed of several wooden panels is without any direct involvement of human collaboration during the construction process. ICD/ITKE's project Hive[6] involves

[1] The Robotics in Architecture map (accessed from http://www.robotsinarchitecture.org/map-of-creative-robots) shows that more than 100 creative industry related institutions and start-ups are using robots worldwide.

[2] Link to McKinsey report: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/digital%20disruption/harnessing%20automation%20for%20a%20future%20that%20works/a-future-that-works-executive-summary-mgi-january-2017.ashx.

[3] Ibid link to McKinsey report and link Eurostat report: https://ec.europa.eu/eurostat/statistics-explained/index.php/Accidents_at_work_-_statistics_by_economic_activity.

[4] HRI/ C is a multidisciplinary field focusing on the development of safe human–robot interaction by involving AI, robotics, interaction design, ergonomics, and psychology.

[5] Link ETH project: https://ethz.ch/en/news-and-events/eth-news/news/2021/11/robots-build-new-hanging-gardens.html.

[6] Link to ICD project: https://www.icd.uni-stuttgart.de/projects/hive-a-human-and-robot-collaborative-building-process/.

✉ Henriette Bier
h.h.bier@tudelft.nl

[1] Technische Universiteit Delft, Delft, The Netherlands

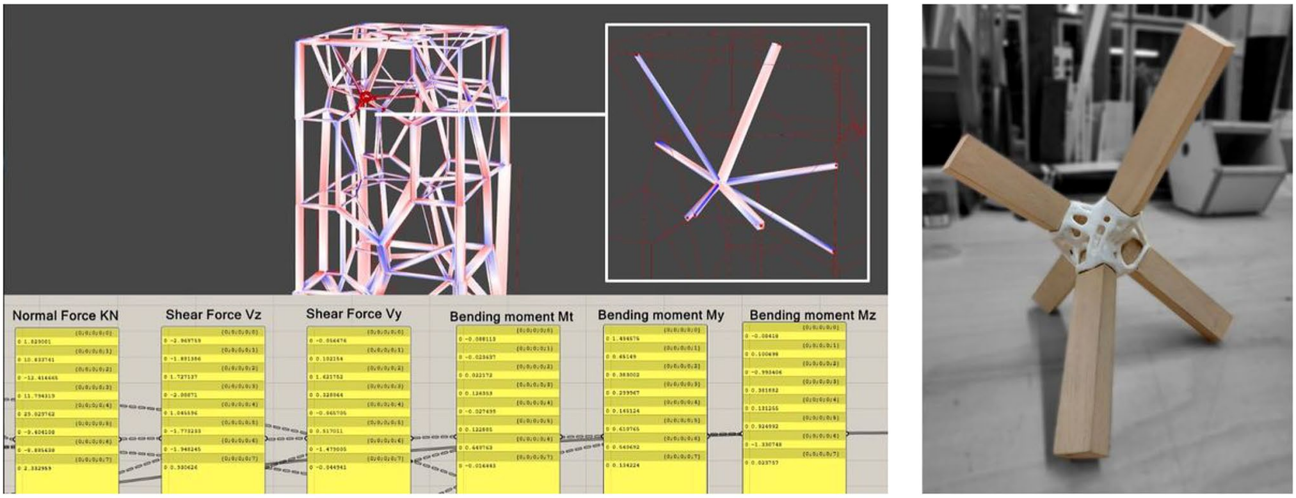

Fig. 1 Node (right) developed by structural optimisation at macro and meso level (left)

humans working alongside the robots during the building of a large structure, however, there is no direct physical interaction or coordination between them. These systems involve a combination of independent agents, who do not communicate or adapt to each other (Bier et al. 2018).

