Human-robot Collaboration for Bin-picking Tasks to Support Low-volume Assemblies

Krishnanand N. Kaipa, Carlos W. Morato, Jiashun Liu, and Satyandra K. Gupta
Maryland Robotics Center
University of Maryland, College Park, MD 20742
Email: skgupta@umd.edu

Abstract—In this paper, we present a framework to create hybrid cells that enable safe and efficient human-robot collaboration (HRC) during industrial assembly tasks. We present our approach in the context of bin-picking, which is the first task performed before products are assembled in certain low-volume production scenarios. We consider a representative one-robot one-human model in which a human and a robot asynchronously work toward assembling a product. The model exploits complementary strengths of either agents: Whereas the robot performs bin-picking and subsequently assembles each picked-up part to build the product, the human assists the robot in critical situations by (1) resolving any perception and/or grasping problems encountered during bin-picking and (2) performing dexterous fine manipulation tasks required during assembly. We explicate the design details of our overall framework comprising three modules: plan generation, system state monitoring, and contingency handling. We use illustrative examples to show different regimes where human-robot collaboration can occur while carrying out the bin-picking task.

I. INTRODUCTION

Assembly tasks are integral to the overall industrial manufacturing process. After parts are manufactured, they must be assembled together to impart the desired functionality to products. Pick-and-place, fastening, riveting, welding, soldering, brazing, adhesive bonding, and snap fitting constitute representative examples of industrial assembly tasks [1].

Humans and robots share complementary strengths in performing assembly tasks. Humans offer the capabilities of versatility, dexterity, performing in-process inspection, handling contingencies, and recovering from errors. However, they have limitations w.r.t. factors of consistency, labor cost, payload size/weight, and operational speed. In contrast, robots can perform tasks at high speeds, while maintaining precision and repeatability, operate for long periods of times, and can handle high payloads. However, currently robots have the limitations of high capital cost, long programming times, and limited dexterity. Owing to these reasons, small batch and custom production operations predominantly use manual assembly. However, in mass production lines, robots are often utilized to overcome the limitations of human workers.

Small and medium manufacturers (SMMs) represent a pivotal segment of the manufacturing sector in US (National Association of Manufacturers estimates that there are roughly 300,000 SMMs in US). High labor costs make it difficult for SMMs to remain cost competitive in high wage markets. Purely robotic cells is not a solution as they do not provide the needed flexibility. These reasons, along with short production cycles and customized product demands, set SMMs as primary candidates to benefit from hybrid cells that support human-robot collaborations. However, currently shop floors install robots in cages. During robot operation, the cage door is locked and elaborate safety protocol is followed in order to ensure that no human is present in the cage. This makes it very difficult to design assembly cells where humans and robots can collaborate effectively.

In this paper, we present a framework for hybrid cells that enable safe and efficient human-robot collaboration (HRC) during industrial assembly tasks. Advent of safer industrial robots [2, 3, 4] and exteroceptive safety systems [5] in the recent years are creating a potential for hybrid cells where humans and robots can work side-by-side, without being separated from each other by physical cages. The main idea behind hybrid cells is to decompose assembly operations into tasks such that humans and robots can collaborate by performing tasks that are suitable for them. In fact, task decomposition between the human and robot (Who does what?) has been identified as one of the four major problems in the field of human robot collaboration [6].

We consider a representative one-robot one-human model in which a human and a robot asynchronously work toward assembling a product. The model exploits complimentary strengths of either agents: Whereas the robot performs a bin-picking task and subsequently assembles each picked-up part to form the product, the human assists the robot in critical...
situations by (1) resolving any perception and/or grasping problems encountered during bin-picking and (2) performing dexterous fine manipulation tasks required during assembly. We explicate the design details of our overall framework comprising three modules: plan generation, system state monitoring, and contingency handling.

In this paper, we focus on the bin-picking task to illustrate our human-robot collaboration (HRC) model. Bin-picking is one of the crucial tasks performed before products are assembled in certain low-volume production scenarios. The task involves identifying, locating, and picking a desired part from a container of randomly scattered parts. Many research groups have addressed the problem of enabling robots, guided by machine-vision and other sensor modalities, to carry out bin-picking tasks. The problem is very challenging and still not fully solved due to severe conditions commonly found in factory environments. In particular, unstructured bins present diverse scenarios affording varying degrees of part recognition accuracies: 1) Parts may assume widely different postures, 2) parts may overlap with other parts, and 3) parts may be either partially or completely occluded. The problem is compounded due to factors like background clutter, shadows, complex reflectance properties of parts made of various materials, and poorly lit conditions.

