Energy, Ecology and Environment
https://doi.org/10.1007/s40974-024-00327-7

ORIGINAL ARTICLE

AR-enhanced digital twin for human–robot interaction in manufacturing systems

Zhongyuan Liao[1] · Yi Cai[1,2,3]

[1] Division of Emerging Interdisciplinary Areas (EMIA), Academy of Interdisciplinary Studies, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2] Smart Manufacturing Thrust, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
[3] Department of Mechanical and Aerospace Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China

Received: 27 February 2024 / Revised: 29 April 2024 / Accepted: 6 May 2024

Abstract The integration of advanced technologies into manufacturing processes is critical for addressing the complexities of modern industrial environments. In particular, the realm of human–robot interaction (HRI) faces the challenge of ensuring that human operators can effectively collaborate with increasingly sophisticated robotic systems. Traditional interfaces often fall short of providing the intuitive, real-time interaction necessary for optimal performance and safety. To address this issue, we introduce a novel system that combines digital twin (DT) technology with augmented reality (AR) to enhance HRI in manufacturing settings. The proposed AR-based DT system creates a dynamic virtual model of robot operations, offering an immersive interface that overlays crucial information onto the user's field of vision. This approach aims to bridge the gap between human operators and robotic systems, improving spatial awareness, task guidance, and decision-making processes. Our system is designed to operate at three distinct levels of DT functionality: the virtual twin for in-situ monitoring, the hybrid twin for intuitive interaction, and the cognitive twin for optimized operation. By leveraging these levels, the system provides a comprehensive solution that ranges from basic visualization to advanced predictive analytics. The effectiveness of the AR-based DT system is demonstrated through a human-centric user study conducted in manufacturing scenarios. The results show a significant reduction in operational time and errors, alongside an enhancement of the overall user experience. These findings confirm the potential of our system to transform HRI by providing a safer, more efficient, and more adaptable manufacturing environment. Our research contributes to the advancement of smart manufacturing by evidencing the synergistic benefits of integrating DT and AR into HRI.

Keywords Digital twin · Intuitive interface · Augmented reality · Human–robot interaction · Human-centricity

✉ Yi Cai
yicai@hkust-gz.edu.cn

Zhongyuan Liao
zy.liao@connect.ust.hk

1 Introduction

The advent of Industry 4.0 has catalyzed a paradigmatic shift in the manufacturing sector, driven by the assimilation of advanced technological innovations such as the Internet of Things (IoT) (Sisinni et al. 2018), artificial intelligence (AI) (Wan et al. 2020), blockchain technology (Leng et al. 2023), augmented and virtual reality (AR/VR) (Eswaran and Bahubalendruni 2022), and Robotics (Bhatt et al. 2020). Despite Industry 4.0's predilection towards technology-centric approaches, there is an emerging trend towards human-centric objectives, as espoused by Industry 5.0 (Huang et al. 2022). This nascent paradigm emphasizes the pivotal role of human creativity, analytical prowess, and distinctive skills, synergizing with advanced technological systems. In the realm of robotics, the automation of monotonous and hazardous tasks has been instrumental in amplifying productivity and curtailing human error. With the advent of Industry 5.0, a critical focus of robotics advancement is the enhancement of human–robot interaction (HRI).

Given the fluidity of market trends and the complexity of product specifications, the resilience of HRI systems is integral to the adaptability and resilience of contemporary manufacturing infrastructures. To fortify the resilience of HRI, particularly in terms of rapid recovery amidst alterations to the manufacturing system, the establishment of an efficacious and intuitive HRI system is imperative. Such enhancement is contingent upon a nuanced understanding of user engagement paradigms with robotic entities. Researchers have investigated various interaction modalities to facilitate user communication with robots. Based on these modalities, HRI can be classified into seven categories:

- Touch-based Interaction: Users can engage with robots using touch screens on mobile devices, tablets, or interactive surfaces, which may involve dragging or drawing on a tablet (Chen et al. 2021), pointing and touching target positions, or navigating virtual menus on a smartphone (Cao et al. 2019). Touch interactions are beneficial for precise input and programming of robot motion (Fuste et al. 2020).
- Tangible Interaction: Users can modify the physical shape of robots (Lindlbauer et al. 2016), interact with them by handling tangible objects (Pedersen et al. 2011), or control them directly through grasping and manipulation (Özgür et al. 2017).
- Voice Interaction: Some studies have explored the use of voice commands for robot operation, especially in co-located environments (Arevalo et al. 2021).
- Gaze Interaction: Gaze is often used in combination with spatial gestures (Yuan et al. 2019), such as for menu selection (Arevalo et al. 2021).
- Spatial Gesture Interaction: Users can manipulate virtual waypoints (Quintero et al. 2018) or control robots using virtual menus (Cauchard et al. 2019) through spatial gestures, commonly utilized in head-mounted display interfaces. Spatial gestures can also be employed for remote management of swarm robots (Siu et al. 2018).
- Proximity Interaction: Proximity can be employed implicitly for robot communication (Hoang et al. 2022). For example, an AR trajectory can be updated to display the robot's intent as a passerby approaches (Watanabe et al. 2015), or a shape-shifting wall can change the content on the robot based on the user's behavior and position (Takashima et al. 2016).
- Pointer and Controller Interaction: Users can manage robots through spatial interaction or device actions with controllers that provide tactile feedback (Hedayati et al. 2018). Users can implicitly communicate with robots, such as creating a 3D virtual object with a 3D printer using pointer control (Peng et al. 2018).
- Electroencephalography or Surface Electromyography. For example, the Electroencephalography (EEG) signal was utilized to build an online brain-computer interface (BCI) system to control mobile robots (Lu et al. 2022). However, this method is not stable because of the current technology limit.

