Mutual-cognition for proactive human–robot collaboration: A mixed reality-enabled visual reasoning-based method

Shufei Li, Yingchao You, Pai Zheng, Xi Vincent Wang & Lihui Wang

View supplementary material

Published online: 11 Mar 2024.

Submit your article to this journal

Article views: 335

View related articles

View Crossmark data

OPEN ACCESS

Check for updates

Mutual-cognition for proactive human–robot collaboration: A mixed reality-enabled visual reasoning-based method

Shufei Li[a,b] , Yingchao You[c] , Pai Zheng[a] , Xi Vincent Wang[b] , and Lihui Wang[b]

[a]Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region; [b]Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden; [c]School of Engineering, Cardiff University, UK

ABSTRACT

Human-Robot Collaboration (HRC) is key to achieving the flexible automation required by the mass personalization trend, especially towards human-centric intelligent manufacturing. Nevertheless, existing HRC systems suffer from poor task understanding and poor ergonomic satisfaction, which impede empathetic teamwork skills in task execution. To overcome the bottleneck, a Mixed Reality (MR) and visual reasoning-based method is proposed in this research, providing mutual-cognitive task assignment for human and robotic agents' operations. Firstly, an MR-enabled mutual-cognitive HRC architecture is proposed, with the characteristic of monitoring Digital Twins states, reasoning co-working strategies, and providing cognitive services. Secondly, a visual reasoning approach is introduced, which learns scene interpretation from the visual perception of each agent's actions and environmental changes to make task planning strategies satisfying human–robot operation needs. Lastly, a safe, ergonomic, and proactive robot motion planning algorithm is proposed to let a robot execute generated co-working strategies, while a human operator is supported with intuitive task operation guidance in the MR environment, achieving empathetic collaboration. Through a demonstration of a disassembly task of aging Electric Vehicle Batteries, the experimental result facilitates cognitive intelligence in Proactive HRC for flexible automation.

1. Introduction

The primary goal of Industry 5.0 is to create sustainable, human-centric, and resilient manufacturing systems (Xu *et al.*, 2021). Towards human-centric smart manufacturing, enterprises are struggling for existence, due to the following challenges:

1. Transformable production required by mass personalization, such as tight changeover time when new products with variability are introduced to the market (Zhang *et al.*, 2022)
2. Large scale production of complicated and fine-fabricated mechanical components, such as assembly of a multistage car body (Wang *et al.*, 2021)
3. Occupational risk factors, such as musculoskeletal disorders among employees caused by awkward posture excessive effort, and repetitive movements (Carnahan *et al.*, 2001).

To tackle the strict requirements in manufacturing, Human–Robot Collaboration (HRC) provides a prevailing solution, which combines human cognitive flexibility and adaptability and robots' high accuracy, strength, and repeatability (Wang, Liu, Liu, and Wang, 2020). Inside a shared workspace, Proactive HRC systems allow human and robot participators to carry out manufacturing tasks qualified for their capabilities based on a holistic understanding of human–robot–workspace relations and task procedural knowledge, improving overall production efficiency (Li, Wang, Zheng, and Wang, 2021). Characterized by flexible automation, HRC is burrowing deep into today's production architecture.

The successful application of HRC systems relies on its context awareness capability (Wang *et al.*, 2022), which allows humans and robots to understand the surrounding environment and task operation goals. To date, emerging technologies, such as Mixed Reality (MR), Augmented Reality (AR), and computer vision, provide solutions for the perception of symbiotic relationships of the two participators. To eliminate safety risks, Hietanen *et al.* (2020) developed an interactive AR system, from which the human operator could obtain dynamic robot status and safety zone changes in the workspace. For precise robot control and handover, Amorim *et al.* (2021) fused 3D vision sensors and inertial measurement units (IMUs) to realize robust human

position tracking in millimeter precision. Along the task process, multimodal communication is essential for on-demand adjustment of task policy in collaboration, which can be achieved by haptic feedback (Tannous et al., 2020), gestures command (Mazhar et al., 2019), and an intuitive interface (Esengün et al., 2023).

Despite the above research efforts, the context awareness in HRC scenarios is limited to a non-semantic perception level, which fails to provide mutual-cognitive intelligence and knowledge of proactive collaboration desired by humans and robots. In detail, nowadays HRC applications fall into a stiff master–slave mode, in which either a robot or human agent needs to follow pre-defined instructions with task progression. To bridge the gap of cognitive co-working decisions, some previous works attempted to distill manufacturing knowledge (Zheng et al., 2022) for dynamic task fulfillment strategy generation (Li et al., 2022). Nevertheless, how to transmit the generated task planning strategies to humans and robots in a natural manner and what components should be included in a complete HRC system deserve more exploration. In addition, human operators in today's HRC systems lack a perception capability to know what is unknown now and what may happen in the future, such as a robot's next motions. Lastly, the previous study fails to consider ergonomics concerns, which are key elements to understanding human physical states. Robot cognitive intelligence remains unattained without assurance of safe, ergonomic, and proactive co-working with humans. The lack of either enhanced human perception or robot cognition makes it difficult to achieve empathetic teamwork skills in HRC systems, which impedes operation comfortability and adaptability along the overall manufacturing process.

Aiming to fill this research gap, an MR-enabled visual reasoning-based method is proposed to realize mutual-cognition intelligence for Proactive HRC. The mutual-cognitive strategy is derived from real-time scene graphs of human–robot operational sequences and then transmitted to the MR execution loop, where the robot catches on and plans for human needed manipulation, while the human operator is supported with intuitive guidance of manual operations from the MR interface. Meanwhile, the robot manipulation meets ergonomic human posture needs and human common task goals, which reflects empathic teamwork skills. The remainder of this article is organized as follows. Section 2 reviews recent related works for HRC implementation, especially for core techniques. The mutual-cognitive HRC framework, comprising its visual reasoning model, safe and ergonomic robot motion planning, and MR execution loop, is proposed in Section 3. Section 4 evaluates the significant performance of our HRC system in terms of a typical disassembly task of aging Electric Vehicle Batteries (EVBs). Section 5 discuss the achievements of the research. Finally, conclusions and future works are given in Section 6.

2. Related work

In this section, the Proactive HRC paradigm is elicited for true complementary of human and robot skills in manufacturing. Then, cutting-edge technologies including MR-assisted robot skills and visual reasoning-based cognitive computing are reviewed, to discover the research gap and promote Proactive HRC evolvement towards mutual-cognitive intelligence.

2.1. Proactive HRC

Instead of non-configurable large-scale automation, HRC plays a crucial role in flexible manufacturing for improved overall productivity (Keung et al., 2022). In this context, Proactive HRC towards smart, cognitive and more adaptable systems was proposed to promote the evolution of the next waves of manufacturing systems (Li, Zheng, Liu, Wang, Wang, Zheng, and Wang, 2023). Mutual-cognition is one critical concern in the Proactive HRC system. In mutual-cognition HRC, a human operator can on-demand, intuitively interact with a mobile robot. Meanwhile, the robot can proactively plan motions with safety (Pecora et al., 2019) and ergonomics concerns.

For Proactive HRC implementation, numerous efforts have been explored to improve human–robot co–working satisfaction when manipulating complex workpieces. For example, Ajoudani et al. (2018) summarized advanced robot control modalities for physical and bidirectional human–robot interaction. Millot and Pacaux-Lemoine (2013) introduced a situation awareness ability into the human–machine system to cope with unknown situations. Rahman (2019) proposed a mutual trust model, which could control robot motions and simulate human actions. Vernon et al. (2016) discussed cognition in HRC from four perspectives, i.e., attention, action, goals, and intentions. Then, Khatib et al. (2021) estimated the uncertainty of the operator's motion to allow the robot's end-effector to follow a position and orientation desired by the human, achieving optimal robot motion for fluent collaboration while avoiding collisions.

2.2. MR-based communication and robot control

MR in manufacturing encapsulates Digital Twin (DT) models and an AR environment together. Beyond AR, which focuses on displaying objects via visual-physical fusion, the MR can analyze system physical states, simulate the system's condition in the future via the DT models, and further present the simulation information via an AR manner. Therefore, advanced MR technologies find widespread applications in HRC (Wang, 2022). Hietanen et al. (2020) developed an HRC system on a projector and wearable MR glasses, respectively. With the MR interface, the human operator obtained real-time robot states and safety zone changes in the shared workspace. For example, the MR-based execution loop provided human operators with online support (Kousi et al., 2019). The human user naturally communicated assembly status information to the robot, without needing any expertise in robotics. Hence, the MR-based communication can allow seamless information exchange between the two participants and intuitive domain knowledge support for operator assistance.