Owing to the complicated nature of these problems, most manufacturing solutions resort to kitting where parts are manually sorted into a container by human workers. This thereby simplifies the perception task and allows the robot to pick from the kitted set of parts before proceeding to final assembly. However, kitting takes significant time and manual labor.

Whereas robots can repetitiously perform routine pick-and-place operations without any fatigue, humans excel at their perception and prediction capabilities in unstructured environments. Their sensory and mental-rehearsal capabilities enable humans to respond to unexpected situations, quite often with very little information. We exploit these complementary strengths of either agents in order to design a deficit-compensation model that overcomes the primary perception and decision-making problems associated with the bin-picking task. An overview of the hybrid cell is shown in Fig. 1.

II. RELATED WORK

Different modes of collaboration between humans and robots have been identified in the recent past. A concept feasibility study conducted by the United States Consortium for Automotive Research (USCAR) in 2010-2011—defined three levels of human and robot collaboration. Shi et al. used this standard to categorize current robotic systems in automotive body shop, powertrain manufacturing and assembly, and general assembly. In low-level HRC, the human does not interact directly with the robot. Instead, the human loads the part into a transfer device (e.g., conveyor belt, rotary table, etc) by standing outside of the working range of the robot. Medium-level HRC differs from the low-level in that the operator directly loads the part into the end-of-arm-tooling (EOAT). However, the robot is de-energized until the human exits the working range and initiates the next action. In high-level HRC, both robot and human simultaneously act within the working range of the robot. The human may or may not come in physical contact with the robot.

Shi and Menassa proposed transitional and partnership based tasks that are examples of medium level and high-level HRC, respectively. The transitional manufacturing task consists of a welding subtask by the robot and a random quality inspection subtask by the human. This task is characterized by serial transitions of the task between the human and the robot (human places part on EOAT before welding and picks up part after welding) during a short duration. In the partnership based assembly task, an intelligent assist device (IAD) plays the role of the robot partner. The task consists of a moonroof assembly: whereas the IAD transports and places a heavy and bulky moon-roof, the human performs dexterous fine manipulative assembly like fastening and wiring harness.

The transitional and complementary nature of the first manufacturing task is similar to that of the HRC task chosen in our paper. In their approach, the human loads the part into the EOAT. The human feeds the part to the robot in our approach too. However, we exploit the human’s perception ability—in recognizing parts in a wide variety of complex part-bin scenarios—to fill the deficit-compensatory role in the collaborative assembly task chosen in our work.

Sisbot and Alami developed a human-aware manipulation planner that generated safe and “socially-acceptable” robot movements based on the human’s kinematics, vision field, posture, and preferences. They demonstrated their approach in a “robot hand over object” task. Jido, a physical 6 DOF robot, used the planner to determine where the interaction must take place, thereby producing a legible and intentional motion while handing over the object to the human. They conducted physiological user studies to show that such human-aware collaboration by the robot decreases cognitive burden on the human partner. The aspect of achieving safe robot motions based on sensed human-awareness is similar to the one achieved in our work. Their robot is used in an assistive role. In contrast, the robot and the human partners are used in complimentary roles in our approach.

Nikolaidis et al. present methods for predictable, joint actions by the robot for HRC in manufacturing. The authors used computationally-derived teaming models to quantify a robot’s uncertainty in the successive actions of its human partner. This enables the robot to perform risk-aware collaboration with the human. In contrast, our approach uses real-time sensing and replanning to address contingencies that arise during collaboration between the partners.

Duan et al. developed a cellular manufacturing system that allows safe HRC for cable-harness tasks. Two mobile manipulators play an assistive role through two functions: (1) deliver parts from a tray shelf to the assembly table and (2) grasp the assembly parts and prevent any wobbling during the assembly process. Safety is ensured using light
curtains, safety fence, and camera-based monitoring system. Assembling directions are provided via visual display, audio, and laser pointers. Some other approaches to hybrid assembly cells can be found in \[18, 19, 20, 21, 22\]. While having most of the primary features of the previous hybrid cell approaches, our framework also provides methods that explicitly handle contingencies in human robot collaboration.