Robotic assembly of timber structures has been explored in recent works, as for instance, Stumm et al. (2018) and Devadass et al. (2019), who used HRI/C to haptically guide the robotic arm during the assembly. However, these methods involved no vision system so the robot was dependent on the human for most of the task. On the other hand, Rogeau et al. (2020) and Kunic et al. (2021) proposed a method for robotic assembly based on visual feedback, but it lacks the high-level HRI/ C part, which is critical when the task becomes cognitively too complex for the robot to execute by itself. Most recently, Kramberger et al. (2022) employed HRI/ C that was only used for humans to teach the robot how to then execute the assembly autonomously and no direct collaboration is implemented during the actual task execution. Also, the use of image processing to translate camera input to metric (distance) for robotic operation, does not focus on the size of the object as a discriminative feature to identify the correct piece to pick up. Moreover, it does not involve a reference (known size) object to estimate the geometry of unknown linear objects as presented in this chapter. It involves fiducial markers as employed by Rogeau et al. (2020) to keep track of the position and orientation of the various timber panels. Although efficient, the use of markers brings a clear disadvantage: unique markers have to be developed for every panel. Furthermore, in processes with differently sized components, geometric information about the overall structure must be either encoded in the marker or cannot be estimated by the robot at all, therefore, disabling the robot to do more complex tasks, which is the goal of the research presented in this chapter.

1.2 Contribution

In the CV and HRI/C supported D2RA approach presented in this paper (a) the robot is in direct physical collaboration with the human during the task execution, (b) it monitors and reacts to human intentions in real-time in order to execute tasks that actually require collaborative effort, and (c) it is able to learn from the collaborating human online. In the proposed approach, the robot uses CV to autonomously execute cognitively less demanding actions, such as moving in the vicinity of the detected object and moving in the vicinity of the assembly point, while the human physically guides the more complex actions, such as grasping the object and then assembling the component. In this context, the use of detection (position, size, and orientation) without the need for markers within a robotic construction process, as presented in this chapter, is unique. Furthermore, HRI during the task execution is often critical to overcome the limitation of robots' cognitive capabilities required for complex actions, such as grasping Fig. 1.

2 Design-to-Robotic-Assembly

D2RA is part of a larger *Design-to-Robotic-Production-Assembly and -Operation* (D2RPA&O) framework that improve process- and material-efficiency as well as embed intelligence in building processes and buildings by (1) computationally optimising material distribution and robotically producing building components (D2RP) and (2) by robotically assembling and operating those components

Fig. 2 Proposal for structure that can be partially dis- and reassembled with various functionalities (left) and structural performances (right)

(D2RA&O) using sensor-actuators (inter al. Bier 2018). These approaches have been tested in a case study wherein the focus has been on D2RA supported by CV and HRI/C methods (Peternel et al. 2018).

The case study involved the development of urban furniture. While furniture components vary in size and functionality, they all were designed with structural, functional, and assembly considerations in mind. The design relied on a Voronoi-based approach displaying degrees of porosity, where the degree and distribution of porosity i.e., density were informed by functional, structural and assembly requirements. The prototyped fragment (Fig. 1) consisted of an optimized node connecting linear elements of various length and thickness.

The project explored the notion of hybrid componentiality involving 3D printed nodes and cut and/or milled linear elements assembled into a larger structure. It involved computational design using Grasshopper scripts for variable Voronoi-cells distribution and structural optimisation. While the variation in cell size and distribution corresponds to functional requirements facilitating activities such as lounging, seating, and climbing, the variation in length and size of the linear elements resulted in the structural optimisation process (Fig. 2) presenting thicker members at the bottom of the structure.

3 Computer Vision

Computer Vision (CV) is used to recognise the location from which the building components, in this case linear elements of various sizes, are to be picked up by the robotic arm. The goal is to use only a (small set of) camera(s) at the construction site and no individual markers. In this project, a digital camera is placed above the scene where the design components are placed. A top view is chosen to make sure that the picking place is effectively represented in 2D plane, therefore, the physical distance (measured in terms of meters) is directly proportional to the distance in the image (measured in terms of pixels). From such an image, image processing and manipulation techniques are used to detect the building components, estimate their sizes, and compute their center points.

In this project, four main steps are distinguished: pre-processing, warping, detection and size estimation, and computation of the center point. Fundamental techniques that are used for pre-processing involve Gaussian blurring to de-speckle the image, Canny filter to extract the edges, and a sequence of morphological operations (dilation and erosion) to enhance the edges. For detecting the rectangular frame, a contour finding technique (Suzuki and Abe 1985) is applied, from which the bounding box (tightest fitting rectangular frame) can be easily extracted. The image is perspectively warped, and its modified version is perfectly encompassed by the frame. The pixel per metric is computed based on the (known) sizes of the frame Fig. 3.