Various interaction techniques are commonly used to operate industrial robots, such as joystick-based, screen-based, and tangible-based kinesthetic teaching methods. Tangible-based kinesthetic instruction, which enables workers to directly manipulate a robot to record its trajectory, is a valuable hands-on teaching method. However, its application in the context of industrial robotics raises critical safety concerns. The considerable size of these robots, coupled with their high-powered actuators and the possible inclusion of hazardous tools, creates a risk-laden environment for human operators. The consequences of mismanagement or inadequate maintenance of robotic systems can be severe, leading to serious injuries or even fatalities. Notably, the occurrence of accidents involving industrial robots is a documented concern, particularly in situations where robotic functionalities are deeply integrated with human tasks. The integration of tangible-based programming techniques for large-scale industrial robots thus necessitates thorough examination. The close proximity between humans and robots inherent in such interactive environments amplifies the potential for injury, and the enforcement of strict safety regulations (Villani et al. 2018) imposes a significant financial cost. Furthermore, the effectiveness of kinesthetic teaching is limited by the user's knowledge and ability to determine and reproduce the ideal trajectory for a given task. This gap in knowledge or execution can introduce operational inefficiencies and elevate the risk of safety incidents if the robot's performance deviates from the intended outcome. Consequently, the adoption of tangible-based kinesthetic programming in industrial settings must be approached with caution to mitigate risks. The use of joystick-based control methods necessitates users to have a strong familiarity with directional movements, which limits the generation of complex pathways and makes it challenging for users to produce intricate trajectories. On the other hand, screen-based techniques utilize 2D visual interfaces for interaction and control, allowing touch screen, keyboard, or mouse input. However, the movement of robotic arms occurs in three-dimensional space, and 2D screens are unable to accurately represent spatial trajectories. Additionally, both joystick-based and screen-based methods require users to memorize complex commands from external platforms, lacking a direct connection to the robot and increasing cognitive demands. While these methods offer a higher level of safety compared with tangible-based methods, they may lack sufficient intuitiveness.

The integration of augmented reality (AR) into robotics has the potential to significantly enhance human–robot interaction by offering an immersive and intuitive experience. AR technology enables multi-modal interactions, presenting vital information about the robot's status, environment, and tasks, thereby facilitating human understanding and control of robotic actions. Employing AR in robotics can improve programming and control while also fostering greater understanding, interpretation, and communication, as highlighted in recent surveys and studies (Mukherjee et al. 2022; Yin et al. 2023). AR interfaces can convey a range of information types, such as the robot's internal state, environmental context, planned activities, and additional supportive content. This diverse information delivery is instrumental in aiding users' comprehension and interpretation of robot behavior (Suzuki et al. 2022). Despite these advancements, challenges persist, such as creating intuitive and user-friendly interfaces for non-experts and integrating AR technology with existing manufacturing systems and workflows.

Similarly, the digital twin (DT) concept has emerged as a transformative technology with the potential to redefine smart manufacturing and Industry 5.0. DT's defining characteristic is its ability to merge physical and digital worlds, creating a real-time virtual counterpart of physical entities, processes, or systems. For example, in manufacturing, DT models simulate production line behaviors, offering engineers valuable insights for performance optimization and preemptive issue resolution (Tao et al. 2019). The energy sector uses DT models to anticipate and mitigate power outages by simulating grid behaviors (Sifat et al. 2022), and in healthcare, DT models replicate patient physiology, aiding in more accurate diagnosis and treatment planning (Okegbile et al. 2022). Despite the widespread application of DT in various domains, the confluence of DT and AR technologies in the HRI domain remains in the early stages of development and application.

The slow adoption of DT and AR in HRI can be attributed to several factors. The complexity of these technologies and the need for interdisciplinary expertise for their successful integration have been significant barriers. Additionally, industrial inertia, cost concerns, and the required evidence for the long-term benefit of these technologies have slowed their implementation in manufacturing. Moreover, the development of accurate and real-time DTs for effective AR integration, ensuring system safety and reliability, and understanding human factors in technology interaction are challenges that researchers continue to address.

In our work, we propose an innovative AR-based DT system designed to enhance the quality and experience of HRI within the manufacturing sector. Our system is structured around three levels of DT functionality: the virtual twin for real-time monitoring, the hybrid twin for intuitive interaction, and the cognitive twin for optimized operation. By integrating AR technology with DT, we create a virtual representation of a robot arm that provides real-time data and insights into its performance, leading to more efficient and adaptable manufacturing operations. The implementation of this AR-based DT interface showcases significant improvements in operational efficiency and flexibility. Furthermore, our research offers valuable insights into the challenges and opportunities of merging AR and DT technologies, contributing to the broader discourse on technological integration in the era of Industry 4.0.

The paper is organized as follows: Sect. 2 provided the literature review of AR and DT technology and their industrial applications. The system design and implementation are introduced in Sect. 3. The demonstration for three levels of DT is elaborated in Sect. 4. The human-centric user study is presented in Sect. 5. Finally, some concluding remarks are provided in Sect. 6.

2 Literature review

The use of AR technology in HRI has become increasingly prevalent in various applications (Baroroh et al. 2021). Suzuki et al. (2022) classified existing AR applications for robots into 12 high-level clusters based on their types of applications. These clusters include data physicalization, robots for workspaces, search and rescue, mobility and transportation, telepresence and remote collaboration, medical and health, design and creative tasks, social interaction, education and training, entertainment, industry applications, and domestic and everyday use. The largest category of applications is industry, which includes manufacturing, assembly, maintenance, and factory automation. In many of these cases, AR can help reduce the workload of assembly or maintenance or program robots for automation.