On the other hand, the MR-based robot programming approach frees HRC from predefined motion and allows for dynamic robot path adjustment, achieving accurate robot control (Wang, Wang, Lei, and Zhao, 2020). Yuan *et al.* (2020) developed a portable Virtual Reality (VR) system, where human operators could modify 3D points and guided the paths of robots for surface taping tasks. Bottani and Vignali (2019) utilized MR techniques to let humans directly guide or teach manipulation to the robots. Users can define 3D points and plan the robot path with an MR interface (Ong *et al.*, 2020). Besides, Hernández *et al.* (2020) exploited robotic motion planning to deal with users' high-level requests for robot manipulation, rather than low-level specific movements. The MR-based robot programming methods open the door to Proactive HRC systems which can dynamically plan proactive robot motions.

2.3. Visual reasoning for cognitive collaboration

The visual reasoning approach (Cooray *et al.*, 2020) aims to learn the relationships of perceived objects, which facilitates HRC scene parsing from a perception level to a cognitive level. Tang *et al.* (2019) composed dynamic tree structures to capture task-specific contexts for visual relationship cognition and the answering of questions. To reason about a visual question, Kim and Lee (2019) proposed a model of dynamics attention for focus transition, which obeyed the human prior towards shorter reasoning paths and produced more interpretable attention maps. Furthermore, the scene graph was introduced to learn structured knowledge between objects and their relationships (Shi *et al.*, 2019). These visual reasoning methods facilitate explainable semantic understanding of different scenarios, which builds a bridge for mutual-cognition generation in HRC tasks.

For cognitive HRC, Ahn *et al.* (2018) leveraged a Text2Pickup network to allow robots to generate proactive decisions based on visual observations of picking objects. When confusing which objects were desired by the human, the robot generated interactive questions to the human for further communication. Besides, with visual and language cues, Venkatesh *et al.* (2020) proposed a neural network to allow the robot to reason about object coordinates in picking and placing tasks. In these systems, the robot can infer human intentions and target objects for mutual-cognitive co-working.

From the literature, one can find that mutual-cognitive intelligence allows HRC systems to distill production knowledge for bidirectional desired collaboration, which is critical to the evolution of an HRC. Our previous works have explored the scene graph (Li *et al.*, 2022) and knowledge graph (Zheng *et al.*, 2022) methods to make task-planning decisions in HRC systems. However, these previous studies focus on the task allocation part, while seldom considering human-centric needs in the execution process. Firstly, a human cannot perceive a robot's next operation goal and obtain on-demand knowledge support in an intuitive manner. Then, a robot fails to adjust operation postures for easy and comfortable human interaction, lacking ergonomic concerns. Motivated by this situation, this work demonstrates a mutual-cognitive HRC system that integrates intuitive human assistance, proactive robot motion, and ergonomic interaction, by integrating perception, decision-making, and control modules.

3. Methodology

This section depicts an architecture of MR-enabled mutual-cognitive HRC, followed by a visual reasoning approach for cognitive co-working strategy generation, and robot motion planning.

3.1. MR-enabled mutual-cognitive HRC architecture

The system architecture of MR-enabled mutual-cognitive HRC is presented in Figure 1, which consists of ergonomic collaboration in physical spaces, visual reasoning modules and virtual replicas in cyber spaces, and cognitive services in MR spaces. The combination of physical and cyber spaces is the HRC DT, which updates physical system changes, previews digital states, and makes co-working decisions. The HRC DT is embedded into the MR space for virtual-physical tracking and registration. Meanwhile, the MR system translates co-working decisions that are the response to human–robot mutual operation needs and task properties into cognitive services. These cognitive services enhance human flexibility (e.g., intuitive suggestions) and ensure proactive robot manipulation (e.g., robot trajectory preview). The proposed architecture allows empathetic HRC, whose connotation is represented by mutual-needed operation support (Li, Zheng, Pang, Wang, and Wang, 2023), ergonomic interaction, and an immersive teamwork environment between a human and a robot. In this context, the mutual-cognitive HRC can maximize human wellbeing and sustain production excellence in manufacturing tasks.

In the physical space, a sensing and monitoring system is developed to perceive human–robot states and surrounding environment changes. In detail, human skeleton joints, industrial parts, and geometric point clouds are detected by Resnet 50 (Li *et al.*, 2020), OpenPose (Li, Fan, Zheng, and Wang, 2021), and OctoMap (Duberg and Jensfelt, 2020), respectively, based on the output of a visual sensor. ROS (Robot Operating System) is deployed in an edge server to collect robot status and feedback control commands on-site.

The cyber space updates physical HRC settings to virtual replicas for visualization and preview in the MR environment. For instance, dynamic changes of human actions, robot operations, and task stages are transmitted to digital HRC models. In turn, proactive robot path planning can be verified in digital models, then translated into physical execution. At the same time, a visual reasoning module is utilized to construct relations between humans, robots, environment, and task structures. Co-working decisions can be inferred from the mutual-cognitive understanding of HRC relations in task processes. Thus, the decisions meet bidirectional human–robot operation needs and dynamically assign human and robot roles in HRC tasks.
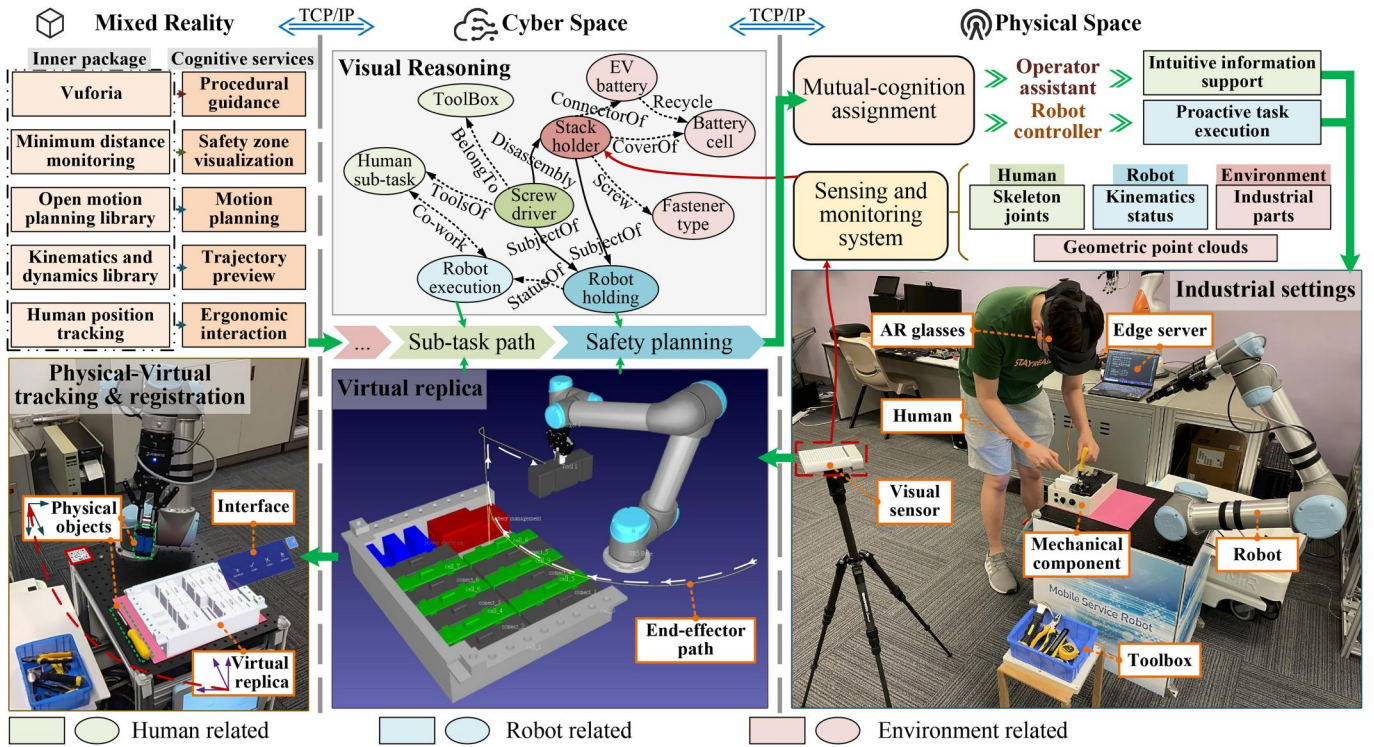
Figure 1. The architecture of MR-enabled mutual-cognitive HRC systems.