### III. Hybrid Cell Framework

Our approach to hybrid cells considers a representative one-human one-robot model, in which a human and a robot will collaborate to assemble a product. In particular, the cell will operate in the following manner:

1. The cell planner will generate a plan that will provide instructions for the human and the robot in the cell.
2. Instructions for the human operator will be displayed on a screen in the assembly cell.
3. The robot will pick up parts from an unstructured bin and assemble them into the product.
4. The human will be responsible to assist the robot by resolving any perception and/or grasping problems encountered during bin-picking.
5. If needed, the human will perform the dexterous fine manipulation to secure the part in place in the product.
6. The human and robot operations will be asynchronous.
7. The cell will be able to track the human, the locations of parts, and the robot at all time.
8. If the human makes a mistake in following an assembly instruction, re-planning will be performed to recover from that mistake and appropriate warnings and error messages will be displayed in the cell.
9. If the human comes too close to the robot to cause a collision, the robot will perform a collision avoidance strategy.

The overall framework of the hybrid cell consists of the following three modules: (1) Plan generation, (2) system state monitoring, and (3) contingency handling.

#### A. Plan generation.

We should be able to automatically generate plans in order to ensure efficient cell operation. This requires generating feasible assembly sequences and instructions for robots and human operators, respectively. Automated planning poses the following two challenges. First, generating precedence constraints for complex assemblies is challenging. The complexity can come due to the combinatorial explosion caused by the size of the assembly or the complex paths needed to perform the assembly. Second, generating feasible plans requires accounting for robot and human motion constraints.

**Assembly sequence generation:** Careful planning is required to assemble complex products (\[23\]). We utilize a method developed in \[25, 24\] that automatically detects part interaction clusters that reveal the hierarchical structure in a product. This thereby allows the assembly sequencing problem to be applied to part sets at multiple levels of hierarchy. A 3D CAD model of the product is used as an input to the algorithm.

The approach described in \[25\] combines motion planning and part interaction clusters to generate assembly precedence constraints. The result of applying the method on a simple assembly model (Fig. 2(a)) is a feasible assembly sequence (Fig. 2(b)).

**Instruction generation for humans:** The human worker inside the hybrid cell follows an instructions list to perform assembly operations. However, poor instructions lead to the human committing assembly related mistakes. We address this issue by utilizing an instruction generation system developed in \[26\] that creates effective and easy-to-follow assembly instructions for humans. A linearly ordered assembly sequence is given as input to the system. The output is a set of multimodal instructions (text, graphical annotations, and 3D animations). Instructions are displayed on a big monitor located at a suitable distance from the human. Text instructions are composed using simple verbs such as Pick, Place, Position, Attach, etc. Data from each step of the input assembly plan is used to simulate the corresponding assembly operation in a virtual environment (developed based on Tundra software). Animations are generated based on visualization of the computed motion of the parts in the virtual environment.

Humans can easily distinguish among most parts. However, they may get confused about which to pick when two parts look similar. To address this problem, we utilize a part identification tool developed in \[26\] that automatically detects such similarities and present the parts in a manner that enables the human to select the correct part.

#### B. System state monitoring.

To ensure smooth and error-free operation of the cell, we need to monitor the state of the assembly operations.
Accordingly, we present methods for real-time tracking of the human operator, the parts, and the robot.

**Human tracking:** The human tracking system is based on an N-Kinect based sensing framework developed in [5, 27]. The system is capable of building an explicit model of the human in near real-time. It is designed for a hybrid assembly cell where one human interacts with one robot to perform assembly operations. The human has complete freedom of his/her motion. Human activity is captured by the Kinect sensors that reproduce the human’s location and movements virtually in the form of a simplified animated skeleton. The sensing system consists of four Kinects mounted on the periphery of the work cell. Occlusion problems are resolved by using multiple Kinects. The output of each Kinect is a 20-joint human model. Data from all the Kinects are combined in a filtering scheme to obtain the human motion estimates.

**Part Tracking:** The assembly cell state monitoring uses a discrete state-to-state part monitoring system that was designed to be robust and decrease any possible robot motion errors. A failure in correctly recognizing the part and estimating its pose can lead to significant errors in the system. To ensure that such errors do not occur, the monitoring system is designed based on 3D mesh matching with two control points—the first control point detects the part picked-up either by the robot or the human and the second control point detects the part’s spatial transformation when it is placed in the robot’s workspace. The detection of the selected part in the first control point helps the system to track the changes introduced by the robot/human in real-time and trigger the assembly re-planning and the robot motion re-planning based on the new sequence. Moreover, the detection of the posture of the assembly part related to the robot in the second control point sends a feedback to the robot with the "pick and place" or "wait" flag.