The concept of DT was first introduced by Michael Grieves in 2002 as a model for product lifecycle management (Grieves and Vickers 2017). The application of DT as a data-driven and multi-physics model-based simulation and prediction tool for predictive maintenance of aircraft was first introduced by NASA and USAF (Tuegel et al. 2011). The original concept of DT, as proposed by Grieves, consisted of three fundamental elements: a physical entity, a virtual model, and their interconnection. The International Organization for Standardization (ISO) also emphasizes the importance of the synchronization property of the observable manufacturing element (OME) and its digital representation (ISO 2020). As a cutting-edge technology in modern industry, DT is gaining widespread recognition from both academia and industry, as it facilitates effective monitoring, maintenance, and optimization of industrial systems, thereby enhancing their productivity and efficiency (Liu et al. 2022c).

Augmented reality (AR) and digital twins (DT) are sophisticated technological paradigms that serve as conduits between the digital and physical realms, representing a salient trend in the evolution of future industrial landscapes. AR and DT exhibit a symbiotic relationship, wherein AR enhances DT through the facilitation of information visualization and interactive modalities. The convergence of AR and DT has engendered a plethora of applications within the domain of human–robot interaction (HRI). For example, Li et al. (2023) have conceptualized a DT-based HRI system that integrates visual augmentation for workers, robot velocity modulation, preemptive motion visualization, collision detection, and a deep reinforcement learning algorithm for robot collision avoidance, all orchestrated within an AR-augmented framework. Paripooranan et al. (2020) proposed an AR-enabled DT for 3D Printing. Kuts et al. (2018) used AR to integrate decision trees with machine learning and computer vision for self-learning robot programming, including simulating and testing robot manipulation with the DT concept. Cai et al. (2020) have executed a layout design method for a multi-robot additive manufacturing system, employing AR-based DT to aid operators in the spatial arrangement of robots. Müller et al. (2021) built the DT system for robot programming by demonstration via AR interaction. Amtsberg et al. (2021) built the DT of robots and work objects, and utilized AR to facilitate interaction methods that enhance communication and collaboration between users and robots, ultimately leading to more effective task sharing. To elucidate the interplay between AR and DT, Yin et al. (2023) have categorized AR-assisted DT systems into three distinct levels based on their functional characteristics and data flow dynamics:

- *Virtual Twin*, involves the transmitting of physical to virtual data, sensor data-based functions of monitoring and alerting, and non-registered visualization. For instance, it can visualize a robot's state and behavior in real-time, aiding in monitoring, diagnosing performance, and identifying potential issues before they become critical. Current research predominantly utilizes virtual twins to facilitate such geometry and data visualization, simulation, and inspections in manufacturing settings.
- *Hybrid Twin*, on the other hand, focuses on the analysis and feedback of virtual to physical systems, which encompasses various components such as multi-modal interaction and control, visual registration, context information-related analysis, and the functions that are realized based on them. For instance, in the context of HRI assembly scenarios, the AR-based DT utilizes scene understanding capabilities and object detection to analyze workspace information collected by the HoloLens sensing system. With an understanding of the assembly process, the virtual assembly step can be visualized beforehand through AR (Liu et al. 2022a). Additionally, the cyber-physical interaction includes collision detection between the physical entity and the virtual model, which can be applied in HRI (Li et al. 2023). This integration facilitates an intuitive and interactive HRI, paving the way for more dynamic and responsive control processes.
- *Cognitive Twin*, regarded as the high-level cognitive DT, involving both human and machine intelligence. A cognitive twin can be instructed by human operators and study how to perform better and fulfill human needs. For instance, in human–robot collaboration tasks (Dimitropoulos et al. 2021), the users can correct the robot poses through gesture interaction with the robot's DT, which can lead to improved human ergonomics. In the prefabrication process for timber, AR provides suitable interaction methods that enable users and cobots to better understand each other and ultimately achieve harmonious task sharing (Amtsberg et al. 2021). Cognitive twins possess the distinct advantage of effectively addressing complex and unpredictable situations by harnessing the power of human intelligence. This capability allows DT system to dynamically adapt and respond to intricate scenarios, surpassing the limitations of traditional DT, which can be applied in operation optimization, and system evolution in HRI scenarios.

In current research on AR-based HRI within the manufacturing sector, the majority of interaction methods are focused on the first level, utilizing the virtual twin to visualize and inspect machines (Müller et al. 2021; Choi and Cai 2014). Some studies examine the hybrid twin, employing the virtual domain to control the machines (Ostanin et al. 2020), while only a few investigate the cognitive twin (Liu et al. 2022b). Moreover, even within level 3, the emphasis is often on machine intelligence, with human intelligence being overlooked (Yin et al. 2023). In the context of the future human-centric industrial paradigm, the well-being of individual workers remains a primary concern. This integration of AR and DT aligns with the human-centric concepts wave and benefits Industry 5.0 (Breque et al. 2021), human-centered intelligent manufacturing (Baicun et al. 2020), and human cyber-physical system (HCPS) (Wang et al. 2022). HRI has surfaced as a promising and challenging research area in the pursuit of human-centric manufacturing systems (Huang et al. 2022). Further exploration in this field may include developing advanced and intuitive systems that allow human operators to engage with DT models of robots. These systems could potentially enhance the efficiency of monitoring, maintenance, and optimization of robotic systems while simultaneously improving users' comprehension of the robot's behavior and performance.

3 System design and implementation

The overall structure of the 3-level DT construction for AR-based HRI is depicted in Fig. 1. In this proposed system, the robot functions as the physical platform that executes the desired tasks, while the HoloLens device serves as the interface between the operator and the robot. The human operator is in charge of controlling and overseeing the robot's actions and interactions.