An equivalent contour finding technique is used for detecting the linear elements (now as a set of elements instead of one element). The set of associated bounding boxes is compared to a query i.e., the particular linear element to be 'picked-and-placed'. The element with the bounding box that is the closest—based on the minimization of the squared distances—to the query is taken to be the one to be picked up. The coordinates of the center point, w.r.t. the known corner of the rectangular frame, is computed for this element only and sent to the robotic arm.

All consecutive steps are implemented in Python. The OpenCV library[7] was among several others used for image processing. The publicly available Github repository[8] shows in-depth how the pipeline can be implemented with two ready-to-be-used Google Colaboratory guiding notebooks and a self-constructed Python library.

[7] OpenCV. (2015). Open Source Computer Vision Library. https://opencv.org/.

[8] Khademi, S., van Engelenburg, C., Github repository for "1:1 Interactive Architecture Prototypes" https://github.com/caspervanengelenburg/1on1-prototyping-IA-CV-sessions.

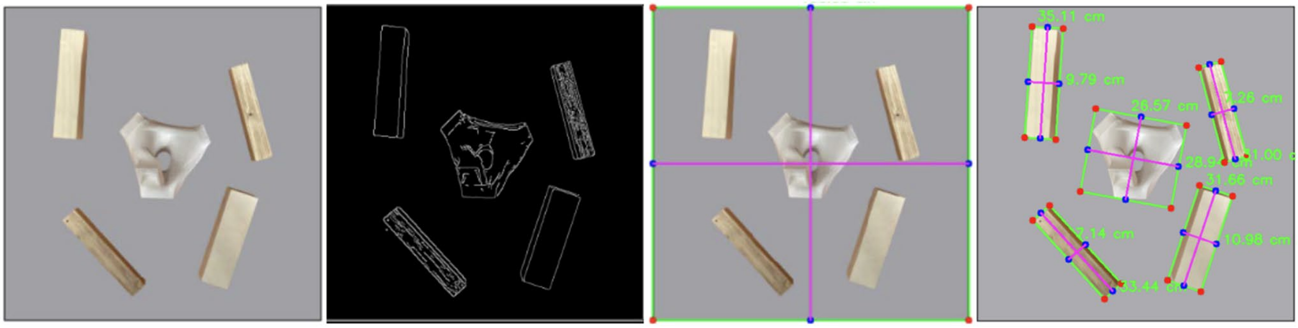

Fig. 3 CV to detect the location and center point of a query

Fig. 4 Automated and HRI/C supported assembly

4 Human–Robot Interaction/Collaboration

In the automated and HRI-supported assembly phase, the CV script links with the robotic process. This allows the robot to distinguish, locate, and measure the objects, and identify a work frame and its elements. While the picking of the linear elements is fully automated, assembling them into the node is HRI-supported (Fig. 4). The robot in this task supports the human user by bearing most of the load during assembly. The role of the human is to use their superior cognitive capabilities to handle complex aspects of the task, such as, supervising the grasping and manipulation of pieces, orienting the pieces that are being manipulated by the robot and thus ensure completion of the task. Before attempting assembly, a calibration step is performed to correctly identify the object and target extrinsic pose information relative to the robot. All programming was implemented in Python 3.7 and communication between the robot and the host computer was accomplished via the Robotic Operating System (ROS).

4.1 Extrinsic calibration

The robot is first put into a free-floating (gravity compensation) mode of operation, only applying enough torque to the robot's motors to compensate for its own weight. This allows the robot to be free and easy to move/manipulate by a human user. The robot is then moved by the user such that its end-effector[9] is in contact with the corner of the table that holds the components. This location is then saved via a keypress by the user such that the robot will always know the location of the table's coordinate system relative to its own base frame. While this is a simple step, it allows for the robotic and CV systems to operate completely independent of one another such that the robot does not require any knowledge of the camera or its location to complete the task. Once the table location has been stored, the robot is manually moved to its goal location (i.e., the final assembly position at which the user expects to assist the robot with the assembly). This goal location can be anywhere within the feasible workspace of the robot and in this case is defined by a mounting point where the assembly node was secured. This extrinsic pose information of the table and node, along with the object detection of the individual components lying upon the table

[9] SoftHand robotic gripper (qb-robotics: Cascina, Italy).

(as afforded by the CV system) is all that the robot needs to plan trajectories to each of the objects on the table and from each object location to the target node position (Fig. 4).