Based on physical-virtual tracking and registration, MR-based cognitive services are provided for HRC systems, which consist of intuitive information support for human operators and proactive task execution for robots. In detail, procedural guidance including text, videos, and visualized operation sequences can be delivered to humans based on Vuforia Toolbox. By continuously calculating the minimum distance between human and robot ontology, the safety zone for human operation is visualized in the MR environment. For robot control, the open motion planning library in ROS can proactively plan robot motions for different task execution, while the kinematics and dynamics library can achieve robot trajectory preview by a physical-virtual fusion manner in the MR space. With online human position tracking in OpenPose, the system can analyze ergonomics risks of human skeleton poses and plan ergonomic robot operations for easy interaction. These cognitive services are on-demand delivered to human and robotic agents based on co-working decisions, for human–robot empathetic teamwork.

3.2. Visual reasoning for mutual-cognition generation

To enable empathic understanding of the teamwork required between humans and robots, a scene graph-based visual reasoning module is utilized to infer their operation needs along task fulfillment and generates mutual-cognitive co-working strategies. As shown in Figure 2, the visual reasoning module contains four parts, (i) scenario perception, (ii) temporal node updating, (iii) dynamic graph construction, and (iv) cognitive strategy mapping. The scenario perception part consists of object detection and human body skeleton estimation, which are leveraged to locate industrial parts among the workspace

and track the motion of the joints in the human skeleton. With the perceptual results, nodes of working-in-progress objects are activated and their attributes are updated. Then, scene graphs are dynamically constructed by connecting perceived objects (nodes) with corresponding relations (edges). Lastly, different scene graphs are mapped to mutual-cognitive co-working strategies by learned graph embeddings, which represent an interpretation of current human–robot operations. In terms of this workflow, the stepwise procedures to achieve the visual reasoning approach are depicted in the figure.

3.2.1. Scenario perception for temporal node updating

Scenario perception is the prerequisite of semantic knowledge inference. As presented in the left corner of Figure 2, Resnet 50 is utilized to detect objects in HRC scenarios, including various industrial parts and the motion of robots. The Resnet model predicts spatial locations and categories of different objects in input images. The output of the object detector is denoted by a bounding box $v_i = [x_i, y_i, w_i, h_i]$ and a label $c_i \in \{1, ..., k\}$, where k is the number of object categories. Meanwhile, OpenPose is introduced to track the human skeleton from images, where the output of human hands is similarly formulated to location v and categories c. These temporal perceptual results are fed into the subsequent procedure and activated as nodes V in scene graphs. The attributes of nodes are updated by the matrices $v \in \mathbb{R}^{n \times 4}$ and $c \in \mathbb{R}^{n \times k}$, with different objects perceiving along the time.

3.2.2. Link prediction for dynamic graph construction

The link prediction is proposed to connect perceived objects with the most related relation, i.e., node pairs. The relation
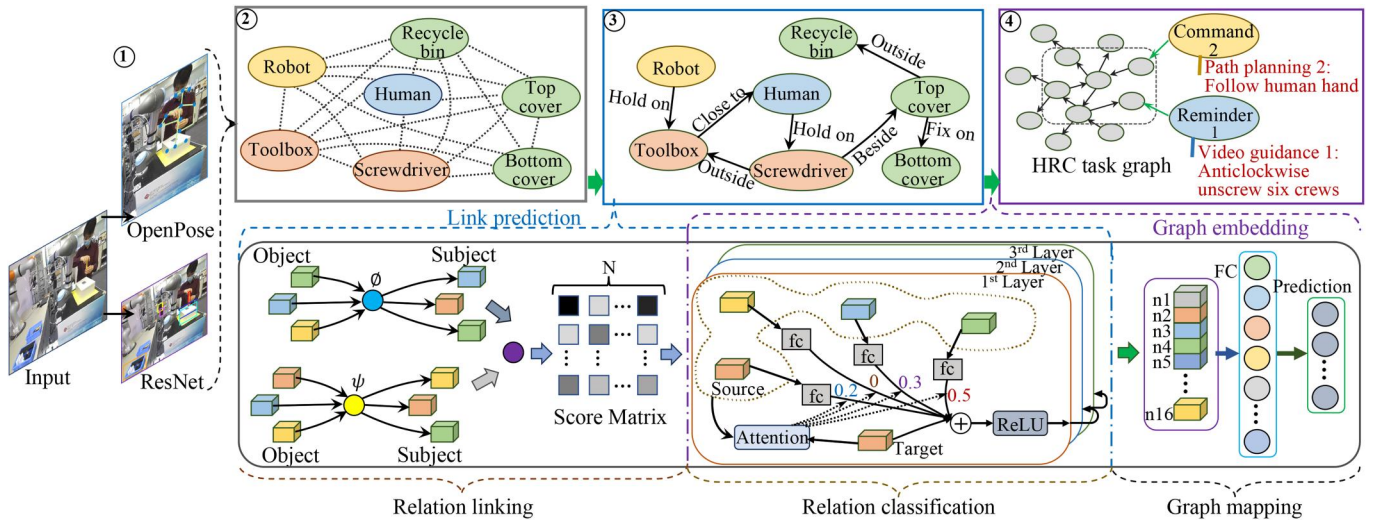
Figure 2. The workflow of visual reasoning-based mutual-cognitive strategy generation.

is an edge between a subject and an object of a pair of nodes. The process of link prediction contains relation linking and relation classification, as shown in the left bottom corner of Figure 2. The scene graph is dynamically constructed by linking edges between nodes. In this context, a two-layer perceptron is introduced to prune superfluous node pairs. The relatedness r_{ij} of $n \times (n-1)$ node pairs $\{x_i, x_j | i \neq j\}$ is defined as following,

$$r_{ij} = f(x_i, x_j) = \langle \phi(x_i), \psi(x_j) \rangle, i \neq j \tag{1}$$

where the relatedness function $f(\cdot, \cdot)$ is computed by a matrix multiplication of $\phi(\cdot)$ and $\psi(\cdot)$. A two-layer perceptron is utilized for the projection process of x and output $\phi(\cdot)$ and $\psi(\cdot)$, respectively. The vector x includes categories c and location v of an object. Then, a sigmoid function is applied on the r_{ij} to generate the relatedness score from zero to one. The top K node pairs are obtained by ranking the relatedness scores in descending order. Among these candidates, nodes that half overlap other nodes in spatial regions are filtered out. The followed by connecting the remaining node pairs with correct relation types in a scene graph.

A three-layer attentional Graph Convolutional Network (GCN) is proposed to extract contextual information between node pairs and predict the type of edges in scene graphs, as presented in the middle bottom corner (i.e., relation classification) of Figure 2. Firstly, a linear transformation w is used to extract features of neighboring nodes x_j for a target node x_i. These features are adjusted via weights α and added together, then are activated by a non-linear function σ, i.e., ReLU. The propagation of feature representations across layers of GCN is denoted as follows,

$$x_i^{(l+1)} = \sigma\left(x_i^{(l)} + \sum_j \alpha_{ij} w x_j^{(l)}\right) \tag{2}$$

where α adjusts the attention to node features, which allow the capture of information key node pairs, such as the robot node and a grasped object. The attention between a target node x_i and its source neighboring node x_j can be calculated by,

$$u_{ij} = w_h^T \sigma(W_a[x_i, x_j])$$
$$\alpha_{ij} = \text{Softmax}(u_{ij}) \tag{3}$$

where w_h and W_a are parameters of a two-layer perceptron, respectively. With the obtained node pairs and their relation types, a scene graph is dynamically constructed by linking edges E to nodes V, as part of the HRC task graph shown in the right upper corner of Figure 2.

3.2.3. Graph embedding for cognitive strategy mapping

With a scene graph dynamically constructed from perceived objects, the next step is to learn the graph embedding and map it to human reminders and robot commands as mutual-cognitive task strategies. The graph embedding module involves a relation classification network and a graph mapping part, as presented in the right bottom corner of Figure 2. As mentioned above, some node pairs in scene graphs integrate implicit interpretations of human–robot teamwork. For example, the pair of a human node and a manipulated industrial part contains human operation intentions, whereas contact hazard may be reflected in a pair of a human node and a robot node. In this context, skip-connect edges are also added among all nodes, which are utilized to directly extract information between nodes. Therefore, the scene graph consists of three different kinds of connections, namely, from *subject* to *relation*, from *relation* to *object*, and from *object* to *object*. The three-layer attentional GCN is leveraged to extract feature representations across these various connections. With the neighboring nodes x_j represented by a matrix $X \in \mathbb{R}^{d \times T_n}$, (2) can be re-formulated to $x_i^{(l+1)} = \sigma(W X^{(l)} \alpha_i)$, where d and T_n are the dimension and the amount of x_j, respectively. Following this notation, the feature transformation of nodes among GCN layers is defined as,

$$X_i^o = \sigma(\overbrace{W^{skip} X^o \alpha^{skip}}^{\text{Other nodes}} + \overbrace{W^{sr} X^r \alpha^{(sr)} + W^{or} X^r \alpha^{(or)}}^{\text{Neighboring Relations}}) \tag{4}$$

where s=*subject*, r=*relation*, and o=*object*. The first part in (4) concerns the features of the skip-connect nodes, whereas

the other one is for neighboring relations. Similarly, the representations of relations are propagated as,

$$X_i^r = \sigma(X_i^r + \overbrace{W^{rs}X^o\alpha^{(rs)} + W^{ro}X^o\alpha^{(ro)}}^{\text{Neighboring nodes}}) \qquad (5)$$

The last procedure is graph mapping, as shown in the right bottom corner of Figure 2. A Fully Connected (FC) layer is stacked on the three-layer attentional GCN, to linear transform the extracted feature representations. Lastly, a Softmax function is connected to the FC layers to learn the graph embedding and map it corresponding human prompts and robot commands, respectively.