The 3-level DT is organized as progressive layers. The most fundamental level is the virtual twin, where real joint states, trajectories, collisions, sensor data, and other physical messages can be transferred to the virtual counterparts. This allows for real-time monitoring, visualization, and diagnosis. Beyond the single-directional data transmission from the physical to the virtual, the hybrid twin enables two-way communication between the physical asset and its corresponding virtual counterpart. Consequently, goal positions, limitations, desired joint angles, and some system commands can be sent to the physical counterpart. At level 3, human involvement is introduced, and human intelligence is utilized to control and optimize the robot's operation, such as collision avoidance and trajectory optimization through the AR interface. The study expands upon the theoretical framework outlined in reference (Yin et al. 2023) by providing a detailed and contextualized application, thereby offering a tangible demonstration of the functionality and interactions across all three levels of AR-based Digital Twins. Through this practical example, we illustrate the seamless integration of three levels and highlight the progressive relationship that extends from the virtual to the cognitive level, emphasizing how each subsequent level builds upon the foundation laid by its predecessors.

The framework comprises three primary components: communication between the robot and HoloLens 2, intuitive interface design, and implementation of essential algorithms such as spatial anchor, object recognition and tracking, and robot kinematic model.

3.1 Intuitive interface design

The intuitive interface is a crucial aspect of HRI, which can greatly impact the efficiency and ease of use of the system. An ideal interface for HRI should have three essential functions: visualization, simulation, and control. Visualization provides real-time feedback on the robot's position, orientation, and movements. The simulation module should accurately represent the robot and its environment, allowing the user to visualize and interact with the robot in a virtual environment, accommodating different scenarios and variations. This allows for experimentation and optimization of the robot's behavior before implementing it in the real world. Control enables the user to efficiently and effectively manipulate the robot's movement and behavior to accomplish various tasks and operations, such as adjusting position, orientation, and motion trajectory and

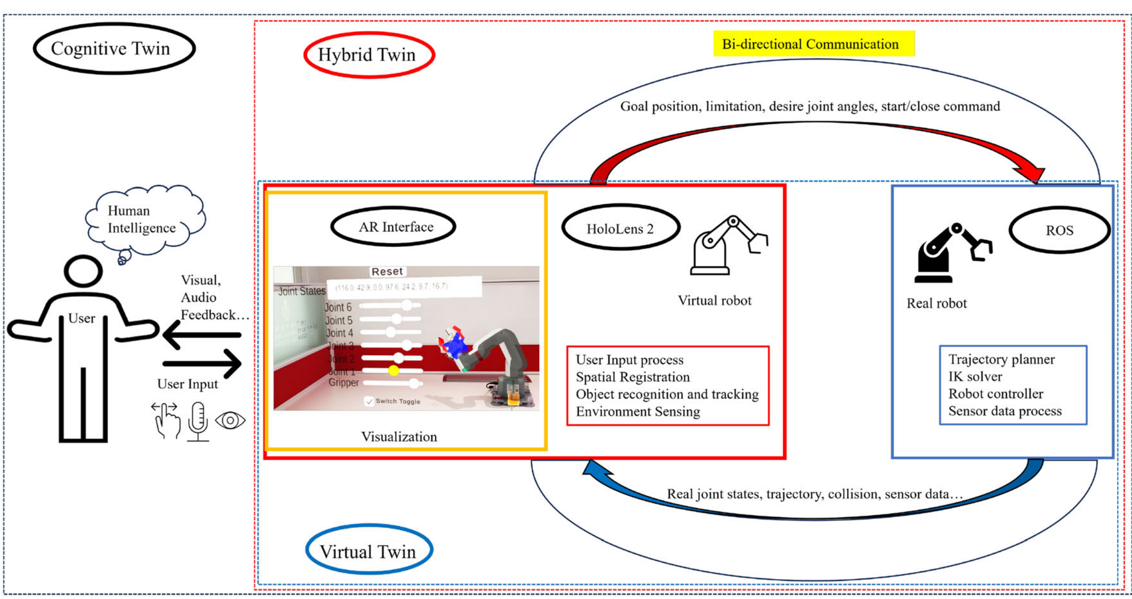


Fig. 1 The framework of DT for AR-based HRI

controlling speed and force. Incorporating these three functions enhances the user's experience, and increases the efficiency and effectiveness of the overall system (Fig. 2).

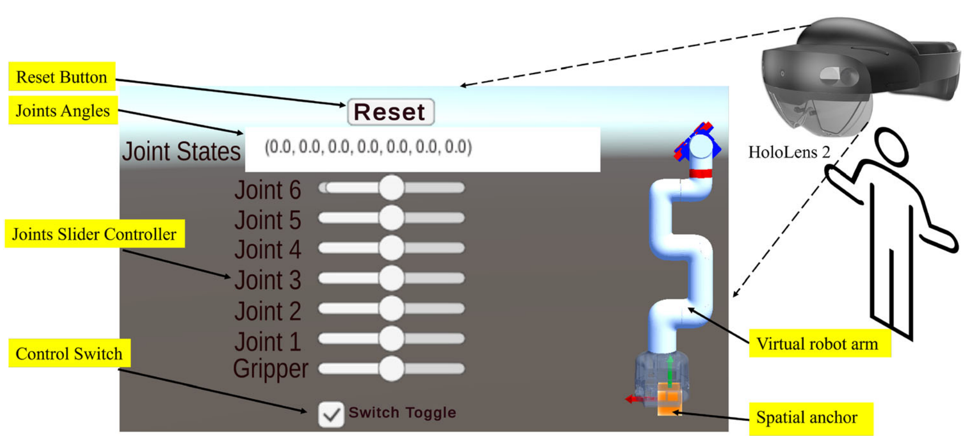
Hololens 2 is a mixed-reality headset developed by Microsoft, designed to provide an immersive experience for the user by blending the virtual and real world. The Hololens 2 is equipped with advanced sensors and cameras, including depth-sensing cameras and eye-tracking sensors, which enable the device to recognize and track the user's head and hand movements in real-time. This technology allows the user to interact with virtual objects and environments in a natural and intuitive way, making it an ideal platform for HRI. Unity, a popular game development engine, is utilized to create the AR environment and user interface for controlling the robot arm. Unity provides a range of tools and features to simplify the development process and enable the creation of immersive and interactive applications. The author used the Mixed Reality Toolkit (MRTK) for Unity, which is an open-source toolkit designed to simplify the development of mixed-reality applications. The MRTK provides a set of pre-built components and tools that can be used to create immersive experiences using the Hololens 2 headset. The interface design based on Hololens 2 incorporates several components that enable the user to control the robot arm using AR technology. These components include:

- Reset button: allows the user to reset the robot to its initial position.
- Joint angles: the joint angles of the robot, along with the gripper status, are presented and can be modified through the use of either the joint slider control or gesture control with the DT robot arm. The parameters are arranged from left to right and correspond to the 6th, 5th, 4th, 3rd, 2nd, and 1st joints, as well as the gripper status. The joint range is $[-180°, 180°]$, except for the 2nd joint, which has a range of $[0°, 180°]$ to prevent collisions with the ground.
- Joints slider controller: enables the user to adjust the angles of each joint of the robot using a slider.
- Control toggle: enables intuitive gesture interaction with the DT model.
- Spatial anchor: anchors the robot to a real location within the real world, allows the user to alter its position, resulting in a corresponding change in the virtual robot's location within the real world.
- Virtual robot arm: provides a visual representation of the robot arm in the AR environment. The users can control it using gestures. The virtual robot DT model can be overlaid with its physical counterparts to provide a more intuitive and accurate representation.

In order to facilitate intuitive and efficient control of the robot arm, the proposed interface incorporates gesture recognition technology using Hololens 2. This new interaction method allows users to directly manipulate the robot arm with intuitive gestures, such as pinching fingers or making a fist. The user can also use other gestures or modalities, such as voice and gaze based on preference or behavior.

The interface's operation process includes far manipulation and near manipulation, see Fig. 3. In far manipulation, the user uses their hand to select the robot joint and then pinches their finger, causing the hand ray to become solid and indicating that the joint has been selected. The user can then move their hand left or right to move the corresponding robot joint. In near manipulation, the user needs to get close to the virtual robot arm, touch the desired robot joint with their hand, and then pinch their finger to select the joint before moving their hand left or right to manipulate it. In contrast to far manipulation, near control involves an additional implicit modality, namely proximity, wherein the user approaches the object, leading to the suspension of far manipulation and activation of near control. Consequently, near manipulation is a multi-modal interaction that encompasses both gesture and proximity.

Fig. 2 The AR intuitive interface based on HoloLens 2



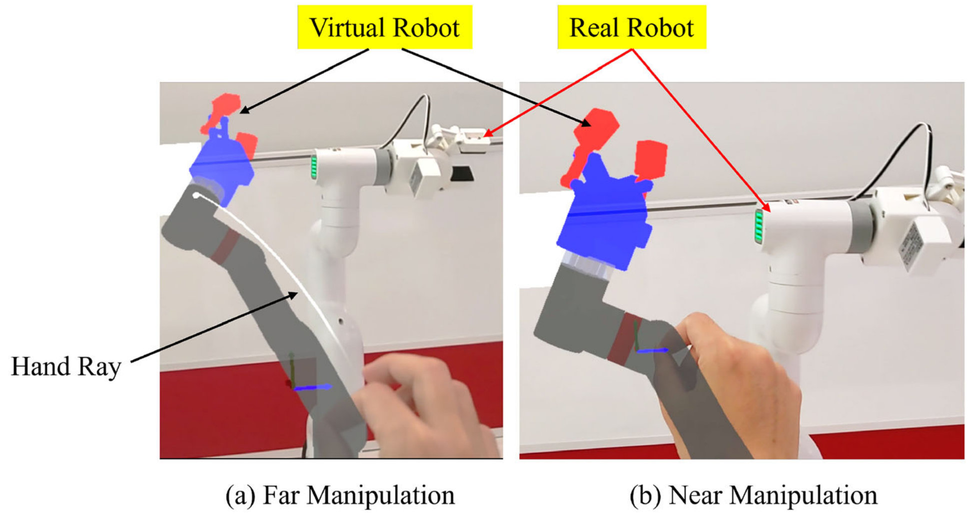


Fig. 3 Two interaction ways: **a** far manipulation and **b** near manipulation

The control logic for intuitive HRI entails several critical components. Firstly, the DT system detects the initiation of manipulation, whereby the user's finger pinch triggers a transition in the select line from a dotted to a solid form. This signifies the commencement of the manipulation process. Subsequently, the HoloLens 2 begins to meticulously track the movement of the user's finger, as explicated in Sect. 3.3.2. The system then proceeds to translate the detected movement of the user's thumb to the corresponding alteration in the robot joint's angle. Notably, once the user has concluded the intended operation, such as releasing fingers, the system ceases tracking the finger's movement, thereby preventing further alterations from being transmitted to the robot. This design enables users to focus on the robot arm and maneuver their hand to effectuate changes in the joint angle with precision and ease, reducing the mental workload required for HRI.

Another method of control available in the interface is slider control. The user can move the slider on the left side to control the specified joint's movement. However, as the control slider is not directly related to the robot arm and lacks intuitive feedback, it may not be as effective for HRI as gesture recognition technology. It is noted that the sliders in the AR interface are mainly used for visualization and debugging.

3.2 Communication

How to establish the communication between the HoloLens 2 and the robot arm is critical for the DT. A viable approach for achieving this objective is to utilize the Robot Operating System (ROS) middleware system. ROS is a popular open-source framework that provides an infrastructure for controlling both software and hardware components of robotic systems. It also includes a messaging system that facilitates communication among the different elements of a robotic system. In addition, the Rosbridge package provides a JSON API to ROS functionality for non-ROS programs. There are a variety of front ends that interface with Rosbridge, in this work we adopt the WebSocket server for web browsers to interact with. In the present study, a Wi-Fi connection is employed as the means of establishing a connection, while the Rosbridge is utilized to establish a communication protocol between the HoloLens and the robot arm, thus facilitating real-time monitoring and control of the robot's performance during operational processes.