4.2 Robotic control and motion planning

A Cartesian impedance controller (Hogan 1985) is used to regulate all robot poses and to track the planned trajectories during the pick and place tasks. This controller is ideal for accommodating safe human–robot collaboration as it does not require precise position or force control (which can result in fast/powerful movements that might endanger a user within the robot's workspace) to accomplish its goals. Instead, this control strategy relies on desired reference position inputs for the end-effector with settable stiffness and damping components. Thus, the end-effector will behave as if spring and damper have been fixed between its actual position and input reference position. Using relatively low stiffness, the robot will attempt to maintain a pose or track a trajectory,[10] while also accommodating physical collisions or outside forces from its environment in a gentle (low contact forces) and human-friendly manner (inter al. Peternel and Ajoudani 2017; Peternel et al. 2018, 2019; Lamon et al. 2019). The stiffness of this virtual spring/damper can be set with a trade-off between precise but less safe (high stiffness), and less precise positioning but safer (lower stiffness). To enable the pick and place task, trajectories must also be planned for the robot to move from its current position to each component and from each component's position to the node/goal position. To plan these point-to-point trajectories, a 5th order polynomial approach was implemented (Angeles and Alivizatos 1987) that both determines a reference pose for the end-effector at every timestep along the path and ensures that the acceleration and velocity of the end-effector will be zero at the beginning and end of each trajectory. This is crucial for smooth and safe motion of the end-effector and, when combined with the soft (low stiffness) impedance controller, it allows for safe human collaboration with the robot during the entire assembly task.

4.3 Pick and place task

With this extrinsic calibration step, complete and the robot control system established, the robot and CV system can now operate together to complete the construction task. Once the components have been placed on the table, an image is captured showing the objects and tabletop. This image is then used to provide a clickable interface for the human user. The steps to completing the task are as follows. (i) the CV system detects and identifies the components placed on the table along with the components' Cartesian coordinates relative to the corner of the tabletop (which corresponds to the table's pose as set during the calibration step). (ii) The image and components are displayed on a screen and the human user can click any of the components to select them for assembly. The spatial location of that component is then sent to the robot which will automatically plan a trajectory to that position and guide the hand near the selected component. (iii) Once the robot has reached the component the human can adjust the hand position until it is in the ideal grasping position, (iv) the human then commands the robot to close the hand and the robot plans a trajectory from that point to the calibrated node position. (v) Upon reaching the node position, the robot reduces the stiffness of its controller allowing the human user to manipulate the component into the correct place within the node.

[10] In essence, a trajectory is composed of a series of desired reference positions for the impedance controller.

5 Results and discussion

This methodology was tested with three different student groups who each tested the complete workflow from design to final HRI assembly. Each of the three groups was successful in assembling their final design and the approach has proven to be effective with respect to establishing a robotic construction process, where humans and robots work side by side, however, various limitations were identified. An example trajectory can be seen in Fig. 5 where the robot is shown in both the pickup and assembly positions for a single component. The reference trajectory is tracked quite closely during the robot's motion, however, larger deviations can be seen when the student moves and modifies the end-effector pose during pickup/assembly phases of the task and/or updates the reference position of the robot's end-effector to better meet their needs.

The presented HRI/C implementation aims to simplify several key aspects of the perceptual and automation tasks, and this presents both limitations and opportunities for future work. Rather than attempting to fully automate the assembly process, via real-time visual feedback and precise position control, parts of these roles were left with the human user. A precisely controlled robot for example could achieve the objective of moving a component and fitting it into the desired node itself, but this would require: (i) high gain control (dangerous for humans within the vicinity of the robot), (ii) exceptional sensory feedback on component position, node position, etc., and (iii) careful calibration that goes far beyond what is necessary in the required extrinsic calibration step, while also needing to account for environmental and sensor noise. Leaving some of these responsibilities with the human user allows for a robust system capable of safely handling a diverse range of uncertainty while