The scene graph construction and embedding process are learned with three stepwise supervision training procedures. For the relation linking, a binary cross entropy loss is deployed during the training process. For the relation classification, a multi-class cross entropy loss is used. For the graph mapping, two other multi-class entropy losses are developed for the mapping of human prompting and robot commands, respectively.

3.3. Safe, ergonomic, and proactive robot motion planning

With task planning strategies inferred from the visual-reasoning module, a robot can perform operations desired by humans in a shared workspace. As presented in Figure 3, to improve human wellbeing and ensure mutual-cognitive capabilities among the co-working agents, a robot executes manipulation following safe, ergonomic, and proactive standards, which are achieved by the fusion of three modules, (i) real-time collision avoidance, (ii) ergonomic interactive actions, and (iii) proactive trajectory generation. Firstly, a real-time collision space is obtained from RGBD (i.e., color images and depth information) data of on-site

workspaces. The collision space provides constraints when generating robot action trajectories. Then, interactive actions between human and robotic agents (e.g., handover) are designed within ergonomic requirements to alleviate a worker's fatigue. With these concerns and assigned robot tasks, a rapid robust motion planning algorithm is adapted to proactively generate robot trajectories. The detailed methodologies of robot control are depicted as follows.

3.3.1. Collision avoidance based on real-time obstacle space

To ensure the safety of both humans and robots, it's necessary to determine contact hazard regions to which a robot cannot move, i.e., an obstacle space. An obstacle space indicates potential collisions between a robot and static obstacles (e.g., tables) and dynamic obstacles (e.g., human body). In a motion planning process, a real-time obstacle space is normally built following three steps. Firstly, a 3D occupancy grid mapping approach, OctoMap, is utilized to realize the representation of an obstacle space in HRC systems. With RGBD data of the on-site workspace, the OctoMap algorithm updates a real-time 3D map of static and dynamic obstacle spaces. Then, the manipulated object is eliminated from the obstacle space to allow the robot to manipulate the target object. The step is achieved by removing surrounding areas of manipulated objects in the obstacle space, based on the position and size of the manipulated object perceived by the object detector. Finally, the kinematic information of robots is obtained from ROS and then visualized on the 3D map. The map indicates collision regions to be avoided for robot motion planning.

3.3.2. Ergonomic interactive action design

The ergonomic interactive action design aims to improve teamwork comfort and eliminate occupational health risks
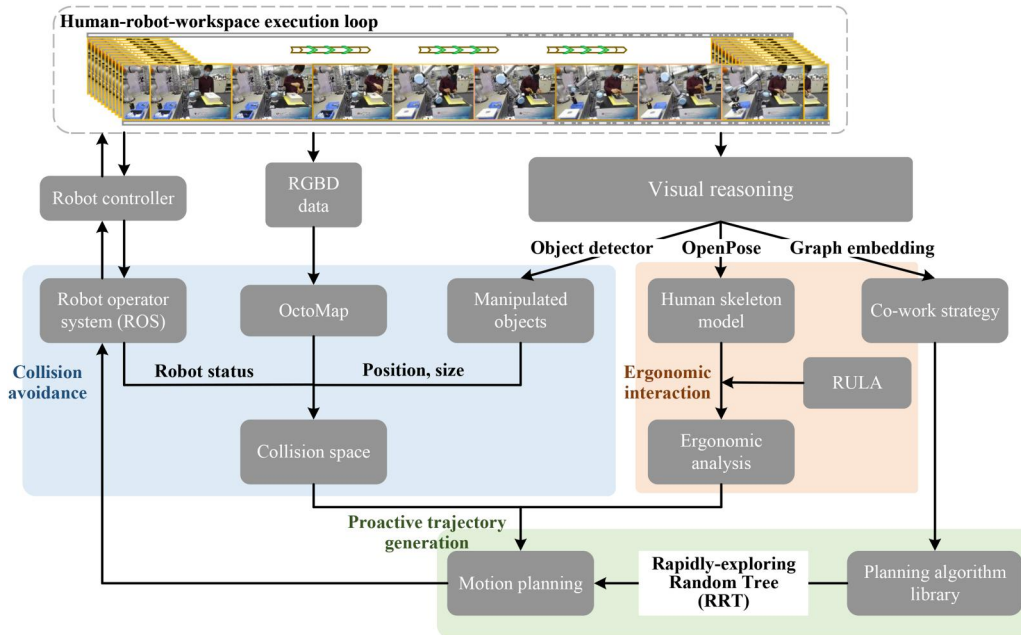

Figure 3. The procedural process of robot motion planning.

for human operators in HRC systems. The interactive actions contain direct contact between human–robotic agents and the handover of manipulated objects, which are essential operations in HRC tasks. To achieve ergonomic interaction, an upper limb assessment method RULA (McAtamney and Corlett, 1993) is leveraged to design the interactive space of the robot, such as the position and orientation of a handover point. The interactive space can be reached by human hands easily and comfortably, whose setting meets the following rules: (i) the range of movement of the upper arm is from $20°$ extension to $20°$ of flexion; (ii) the range of the lower arm is in 60-100° flexion; and (iii) the wrist is in a neutral position.

Following these requirements, a 5-DOF kinematic model of the human arm is introduced to obtain the robot's interactive space, as presented in Figure 4. In detail, the shoulder joint has three degrees of freedom, respectively represented by shoulder adduction R_A, shoulder flexion R_F, and shoulder rotation R_R. The elbow is defined as a joint R_E with one degree of freedom. The wrist is denoted as a joint R_W. The coordinate of one above joint i is denotes as p_i, whereas the angle is represents as θ_i. The upper arm, the lower arm and the hand are represented by linkages. Then, a three-dimensional cartesian coordinate system is built, with the human neck point n as its origin. The body's relative direction is used as the axis direction of the coordinate system. Forwards of the human is the Y-axis direction, right is the X-axis direction and up is the Z-axis direction. Based on the forward kinematic of linkage, the coordinate of the palm, which is the human–robot interactive space, can be obtained through the following equation:

$$p_h = d_n + A_a A_f A_r (d_h + A_e (d_f + A_w d_w))\qquad(6)$$

where d_n is the distance between neck and shoulder, whereas d_w is the distance from wrist to palm. d_h and d_f are

the length of the upper arm and the lower arm, respectively. A_i denotes the rotation matrix of joint i.

The value of the d_n, d_h, d_f, and d_w are obtained through real-time estimation of human skeleton joints. The rotation angle of the upper arm is denoted as $\theta = \arccos(\cos\theta_a \cos\theta_f)$, as the two joints θ_a and θ_f are perpendicular to each other. The rotation range of the lower arm is denoted as θ_e, while the range of the wrist is θ_w. In addition, to meet the requirement of RULA, the rotation range of human arm joints is suggested to be set as:

$$\begin{aligned} 0 &\leq \theta \leq 20° \\ 60° &\leq \theta_e \leq 100° \\ \theta_w &\approx 180° \end{aligned}\qquad(7)$$

By the calculation of the forward kinematic equation, the human-robot interactive space can be obtained. Then, the robot moves the end-effector to the interactive space, where the human can operate handover actions with the robot satisfying ergonomic requirements.