3.3 Fundamental algorithm

In this section, the essential algorithms, including spatial anchor, object recognition and tracking in AR, and robotics forward and inverse kinematics, are introduced.

3.3.1 Spatial anchor for AR

The integration of AR with a physical robot arm poses a challenge in aligning virtual objects with the tangible environment. To address this, a spatial anchor is utilized to establish a virtual point in a three-dimensional space that is affixed to a tangible object or location. This anchor preserves the orientation and location of AR content in the tangible world, even when the camera is in motion. Unlike visual markers that trigger AR experiences, spatial anchors serve as a reference point for all other objects in the AR environment. The illustration of the spatial anchor is shown in Fig. 4.

In order to determine the position of a virtual object in the global coordinate system, it is necessary to know its position and orientation relative to the spatial anchor in the local coordinate system. In this study, the position (P_l) and orientation (R_l) of the virtual object with respect to the

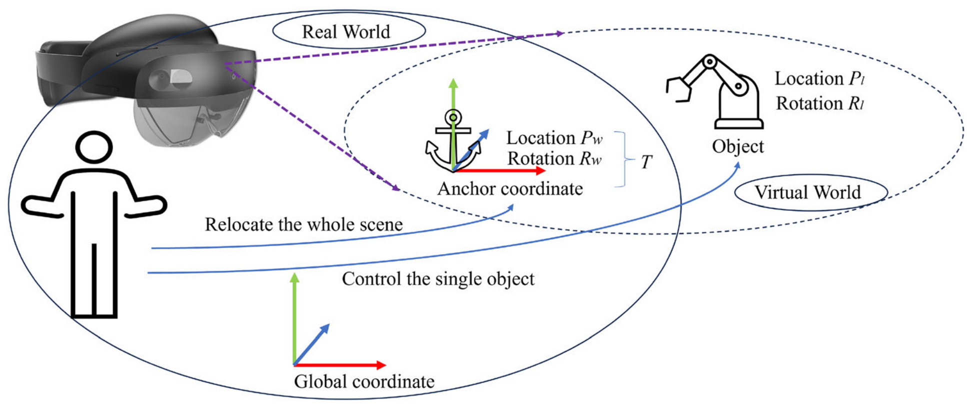


Fig. 4 Illustration of the transformation between the virtual world and the real world using spatial anchor

spatial anchor are predetermined in Unity scenes and can be adjusted through human interaction. The 3×3 rotation matrix R_l represents the orientation of the virtual object relative to the spatial anchor and can be computed using Euler angles or quaternions. Euler angles define the virtual object's orientation through three angles of rotation around the x, y, and z axes, while quaternions provide a more concise representation of orientation using a four-dimensional vector.

A homogeneous coordinate in a 4×4 metric (G_l) can be utilized to represent the local gesture, including the position and orientation with respect to the spatial anchor:

$$G_l = \begin{bmatrix} R_l & P_l \\ O(1 \times 3) & 1 \end{bmatrix} \tag{1}$$

The transformation process that maps the local G_l to the global G_w can be expressed as:

$$G_w = T \cdot G_l \tag{2}$$

where T is a 4×4 homogeneous transformation matrix:

$$T = \begin{bmatrix} R_w & P_w \\ O(1 \times 3) & 1 \end{bmatrix} \tag{3}$$

here R_w is a 3×3 matrix that characterizes the orientation, and P_w is a 3×1 vector that denotes the position of the spatial anchor with respect to the global coordinate system. It should be noted that the global coordinate system is located in the real world, while the local coordinate system is in the virtual scene. Through Eq. 2, the virtual world is aligned with the real world, allowing users to modify the location of the entire virtual scene by adjusting the anchor's position through AR interactions.

3.3.2 Object recognition and tracking

The Hololens 2 is a potent tool for object recognition and tracking that utilizes advanced computer vision algorithms and machine learning techniques to identify and track objects in the real world. One commonly used algorithm for object recognition and tracking is the convolutional neural network (CNN). CNNs are a type of deep learning algorithm that can be trained to classify images based on their features. The Hololens 2 can use a pre-trained CNN to quickly classify objects based on their features, allowing it to quickly identify them in future images.

In this work, the Hololens 2 was used for finger recognition and tracking. This is achieved using a combination of depth sensors and machine learning algorithms to detect and track the position and movement of the user's fingers. One common algorithm for finger recognition and tracking is the Hand Keypoint Detection algorithm, which is based on a CNN architecture. The algorithm works by first detecting the hand in the image, then using a CNN to detect the key points of the hand, such as the tips of the fingers and the base of the palm.

The algorithm for Hand Keypoint Detection can be mathematically articulated through the following set of equations:

$$h_{i,j} = \sigma(W_{i,j}x + b_j) \tag{4}$$

$$y = \mathrm{softmax}(Uh + c) \tag{5}$$

Herein, x denotes the input image, W and U represent the respective weight matrices, b and c are the corresponding bias vectors, σ refers to the activation function (such as the Rectified Linear Unit, ReLU), and softmax is the function employed for generating the output. The culmination of this algorithm is a constellation of keypoints that serve to delineate the position and trajectories of the user's digits.

In the context of this work, the thumb's displacement is leveraged as a proxy for modifying the angular disposition of a joint:

$$\Delta\theta = K_f \Delta y \tag{6}$$

where $\Delta\theta$ is the alteration of the robot's joint angle, Δy is the thumb's movement, and K_f epitomizes the linear factor that correlates the motion of the thumb to the angular variation of the joint. During actual manipulation, the user designates a specific robot joint and moves the finger either to the left or the right; consequently, the thumb's displacement is transduced into a corresponding alteration in the robot's joint angle.