The 3D mesh matching algorithm uses a real-time 3D part registration and a 3D CAD-mesh interactive refinement [31]. Multiple acquisitions of the surface are necessary in order to register the assembly part in 3D format. These views are obtained by the Kinect sensors and represented as dense point clouds. In order to perform a 3D CAD-mesh matching, an interactive refinement revises the transformations composed of scale, rotation, and translation. Such transformations are needed to minimize the distance between the reconstructed mesh and the 3D CAD model. Point correspondences were extracted from the reconstructed scene and the CAD model using the iterative closest point algorithm [32].

The part detection (first control point) and part posture determination (second control point) results are shown in Figs. 3 and 4 respectively. The initial scene in the first experiment is shown in Fig. 3(a). The 3D mesh representation of the initial scene is shown in Fig. 3(c). The human is instructed to pick the 'Cenroll bar' part. However, he picks the 'Rear brace' (Fig. 3(d)). Now, we show how this is detected by the system. Figure 3(f) shows a 3D mesh representation of the scene after the part is picked. Next, point cloud of the current scene is matched with that of CAD model of every part. Since the 'Rear brace' is not present in the live scene mesh, the error associated with its mesh matching with the live scene is greater than that of every other part. This determines that the part picked is 'Rear brace', thereby indicating to the human that a wrong part has been picked (Figs. 3(g) and 3(h)). Human replaces the part and picks a different part in Fig. 3(i). Application of the part matching routine to the new scene is shown in Figs. 3(j)-3(m).

In the second experiment, the human places the part in robot’s workspace (Fig. 4(a-c)). The desired part posture is shown in a virtual environment in blue color (only for illustration purpose) in Fig. 4(d). The 3D reconstruction of the real scene is shown in Fig. 4(e). A 3D reconstruction of the reference scene with the part in the desired posture is shown in Fig. 4(f). A large MSE (shown in Fig. 4(g)) and a low
D. Contingency handling

We consider three types of contingency handling—collision avoidance between robot and human, replanning, and warning generation.

Collision avoidance between robot and human: Our approach to ensuring safety in the hybrid cell is based on the pre-collision strategy developed in: robot pauses whenever an imminent collision between the human and the robot is detected. This stop-go safety approach conforms to the recommendations of the ISO standard 10218 [29, 30]. In order to monitor the human-robot separation, the human model generated by the tracking system (described in the previous section) was augmented by fitting all pairs of neighboring joints with spheres that move as a function of the human’s movements in real-time. A roll-out strategy was used, in which the robot’s trajectory into the near future was pre-computed to create a temporal set of robot’s postures for the next few seconds. Next, the system verified if any of the postures in this set collides with one of the spheres of the augmented human model. The method was implemented in a virtual simulation engine developed based on Tundra software.

Replanning and warning generation: An initial assembly plan is generated before the operations begin in the hybrid assembly cell. We define deviations in the assembly cell as a modification to the predefined plan. These modifications can be classified into three main categories:

i Deviations that leads to process errors. These are modifications introduced by either the human or robot that cannot generate a feasible assembly plan. These errors can generate an error in the assembly cell in a way that will require costly recovery. In order to prevent this type of errors, the system has to detect the presence of this modification by the registration of the assembly parts. Once the system has information about the selected assembly part, it evaluates the error in real-time by propagating the modification in the assembly plan and giving a multi-modal feedback.

ii Deviations that leads to improvements in the assembly speed or output quality. Every single modification to the master assembly plan is detected and evaluated in real-time. The initial assembly plan is one of the many feasible plans that can be found. A modification in the assembly plan that generates another valid feasible plan classifies as an improvement. These modifications are accepted and give the ability and authority to the human operators to use their experience in order to produce better plans. This process helps the system to evolve and adapt quickly using the contributions made by the human agent.

iii Deviation that leads to adjustment in the assembly sequence. Adjustments in the assembly process may occur when the assembly cell can easily recover from the error introduced by the human/robot by requesting additional interaction in order to fix it. Another common mistake in assembly part placement is the wrong pose (rotational and translational transformation that diverges

C. Robot Tracking

We assume that the robot can execute motion commands given to it so that the assembly cell will know the robot state.
Fig. 5. Representative parts affording different recognition and grasping complexities: (a) Part 1 (b) Part 2 (c) Part 3 (d) Part 4.

Fig. 6. Part bin scenarios of varying complexities: (a) Ordered pile of parts. (b) Random pile with low clutter. (c) Random pile with medium clutter. (d) Random pile with high clutter.

from the required pose). Two strategies can be found to solve this issue: a) robot recognizes the new pose and recomputes its motion plan in order to complete the assembly of the part or b) human is informed by the system about the mistake and is prompted to correct it. More details about the replanning system can be found in [28].