3.3.3. Motion planning for proactive trajectory generation

Robots are controlled by the motion planning algorithm to proactively conduct the operations of the co-working strategy generated by the visual reasoning module, such as picking-and-placing objects or handover. A motion planning algorithm, Rapidly-exploring Random Tree (RRT) (LaValle et al., 1998), is utilized to find continuous robot trajectories that move from an origin to a terminus. With concerns about avoiding the collision space, the RRT algorithm grows a tree from starting points to the ergonomic interactive points by using random samples from the configuration space. As each sample is found, a connection is attempted between it and the nearest points on the existing tree. The points will be added to the tree if the connection does not obey any constraints. Finally, a path from the start points to
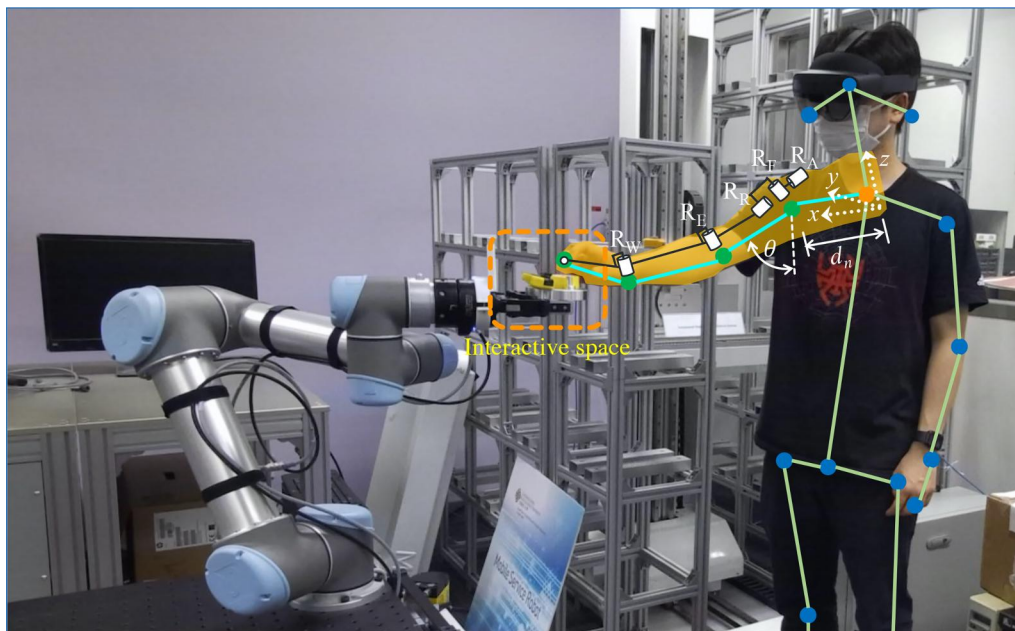

Figure 4. 5-DOF kinematic model of a human arm.

the ergonomic interactive destination can be dynamically generated for proactive robot task execution without collision.

4. Case study and experimental results

In this section, a prototype system of mutual-cognitive HRC is implemented on the disassembly task of EVBs. Then, the generation of cognitive task planning strategies is evaluated with the visual reasoning module. Finally, the mutual-cognitive intelligence in Proactive HRC is tested based on experimental results of intuitive human support and robot safe, ergonomic, and proactive manipulation.

4.1. Mutual-cognitive HRC for disassembly of EVBs

The disassembly task of EVBs remains a challenging problem that needs to be addressed due to the emergence of electric vehicles. In the lab environment, the task mainly consists of 11 substages, from delivering tools, unscrewing screws, opening cover, handover cover, testing electric power, cutting wires, removing glues, recycling PCB modules, recycling Thermo sensors, recycling Ion cells, and disposing of the bottom cover.

HRC provides an efficient solution for the disassembly of EVBs, as a human can complete some agile operations, whereas a robot can conduct dangerous operations. The system setting of mutual-cognitive HRC is presented in Figure 5. The on-site setup, edge server, cloud server, ROS platform, and robot controller are encapsulated in an MR execution loop. The MR glasses are HoloLens 2 produced by Microsoft, Washington. The mobile robot in the HRC system contains UR5 (Universal Robots, Odense) and MiR100 base (Mobile Industrial Robots, Odense). Among the loop, Azure Kinect (Tölgyessy *et al.*, 2021) is used to capture on-site images as the 11 substages progress. Human skeleton joints, industrial parts, and cloud points of the workspace in each disassembly stage are estimated in the edge server. The perceptual results are dynamically constructed to a scene graph via the visual reasoning module in the cloud server. An HRC task graph contains procedural knowledge of all these 11 substages, whereas the scene graph dynamically connects humans, robots, and their operation knowledge for each stage. The linked knowledge contains video guidance of human operations and robot path planning. The video guidance is transmitted to the MR glasses for human operation reminders, which give suggestions on how to uninstall components of EVBs step by step. The path planning commands are delivered to the ROS and a robot controller. Thus, the mobile robot can proactively conduct interactive actions with the human or take over dangerous subtasks, such as picking and placing battery cells. With the on-demand reminder support and proactive robot command, human and robotic agents complete the disassembly process of EVBs in a mutual-cognitive manner.

4.2. Visual reasoning for co-working strategy generation

The visual reasoning module is utilized to generate task planning strategies during the 11 disassembly stages of the EVBs. To evaluate the visual reasoning performance, a dataset is developed covering the 11 subtasks of the overall
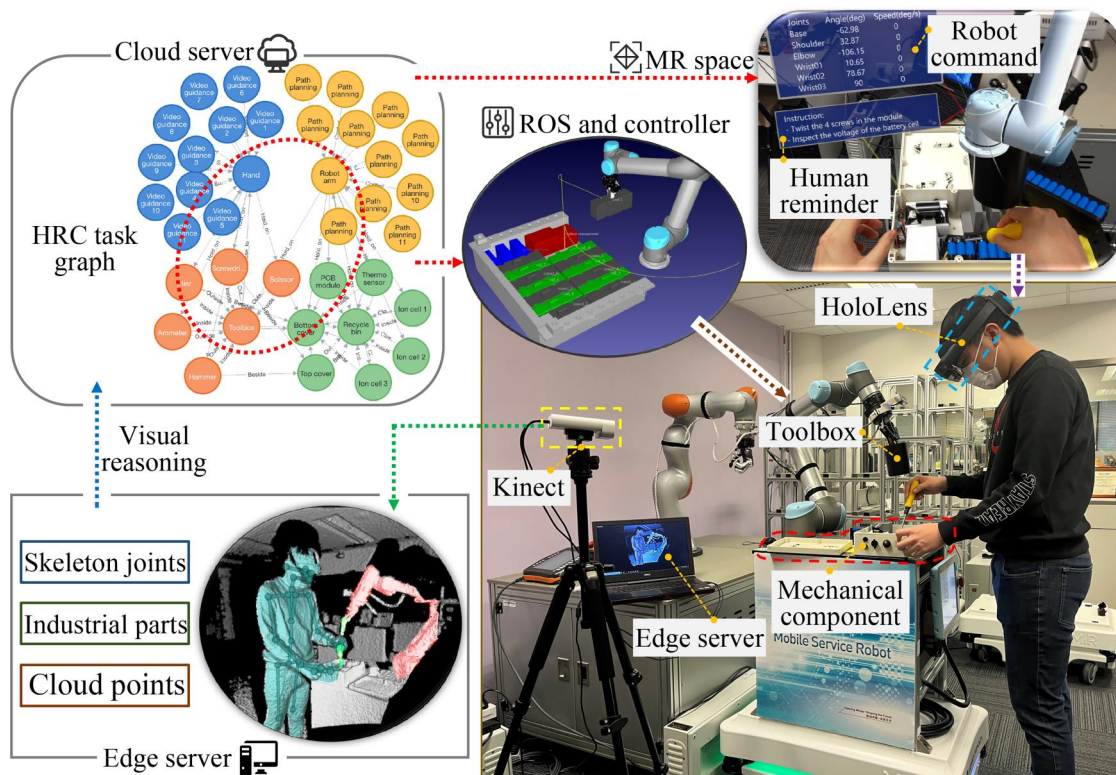

Figure 5. Prototype system setup for EVBs disassembly task.

disassembly procedure, which contains 779 RGB images and their depth information. Along with a human operator, 13 different industrial parts are included in the dataset, namely, Toolbox, Screwdriver, Ammeter, Plier, Scissor, Hammer, Recycle bin, Top cover, Bottom cover, PCB module, Thermo sensor, Ion cell, and Robot arm. For the annotation of the dataset, these industrial parts are labeled with a classified category and four coordinates of a bounding box. The relations between industrial parts in each image are annotated in the dataset. For these 11 disassembly stages, the dataset contains 11 kinds of video guidance and 10 different robot path plans as operation knowledge for various scenarios. The visual reasoning approach learns knowledge of operation intentions of human–robot teams and links suitable human reminders and robot commands for their cognitive disassembly co-working.

For the experiment setup, the dataset is divided into a training part (467 images) and a testing one (312 images). OpenPose is leveraged to estimate coordinates of 18 body skeleton joints from images. The number of categories k in Resnet 50 is set to 13. In this way, human hands and 13 different industrial parts can be firstly detected in the scenario perception part. For the link prediction algorithm, the parameter of node pairs with the most relatedness K is set to 128. The SGD (Stochastic Gradient Descent) optimizer is used to train the algorithm, with a learning rate of 0.001. For the graph embedding, the FC layer extracts features from the 14 graph nodes, i.e., one human node and 13 industrial part nodes. The graph embedding is trained with an SGD optimizer and a learning rate of 0.01. From scenario perception to scene graph embedding, the training processes are deployed on a Tesla V100 GPU (16G). For the testing, the trained model perceives various objects along different disassembly stages, dynamically connects relations of these objects as a scene graph, and triggers video guidance as human reminders and path planning as robot commands.