3.3.3 Robot kinematic model

The robot kinematic model is a mathematical representation of a robot's physical structure and movement capabilities. It provides an understanding of the relationship between the robot's joints and links and how they move relative to each other. The Denavit–Hartenberg (DH) model is a commonly used kinematic model that uses four parameters to describe the relationship between adjacent links in a robot arm, where the DH model for the robot arm used in this work is shown in "Appendix A". The DH model can be used to calculate the position and orientation of the end effector of a robot arm given the joint angles.

The forward kinematics equation for a robot arm using the DH model is expressed as:

$$T_n = T_1 \cdot T_2 \cdot \cdots \cdot T_{n-1} \tag{7}$$

where T_n is the homogeneous transformation matrix that describes the position and orientation of the end effector, and T_1 to T_{n-1} are the homogeneous transformation matrices that describe the position and orientation of each link in the robot arm. The DH parameters including the link length a_i, the link twist α_i, the link offset d_i, and the joint angle θ_i, shown in Fig. 14, are used to calculate these transformation matrices as the transformation matrix T_i:

$$T_i = \begin{bmatrix} \cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

Therefore, the final gesture can be expressed as a combination of the position and orientation vectors, obtained from the translation vector and rotation matrix of the transformation matrix T_n, respectively. The position vector p can be obtained by extracting the first three elements of the fourth column of T_n. That is:

$$p = [\, T_{n,1,4}\ T_{n,2,4}\ T_{n,3,4} \,] \tag{9}$$

where $T_{n,1,4}$, $T_{n,2,4}$, and $T_{n,3,4}$ represent the x, y, and z position coordinates.

The orientation matrix R can be obtained by extracting the first three columns of the first three rows of T_n. That is:

$$R = \begin{bmatrix} T_{n,1,1} & T_{n,1,2} & T_{n,1,3} \\ T_{n,2,1} & T_{n,2,2} & T_{n,2,3} \\ T_{n,3,1} & T_{n,3,2} & T_{n,3,3} \end{bmatrix} \tag{10}$$

From the orientation matrix R, we can get the orientation angles α, β, and γ.

The inverse kinematics equation is used to calculate the joint angles required to achieve the desired position and orientation of the end effector, expressed as:

$$\theta = f^{-1}(q) \tag{11}$$

where θ is the vector of joint angles, q represents $[x, y, z, \alpha, \beta, \gamma]$, x, y, and z are the desired position coordinates of the end effector, and α, β, and γ are the desired orientation angles of the end effector.

The Jacobian matrix can be used in inverse kinematic calculation, which can be calculated using the partial derivatives of the forward kinematics equation:

$$J = \begin{bmatrix} \frac{\partial x}{\partial\theta_1} & \frac{\partial x}{\partial\theta_2} & \cdots & \frac{\partial x}{\partial\theta_n} \\ \frac{\partial y}{\partial\theta_1} & \frac{\partial y}{\partial\theta_2} & \cdots & \frac{\partial y}{\partial\theta_n} \\ \frac{\partial z}{\partial\theta_1} & \frac{\partial z}{\partial\theta_2} & \cdots & \frac{\partial z}{\partial\theta_n} \\ \frac{\partial \alpha}{\partial\theta_1} & \frac{\partial \alpha}{\partial\theta_2} & \cdots & \frac{\partial \alpha}{\partial\theta_n} \\ \frac{\partial \beta}{\partial\theta_1} & \frac{\partial \beta}{\partial\theta_2} & \cdots & \frac{\partial \beta}{\partial\theta_n} \\ \frac{\partial \gamma}{\partial\theta_1} & \frac{\partial \gamma}{\partial\theta_2} & \cdots & \frac{\partial \gamma}{\partial\theta_n} \end{bmatrix} \tag{12}$$

By inverting the Jacobian matrix, the joint angles required to achieve a desired end effector position and orientation can be calculated. This is known as the inverse kinematics equation, which is represented by the equation

$$\Delta\theta = J^{-1}\Delta q \tag{13}$$

where $\Delta\theta$ is the vector of joint angles required to achieve the desired end effector position and orientation, J^{-1} is the inverse of the Jacobian matrix, and Δq is the difference between the desired and current end effector position and orientation.

It is important to note that the Jacobian matrix may not have an inverse in all configurations of the robot arm. In these cases, alternative methods such as numerical optimization or iterative methods may be used to solve the inverse kinematics problem.

In this work, we present novel contributions to AR-based DT systems for enhanced human–robot interaction using the HoloLens 2 platform. Our research introduces unique algorithms enabling intuitive, gesture-based control interfaces, allowing users to manipulate robot joints directly through finger movements. This direct finger-to-joint interaction method, not inherent to the HoloLens 2, represents an innovative approach to bridging the gap between human intent and robotic action. Additionally, we developed the spatial marker algorithms requiring complex transformations for accurate virtual-to-physical world anchoring, and seamlessly integrated object recognition with robot kinematics to translate user gestures into precise robotic movement. These advancements are not only tailored to the specific needs of robot control within a DT environment but also enhance the user experience by providing a more natural and direct method of interaction.

4 Demonstration for three levels of digital twin

According to different functions and data flow directions, DT technology is composed of three levels: the virtual twin, the hybrid twin, and the cognitive twin (Yin et al. 2023). Figure 5 depicts the three levels of DT for AR-based HRI.