IV. HRC MODEL FOR BIN-PICKING

Our HRC model to perform the bin-pick task consists of the following steps. Under normal conditions, the robot uses a part recognition system to automatically detect the part, locate part posture, and plan its motion in order to grasp and transfer the part from the bin to the assembly area. However, if the robot determines that the part recognition is not clear from the current scene, then it initiates a collaboration with the human. The particular bin scenario determines the specific nature of collaboration between the robot and the human. We present four different regimes where human and robot collaboration can occur while performing the bin-picking task.

1) **Confirmative feedback:** If the robot recognizes the part but not accurately enough to act upon (occurs when the system estimates a small recognition error), then it requests the human for confirmative feedback. This is accomplished by displaying the target part and the bin image (with the recognized part highlighted) side-by-side on a monitor. Now, the human responds to a suitable query displayed on the monitor (e.g., Does the highlighted part match the target part?) with Yes/No answer. If the human’s answer is "Yes", the robot proceeds with picking up the part. If the answer is "No", the human then provides a simple cue by casting a laser pointer on the correct part. This thereby resolves the ambiguity in identifying and locating the part.

2) **Positional cues:** Part overlaps and occlusions may lead to recognition failures by the sensing system. In these situations, the robot requests additional help from the human. The human responds by identifying the part that matches the target part and conveying this information to the robot by directly shining the laser on that part.

3) **Orientation cues:** Apart from the part’s location, the robot also requires orientation information in order to perform the assembly operation properly. The human can provide the postural attributes of the part by casting the laser on each visible face of the part and labeling it using a suitable interface. This information can be used by the system to reconstruct the part orientation.

4) **De-clutter mode:** Suppose, the human predicts that in the current part arrangement, there exists no posture in which the robot can successfully grasp and pick up
the part; for example, when the part is positioned in an unstable posture and any attempt to hold the part causes it to tip over and slip from the robot's grasp. The human predicts this in advance and issues a de-clutter command. The robot responds by using raster-scan like movements of its gripper to randomize the part cluster. If the human predicts that the part can be grasped in the changed state, then he/she continues to provide location and postural cues to the robot. Otherwise, the de-clutter command is issued again. De-cluttering can also be used when the target part is completely hidden under the random pile of parts, thereby making itself unnoticeable to both the robot and the human.

V. ILLUSTRATIVE EXPERIMENTS

We create four representative parts (Fig. 5) that afford different recognition and grasping complexities to illustrate various challenges encountered during the bin-picking task. Four copies of each part are made to give rise to a total of sixteen parts. A few bin scenarios of varying complexities generated using these parts are shown in Fig. 6

Illustration 1: The bin scenario shown in Fig. 7(a) represents a case where the robot is tasked with picking Part 1, but it is difficult to detect the part due to overlap with other parts. Therefore, robot requests human assistance and the human responds by directly shining a laser on the part (shown in the figure). The robot uses this information to locate and pick up the part from the bin (Figs. 7(b)-(d))

Illustration 2: The bin scenario shown in Fig. 8 represents a case where the parts are highly cluttered disallowing detection of parts hidden under the pile. In this situation, the human issues a de-clutter command to the robot. Figures 8(a)-(c) show how the robot uses back-and-forth motions to shackle the pile, thereby enabling detection of the hidden parts.

Illustration 3: The bin scenario shown in Fig. 9(a) represents a case where the part to be picked (Part 1) is in an unstable position. That is, any attempt by the robot to pick up the part results in the part tipping over and slipping from the robot’s grasp (Figs. 9(b) and (c)). However, as the human can predict this event before hand, he/she can force the robot to de-clutter the pile before attempting to pick up the part. This is illustrated in Fig. 9(d)-(f).

VI. CONCLUSIONS

We presented design details of hybrid cell framework that enables safe and efficient human robot collaborations during assembly operations. We used illustrative experiments to present different cases in which human robot collaboration can take place to resolve perception/grasping problems encountered during bin-picking. In this paper, we considered bin-picking used for assembly tasks. However, our approach can be extended to the general problem of bin-picking as applied to other industrial tasks like packaging. More experiments based empirical evaluations, using the Baxter robot, are in order for systematically testing the ideas presented in the paper.

ACKNOWLEDGEMENT

This work is supported by the Center for Energetic Concepts Development. Dr. S.K. Gupta’s participation in this research was supported by National Science Foundation’s Independent Research and Development program. Opinions expressed are those of the authors and do not necessarily reflect opinions of the sponsors.

REFERENCES


of the 41st International Symposium on Robotics and 6th German Conference on Robotics (ROBOTIK 2010).