For the demonstration of the visual reasoning module, Figure 6 presents two examples of co-working strategy generation among stages of the testing of electric power and recycling PCB modules. As presented in the left part of Figure 6, the scene graph algorithm first identifies an electric power testing stage for the given HRC settings, then maps a human reminder and a robot command to this scene. The

human worker tests the electric power of three ion cells following video guidance, while the robot holds suitable tools to the human in close proximity. In the next stages, when the human holds a plier, the visual reasoning algorithm can infer human–robot operation intentions of loosening a PCB module. As presented in the right part of Figure 6, the video guidance on removing glue from the PCB module is delivered to the human operator. Meanwhile, the robot puts down the toolbox on a storage table, followed by recycling the PCB module. In this way, human and robotic agents learn about teammate operation goals and proactively conduct actions desired by each other in the EVBs disassembly task. In addition, the performance of the visual reasoning module for these 11 disassembly substages is shown in Table 1. For the SGGen + metric (Yang et al., 2018) in the second row, X/Y evaluates the graph construction accuracy, where X is the predicted result out of Y numbers of nodes, edges, and triplets in a scene graph. In the EVBs disassembly task, the nodes represent different industrial parts, while the edges represent the types of relationships between these nodes. Two nodes and their relation compose a triplet. The last row in Table 1 assesses the accuracy of the mapping between graph embedding to human reminders and robot commands through the visual reasoning module. This accuracy is calculated as the ratio of the correct predictions of the co-working strategy to the total number of predicted values.

4.3. MR-based operator assistance and robot control

In the disassembly process of EVBs, the generated co-working strategy is assigned to human-robot teams via the MR glasses. The human can obtain intuitive information support in the MR environment, as presented in Figure 7. Based on the co-working strategy, the MR glasses provides human operators with procedural guidance with virtual-physical fused visualization, such as video guidance of a manual operation. Meanwhile, safe zones of different levels are visualized in real-time to prompt human operators on safety concerns. The MR glasses also presents the robot trajectory preview before its execution, so that human operators can intuitively learn about the robot's next intended motion. In this context, the human operator is equipped with enhanced flexibility and cognition to make decisions on further
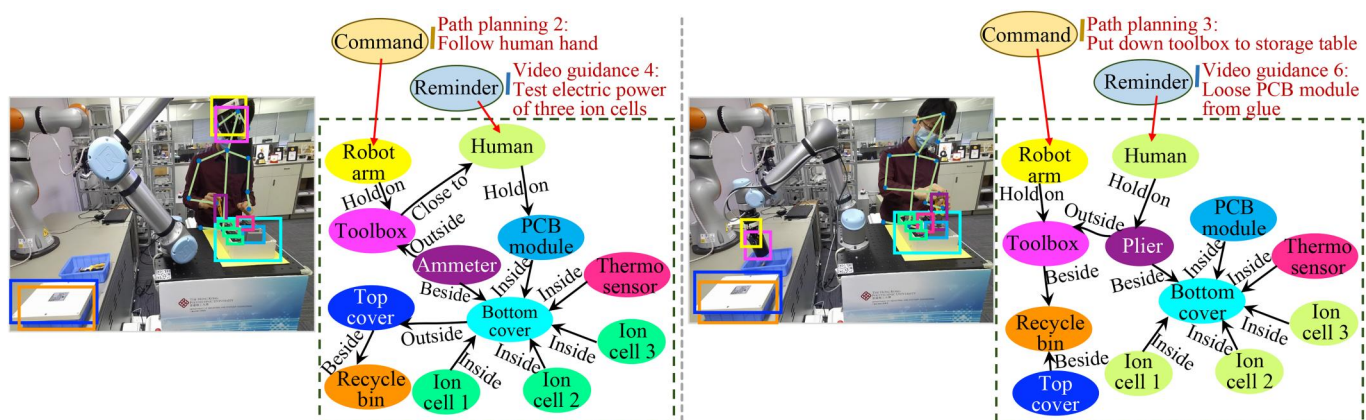

Figure 6. Examples of co-working strategy generation via visual reasoning.

Table 1. Accuracy of scene graph (SG) construction and co-working strategy generation.

HRC SG	SG1	SG2	SG3	SG4	SG5	SG6	SG7	SG8	SG9	SG10	SG11
SGGen+	13/14	14/14	21/21	20/21	34/34	32/34	34/34	33/34	34/34	33/34	32/34
Precision	91.67	95.00	96.36	91.67	94.44	95.74	96.30	92.86	96.00	93.75	92.31

Figure 7. System demonstration of MR-based information support and trajectory preview.

disassembly operations based on the suggestions from the MR environment.

On the other hand, the MR environment can simulate robot motions included in the co-working strategy by the HRC DT in advance. Then the motion planning commands can be transmitted to a robot for proactive task execution. In detail, the performance of robot task execution is evaluated from three aspects, i.e., the feasibility, safety, and ergonomics. The feasibility analysis focus on the validation of set functions that the system is capable of carrying out, such as whether the robot can generate a path to operate an assigned task or not. The safety analysis assesses the robustness of obstacle detection and collision avoidance of the system. The ergonomic analysis aims to assess whether the interactive points figured out by the system can be comfortably reached by human hand. The three considerations are given attention when the robot performs collaborative operations for the disassembly of EVBs. A physical-simulated platform, gazebo, is used as HRC DT to visualize the robot motion planning process in the three concerns.

Feasibility test. Receiving a command from the visual reasoning module, the HRC system generates corresponding trajectories, which are then executed by the robot actuators for human required operation. A widespread subtask of the robot which grasps, moves, and delivers toolbox to a position where human partners can take tools conveniently and comfortably, is used for demonstration. Figure 8(a) shows an execution stage of the subtask, whereas the entire generated trajectory is visualized in Figure 8(b).

Safety test. Collision avoidance is a prerequisite for HRC systems. With the same robot subtask, the planning module can generate a safe trajectory to avoid any collision and ensure the safety of human and robotic agents. As shown in Figure 8(c), an obstacle is perceived and added to the workspace between two agents, and blocks the movement of the robot arm. With the obstacle space dynamically updating based on perceptual results, the robot can circumvent these obstacles for a safe trajectory generation (see Figure 8(d)).

Ergonomic test. This test is designed to validate the comfortability of interactive actions in human–robot teams. When a human is working with different postures, the robot figures out a handover position that the human can reach

easily. As shown in Figure 8(e), the HRC system learns about how the human unfolds the top cover of EVBs and needs an Ammeter to test the electric power of Ion cells. Thus, the robot calculates an ergonomic space position and delivers the Ammeter to the human, with assigned commands. Figure 8(f) shows the position for the handover of an Ammeter from the robot to the worker. The handover points are obtained based on the forward kinematic equation. Specifically, the parameters $\theta_a=5°$, $\theta_f=5°$, $\theta_r=0°$, $\theta_e=80°$, and $\theta_w=180°$.

Ten participants, consisting of six males and four females aged between 23 and 30, with an average height of around 169 cm and an average weight of approximately 60 kg, were invited to take part in the test. Each participant was asked to complete the disassembly task of EVBs three times, for a total of 30 iterations. Each disassembly experiment consists of 11 substages. When the robot arm reaches the handover position, the participants can pick up objects delivered by the robot using various postures. Among the process, the skeleton joints of participants, including the movement of the upper arm, lower arm, and wrists, are obtained by the OpenPose tool. Followed by evaluation of human–robot handover gestures, 80% of participants' skeleton postures across all substages of the disassembly task fall within the suggested rotation ranges of the RULA rules in (7). The results suggest that the robot motion planning can robustly calculate an interactive position for human–robot handover aligning with the ergonomic requirement.