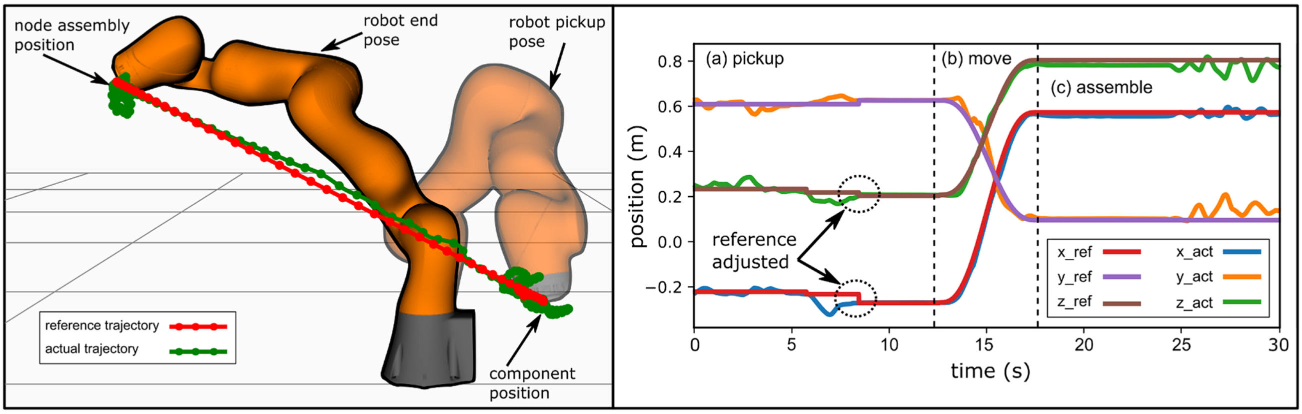

Fig. 5 Trajectory tracking and adjustment. Left: the robot is shown in the pickup and assembly pose with reference and actual trajectory shown. In both positions, the robot is moved and adjusted by the human to better accommodate the picking up and assembly of the components. Right: The end-effector position is shown for the full 30-s trial. **a** During the pickup phase, the human can move the robot as needed and adjust the reference position. **b** The robot moves along the planned trajectory to the assembly pose. **c** In the assembly phase, the human is again able to adjust the robot's pose as needed

offloading the physically difficult aspects of the task (lifting and moving of components) from the human.

While this is a functional division of responsibilities, affording additional capabilities to the robot would allow for improved performance across a range of different metrics. Real-time visual feedback on component and node position for example would enable capabilities such as faulty part detection or error correction in the event that one component or node is moved or misplaced. Likewise, component models (complete with geometry and moments of inertia) could be taken into account within the trajectory planning step to accommodate the weight of each component (for improved gravity compensation) as well as the physical dimensions of the components to avoid collisions with objects, humans or the robot's joints.

In the presented system, Artificial Intelligence (AI) has been used for detecting the scene with known components and for planning the robot actions as demonstrated by the human. Next steps are envisioned with respect to training the system to autonomously detect other/unknown objects in the scene and also identify faulty parts. Furthermore, the robot could employ reinforcement learning (Kober et al. 2013) to optimise trajectories and sequence of assembly beyond what was demonstrated by the human.

Developing industrial applications in the future using this approach involves consideration with respect to scaling up. Also, identifying the challenges and opportunities of prefab vs. on-site construction need to be investigated. This requires the involvement of industrial partners to test the approach in relevant environments and develop first small-scale industrial production. Another challenge in scaling to industrial applications is the payload limitation of existing torque-controlled collaborative robots. Currently available systems are quite limited in their payload capacity and thus may be unable to handle heavier components. While larger position-controlled industrial robots offer significantly greater power and payload capacities, these systems are not collaborative and thus are not safe for humans to engage with physically. Finally, limitations may also exist with regard to component grasping. Depending on the type, size and weight of components, traditional robotic grippers may be unable to safely grasp and hold certain components, thus custom grippers or a variety of grippers might be required to complete tasks on a larger scale.

6 Conclusion

The presented CV and HRI/C supported D2RA methods are linking design to assembly processes in which humans and robots work safely side by side. They rely on CV and HRI/C to pick and move the linear elements and then fit them into the node, respectively. They are proof of concept for construction tasks that are thus partially automated, while also relying to a certain degree on human support. They can be employed in all assembly tasks that involve components of various sizes and require careful calibration.

Future work will focus on the integration of CV and HRI/C with D2RPA&O processes and scaling up to building scale with the goal to automate and HRI/C support all tasks in buildings and building processes. It will include improving trajectory planning to better accommodate a diversity of component geometries and weights and to provide for additional shared responsibility of the robot in more cognitive aspects of the tasks (i.e., assembly failures, faulty part detection, automated selection of components during the build sequence etc.).

CV-specific future work will focus on more robust object detection, which is needed when building sites are 'noisy' due to dust or debris. The CV pipeline as it stands now, can be tweaked separately for various pickup places. The challenge is to deal with varying weather circumstances and pickup places. An option to circumvent the need for a very complex CV pipeline is to train instead data-driven object detection models, which will be addressed in future work.

Acknowledgements This paper has profited from the contribution of researchers and students from the Robotic Building, AiDAPT and Cognitive Robotics labs involved in the project.

Funding There are no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Data Availability CV-related data is available on GitHub (https://github.com/caspervanengelenburg/1on1-prototyping-IA-CVsessions).

References

Agravante D, Cherubini A, Sherikov A, Wiebe P, Kheddar A (2019) Human-humanoid collaborative carrying, IEEE Transactions on Robotics

Amor H, Neumann G, Kamthe S, Kroemer O, Peters J (2014) Interaction primitives for human-robot cooperation tasks, IEEE international conference on robotics and automation

Angeles J, Alivizatos A (1987) The synthesis of smooth trajectories for pick-and-place operations. Technical Report McRCIM, McGill University

Bier HH, Knight T (2014) Data-driven design to production and operation. Footprint 8(2):1–8. https://doi.org/10.7480/footprint.4.1

Bier HH (2018) Robotic Building. Springer Adaptive Environments book series, Springer Cham. https://doi.org/10.1007/978-3-319-70866-9

Bier HH, Cheng AL, Mostafavi S, Anton A, Bodea S (2018) Robotic building as integration of design-to-robotic-production and-operation. Robotic Building (pp. 97–119). Springer. https://doi.org/10.1007/978-3-319-70866-9_5

Devadass P, Stumm S, Brell-Cokcan S (2019) Adaptive haptically informed assembly with mobile robots in unstructured environments. Proceedings of the International Symposium on Automation and Robotics in Construction (IAARC). https://doi.org/10.22260/isarc2019/0063

Hogan N (1985) Impedance control: An approach to manipulation. J Dynam Syst Measur Control. https://doi.org/10.1115/1.3140713

Kober J, Bagnell JA, Peters J (2013) Reinforcement learning in robotics: A survey. Int J Robot Res 32(11):1238–1274

Kramberger A, Kunic A, Iturrate I, Sloth C, Naboni R, Schlette C (2022) Robotic assembly of timber structures in a human-robot collaboration setup. Front Robot AI. https://doi.org/10.3389/frobt.2021.768038

Kunik A, Nabonia R, Krambergerb A, letteb C, (2021) Design and assembly automation of the robotic reversible timber beam. Autom Constr 123(1):103531. https://doi.org/10.1016/j.autcon.2020.103531

Lamon E, De Franco A, Peternel L, Ajoudani A (2019) A capability-aware role allocation approach to industrial assembly tasks, IEEE Robotics and Automation Letters

Mayer H, Gramazio F, Kohler M (2017) Disarmed strategies: new machines and techniques for an era of computational contextualism in architecture. Archit Des. https://doi.org/10.1002/ad.2203

Peternel L, Ajoudani A (2017) Robots learning from robots: A proof of concept study for co-manipulation tasks, IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)

Peternel L, Tsagarakis N, Caldwell D, Ajoudani A (2018) Robot adaptation to human physical fatigue in human–robot co-manipulation. Auton Robot 42(5):1011–1021

Peternel L, Fang C, Tsagarakis N, Ajoudani A (2019) A selective muscle fatigue management approach to ergonomic human-robot co-manipulation, Robotics and Computer-Integrated Manufacturing

Rogeau N, Tiberghien V, Latteur P, Weinand Y (2020) Robotic Insertion of Timber Joints using Visual Detection of Fiducial Markers. Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC). https://doi.org/10.22260/isarc2020/0068

Stumm S, Devadass P, Brell-Cokcan S (2018) Haptic programming in construction. Constr Robot 2(1–4):3–13. https://doi.org/10.1007/s41693-018-0015-9

Suzuki S, Abe K (1985) Topological structural analysis of digitized binary images by border following. CVGIP 30(1):32–46. https://doi.org/10.1016/0734-189X(85)90016-7

Vasey L, Grossman T, Kerrick H, Nagy D (2016) The hive: a human and robot collaborative building process, SIGGRAPH '16. https://doi.org/10.1145/2897839.2927404