5. Discussions

The mutual-cognitive intelligence in Proactive HRC systems stands for empathic understanding between human–robot teams. For task cognition, the visual reasoning approach infers the required bi-directional operations by reasoning knowledge interpretation of human–robot–object relationships among current co-working scenarios from the explainable scene graph base. For enhanced human cognition, the MR execution loop allows for proactive communication among HRC systems, where essential suggestions and supports are transmitted to the human for the worker's
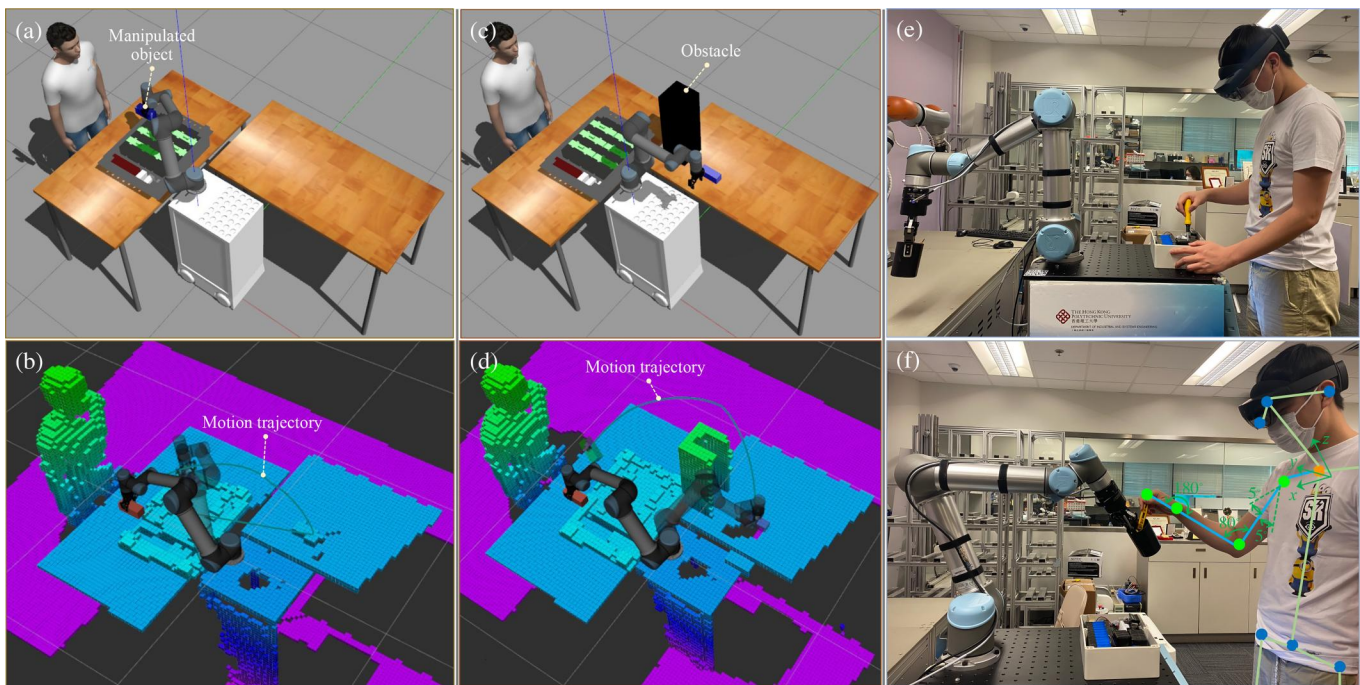
Figure 8. Feasibility, safety, and ergonomic test of robot motions.

improved decision making. For robot cognition, the robot conducts interactions with humans following ergonomics rules, such as handover position and orientation desired by the worker, which improves human wellbeing.

The MR-enabled visual reasoning-based method paves the way to the mutual-cognitive HRC systems, which prompt the next waves of human-centric intelligent manufacturing. Apart from the above advantages, the precision of co-working strategy generation can be improved with further experimental tests, for example, by eliminating the sample imbalance problem via data-augmented techniques. For the ergonomic test, there are two reasons why a few human skeleton models fail to meet RULA rules. One part is visual estimation errors of human skeleton points in OpenPose, whereas the other one is human movement uncertainty when moving towards a position. Lastly, the feasibility of the prototype system of mutual-cognitive HRC should be evaluated with other industrial cases, such as the assembly of complicated mechanical engines.

6. Conclusions

The notable shift to human-centric intelligent manufacturing elicits much interest in mutual-cognitive capability for Proactive HRC systems, which can help achieve trustworthy teamwork for flexible manufacturing automation. An MR-enabled visual reasoning-based architecture is explored to facilitate mutual-cognitive HRC evolution. In this context, the visual reasoning module stepwise perceives the on-site workspace, constructs a scene graph from the perceptual results, and maps task planning strategies by learning the graph embedding. Then, in the MR environment, the human operator receives suggestions and support from the co-working strategy for further suitable operations; meanwhile, the robot

obtains interpretation of current scenarios and conducts ergonomic, proactive operations. To summarize, the main scientific contributions achieved in this article are listed as follows:

1. A visual reasoning approach is proposed in the HRC system to advance its intelligence from perception base to mutual-cognition level. The reasoning module learns knowledge of human–robot relations in co-working processes by contextual scene graph, and infers task planning strategies addressing cooperation needs.
2. Safety, preview, and ergonomics rules of robot motions are established which bridge the gap for empathetic robot skills. The robot's control and manipulation enhance human context-awareness ability and response to human-centric needs through the visualization of safety rules, trajectory preview in the MR environment, and planning interactive positions that are feasible for human reach.

Except for these mentioned achievements, several research efforts should be further taken, which are highlighted here, including (i) mutual-cognitive capability when facing a new, but similar, HRC task, such as the intervention of new or different nodes in a scene graph; (ii) establishment of HRC knowledge base from multi-layers, e.g., task layer, mechanical component layer, and operation process layer; and (iii) predictable HRC task fulfillment with cognitive knowledge support even facing human motion uncertainty.

Funding

Notes on contributors

Shufei Li currently serves as a Postdoctoral Fellow within the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University. In 2023, he earned his PhD from the same department. Prior to this, in 2020, he obtained his MS in industrial and manufacturing systems engineering at Beihang University, following a BE in mechatronic engineering from Shandong Jianzhu University in 2017. His research interests lie in smart manufacturing and intelligent robotics field, including proactive human–robot collaboration, embodied AI, computer vision, and augmented reality.

Yingchao You is a PhD student at the engineering school, Cardiff University, UK. He received a bachelor's degree in industry engineering from Southwest Jiaotong University in Chengdu, China. He is interested in human-robot collaboration, learning from demonstration, and human-centric manufacturing.

Pai Zheng (SM'IEEE/CMES, M'ASME/SME, CIRP Research Affiliate) is currently an assistant professor, Wong Tit-Shing Endowed Young Scholar in Smart Robotics, and lab-in-charge of Digitalized Service Laboratory in the Department of Industrial and Systems Engineering, at The Hong Kong Polytechnic University. He received the dual bachelor's degrees in mechanical engineering (Major) and computer science and engineering (Minor) from Huazhong University of Science and Technology, Wuhan, China, in 2010, master's degree in mechanical engineering from Beihang University, Beijing, China, in 2013, and Ph.D. in mechanical engineering at The University of Auckland, Auckland, New Zealand, in 2017. His research interest includes human–robot collaboration, smart product-service systems, and smart manufacturing systems.

Xi (Vincent) Wang is an associate professor in the IPU Department of Production Engineering, KTH Sweden. He is working as the division head of Industrial Production Systems (IPS). He received his PhD and bachelor degrees in mechanical engineering from the University of Auckland (New Zealand) and Tianjin University (China), respectively in 2013 and 2008. In 2021 Vincent received his Docentship from KTH. Vincent's main research focus includes Cloud-based manufacturing, sustainable manufacturing, robotics, digital twin, computer-aided design, and manufacturing systems.

Lihui Wang is a chair professor at KTH Royal Institute of Technology, Sweden. His research interests are focused on cyber-physical production systems, human–robot collaborative assembly, brain robotics, and adaptive manufacturing systems. Professor Wang is actively engaged in various professional activities. He is the editor-in-chief of *International Journal of Manufacturing Research, Journal of Manufacturing Systems*, and *Robotics and Computer-Integrated Manufacturing*. He has published 10 books and authored in excess of 650 scientific publications. Professor Wang is a fellow of Canadian Academy of Engineering (CAE), International Academy for Production Engineering (CIRP), Society of Manufacturing Engineers (SME), and American Society of Mechanical Engineers (ASME). In 2020, he was elected one of the20 Most Influential Professors in Smart Manufacturing by Society of Manufacturing Engineers.

ORCID

Shufei Li http://orcid.org/0000-0002-5684-6756
Yingchao You http://orcid.org/0000-0002-4193-3304
Pai Zheng http://orcid.org/0000-0002-2329-8634
Xi Vincent Wang http://orcid.org/0000-0001-9694-0483
Lihui Wang http://orcid.org/0000-0001-8679-8049

References

Ahn, H., Choi, S., Kim, N., Cha, G. and Oh, S. (2018) Interactive text2-pickup networks for natural language-based human–robot collaboration. *IEEE Robotics and Automation Letters*, **3**(4), 3308–3315.

Ajoudani, A., Zanchettin, A.M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K. and Khatib, O. (2018) Progress and prospects of the human–robot collaboration. *Autonomous Robots*, **42**, 957–975.

Amorim, A., Guimares, D., Mendona, T., Neto, P., Costa, P. and Moreira, A.P. (2021) Robust human position estimation in cooperative robotic cells. *Robotics and Computer-Integrated Manufacturing*, **67**, 102035.

Bottani, E. and Vignali, G. (2019) Augmented reality technology in the manufacturing industry: A review of the last decade. *IISE Transactions*, **51**(3), 284–310.

Carnahan, B.J., Norman, B.A. and Redfern, M.S. (2001) Incorporating physical demand criteria into assembly line balancing. *IIE Transactions*, **33**(10), 875–887.

Cooray, T., Cheung, N.-M. and Lu, W. (2020) Attention-based context aware reasoning for situation recognition, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE Press, Piscataway, NJ, pp. 4736–4745.

Duberg, D. and Jensfelt, P. (2020) Ufomap: An efficient probabilistic 3d mapping framework that embraces the unknown. *IEEE Robotics and Automation Letters*, **5**(4), 6411–6418.

Esengün, M., Üstündağ, A. and İnce, G. (2023) Development of an augmented reality-based process management system: The case of a natural gas power plant. *IISE Transactions*, **55**(2), 201–216

Hernández, J.D., Sobti, S., Sciola, A., Moll, M. and Kavraki, L.E. (2020) Increasing robot autonomy via motion planning and an augmented reality interface. *IEEE Robotics and Automation Letters*, **5**(2), 1017–1023.

Hietanen, A., Pieters, R., Lanz, M., Latokartano, J. and Kämäräinen, J.-K. (2020) AR-based interaction for human-robot collaborative manufacturing. *Robotics and Computer-Integrated Manufacturing*, **63**, 101891.

Keung, K., Lee, C. and Ji, P. (2022) Industrial internet of things-driven storage location assignment and order picking in a resource synchronization and sharing-based robotic mobile fulfillment system. *Advanced Engineering Informatics*, **52**, 101540.

Khatib, M., Al Khudir, K. and De Luca, A. (2021) Human-robot contactless collaboration with mixed reality interface. *Robotics and Computer-Integrated Manufacturing*, **67**, 102030.

Kim, W. and Lee, Y. (2019) Learning dynamics of attention: Human prior for interpretable machine reasoning. *arXiv preprint arXiv:1905.11666*.

Kousi, N., Stoubos, C., Gkournelos, C., Michalos, G. and Makris, S. (2019) Enabling human robot interaction in flexible robotic assembly lines: An augmented reality based software suite. *Procedia CIRP*, **81**, 1429–1434.

LaValle, S.M. *et al.* (1998) *Rapidly-exploring random trees: A new tool.* Research Report 9811.

Li, S., Fan, J., Zheng, P. and Wang, L. (2021) Transfer learning-enabled action recognition for human-robot collaborative assembly. *Procedia CIRP*, **104**, 1795–1800.

Li, S., Wang, R., Zheng, P. and Wang, L. (2021) Towards proactive human–robot collaboration: A foreseeable cognitive manufacturing paradigm. *Journal of Manufacturing Systems*, **60**, 547–552.

Li, S., Zheng, P., Liu, S., Wang, Z., Wang, X.V., Zheng, L. and Wang, L. (2023) Proactive human–robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives. *Robotics and Computer-Integrated Manufacturing*, **81**, 102510.

Li, S., Zheng, P., Pang, S., Wang, X.V. and Wang, L. (2023) Self-organising multiple human–robot collaboration: A temporal subgraph reasoning-based method. *Journal of Manufacturing Systems*, **68**, 304–312.

Li, S., Zheng, P., Wang, Z., Fan, J. and Wang, L. (2022) Dynamic scene graph for mutual-cognition generation in proactive human-robot collaboration. *Procedia CIRP*, **107**, 943–948.

Li, S., Zheng, P. and Zheng, L. (2020) An AR-assisted deep learning-based approach for automatic inspection of aviation connectors. *IEEE Transactions on Industrial Informatics*, **17**(3), 1721–1731.

Mazhar, O., Navarro, B., Ramdani, S., Passama, R. and Cherubini, A. (2019) A real-time human-robot interaction framework with robust background invariant hand gesture detection. *Robotics and Computer-Integrated Manufacturing*, **60**, 34–48.

McAtamney, L. and Corlett, E.N. (1993) Rula: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, **24**(2), 91–99.

Millot, P. and Pacaux-Lemoine, M.-P. (2013) A common work space for a mutual enrichment of human-machine cooperation and team-situation awareness. *IFAC Proceedings Volumes*, **46**(15), 387–394.

Ong, S., Yew, A., Thanigaivel, N. and Nee, A. (2020) Augmented reality-assisted robot programming system for industrial applications. *Robotics and Computer-Integrated Manufacturing*, **61**, 101820.

Pecora, A., Maiolo, L., Minotti, A., Ruggeri, M., Dariz, L., Giussani, M., Iannacci, N., Roveda, L., Pedrocchi, N. and Vicentini, F. (2019) Systemic approach for the definition of a safer human-robot inter-action. *Factories of the Future: The Italian Flagship Initiative*, Springer, pp. 173–196.

Rahman, S.M. (2019) Cognitive cyber-physical system (c-cps) for human-robot collaborative manufacturing, in *2019 14th Annual Conference system of Systems Engineering (SoSE)*, IEEE Press, Piscataway, NJ, pp. 125–130.

Shi, J., Zhang, H. and Li, J. (2019) Explainable and explicit visual rea-soning over scene graphs, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE Press, Piscataway, NJ, pp. 8376–8384.

Tang, K., Zhang, H., Wu, B., Luo, W. and Liu, W. (2019) Learning to compose dynamic tree structures for visual contexts, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE Press, Piscataway, NJ, pp. 6619–6628.

Tannous, M., Miraglia, M., Inglese, F., Giorgini, L., Ricciardi, F., Pelliccia, R., Milazzo, M. and Stefanini, C. (2020) Haptic-based touch detection for collaborative robots in welding applications. *Robotics and Computer-Integrated Manufacturing*, **64**, 101952.

Tölgyessy, M., Dekan, M., Chovanec, L. and Hubinskỳ, P. (2021) Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors*, **21**(2), 413.

Venkatesh, S.G., Biswas, A., Upadrashta, R., Srinivasan, V., Talukdar, P. and Amrutur, B. (2020) Spatial reasoning from natural language instructions for robot manipulation. *arXiv preprint arXiv: 2012.13693*.

Vernon, D., Thill, S. and Ziemke, T. (2016) The role of intention in cognitive robotics. *Toward Robotic Socially Believable Behaving Systems-Volume I: Modeling Emotions*, Springer, pp. 15–27.

Wang, B., Zheng, P., Yin, Y., Shih, A. and Wang, L. (2022) Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective. *Journal of Manufacturing Systems*, **63**, 471–490.

Wang, C., Zhu, X., Zhou, S. and Zhou, Y. (2021) Bayesian learning of structures of ordered block graphical models with an application on multistage manufacturing processes. *IISE Transactions*, **53**(7), 770–786.

Wang, L. (2022) A futuristic perspective on human-centric assembly. *Journal of Manufacturing Systems*, **62**, 199–201.

Wang, L., Liu, S., Liu, H. and Wang, X.V. (2020) Overview of human-robot collaboration in manufacturing, in *5th International Conference on the Industry 4.0 Model for Advanced Manufacturing, AMP 2020*, Springer, pp. 15–58.

Wang, X.V., Wang, L., Lei, M. and Zhao, Y. (2020) Closed-loop aug-mented reality towards accurate human-robot collaboration. *CIRP Annals*, **69**(1), 425–428.

Xu, X., Lu, Y., Vogel-Heuser, B. and Wang, L. (2021) Industry 4.0 and industry 5.0—inception, conception and perception. *Journal of Manufacturing Systems*, **61**, 530–535.

Yang, J., Lu, J., Lee, S., Batra, D. and Parikh, D. (2018) Graph r-cnn for scene graph generation, in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, pp. 670–685.

Yuan, Q., Weng, C.-Y., Suarez-Ruiz, F. and Chen, I.-M. (2020) Flexible telemanipulation based handy robot teaching on tape masking with complex geometry. *Robotics and Computer-Integrated Manufacturing*, **66**, 101990.

Zhang, Y.-J., Huang, N., Radwin, R.G., Wang, Z. and Li, J. (2022) Flow time in a human-robot collaborative assembly process: Performance evaluation, system properties, and a case study. *IISE Transactions*, **54**(3), 238–250.

Zheng, P., Li, S., Xia, L., Wang, L. and Nassehi, A. (2022) A visual rea-soning-based approach for mutual-cognitive human-robot collabor-ation. *CIRP Annals*, **71**(1), 377–380.