4.1 Virtual twin

The virtual twin level allows for real-time monitoring and data transmission from physical assets to their virtual counterparts. In Fig. 5, the system presents the actual status of the robot arm through a combination of values, sliders, and a virtual robot arm, as depicted in Fig. 6. When the tangible robot's status undergoes modification, the values slider and the virtual robot arm alter correspondingly. It is noted that the virtual robot can be overlaid with the real robot to provide a more accurate and immersive experience. This enables remote analysis of the asset's performance, condition, and behavior. Additionally, it facilitates the simulation of different scenarios and the prediction of outcomes, leading to proactive maintenance and optimization of the asset. The use of simulation is particularly beneficial for robot planning, as it provides essential information about collisions, trajectories, and possible outcomes before the actual operation, thereby increasing efficiency and safety.

4.2 Hybrid twin

The hybrid twin level of DT technology enables bidirectional communication between a physical asset and its corresponding virtual counterpart. In addition to compatibility at the virtual level, the hybrid level allows for control from the virtual twin to the physical twin. This means that changes made to the virtual twin can be transmitted to the physical asset, and vice versa. An intuitive AR interface is designed to enable users to interact with the DT using natural gestures, making it easier to program and control the asset. For example, in this work, the joint angle of the virtual robot arm can be changed by human gestures, such as moving hands to the left or right. The interaction can be through voice, gaze, or a combination of both in HoloLen 2. Upon activation, the physical robot will synchronize with the virtual twin's status, thereby facilitating the transmission of messages from the virtual counterpart to the tangible asset. Despite the presence of bidirectional communication, effectively controlling the robot in unforeseen circumstances, such as collisions with physical objects, delays in robot motion, and operation in intricate environments, remains challenging. Therefore, it is imperative to integrate human intelligence into the control loop.

4.3 Cognitive twin

The cognitive level harnesses both machine and human intelligence to amplify the functionality of assets. Machine intelligence has proven to be particularly effective in object recognition and tracking fields, thereby enabling more precise and intuitive user interactions. Utilizing machine learning algorithms, the system can accurately identify and

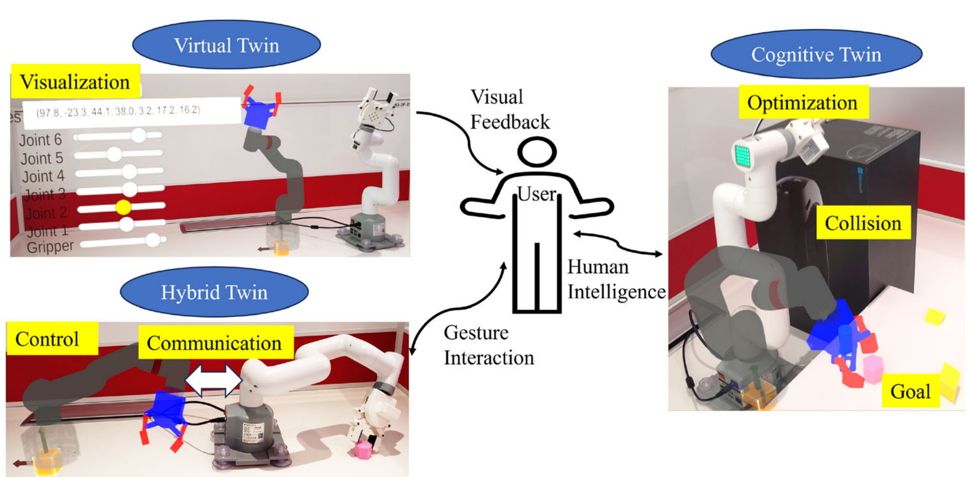
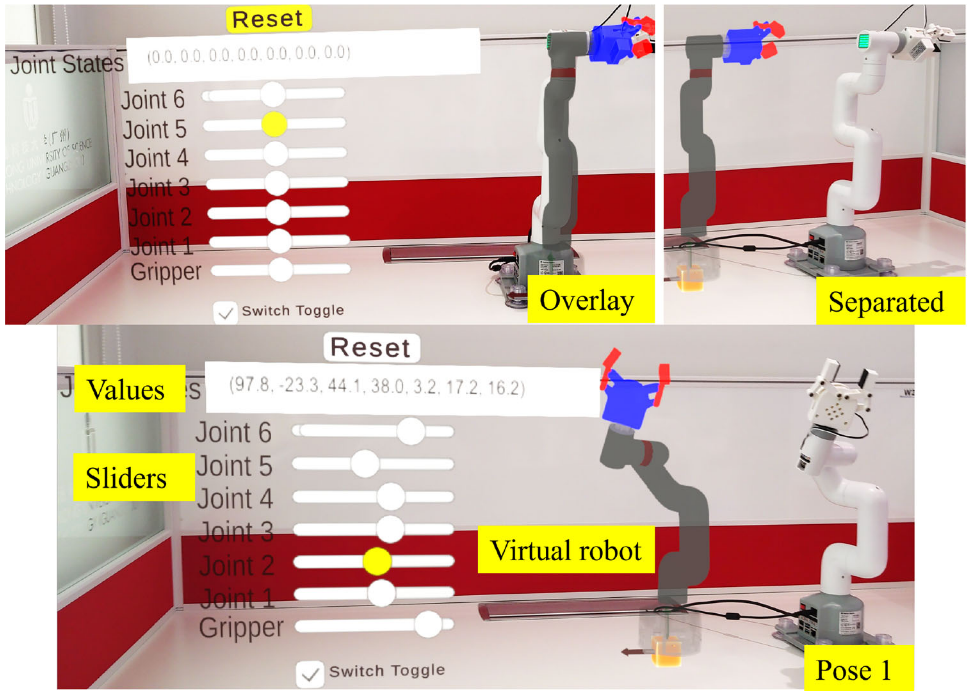


Fig. 5 Illustration for three levels of DT (virtual twin, hybrid twin, and cognitive twin) of the HRI, with the corresponding functionalities for real-time monitoring, intuitive interaction and control, and optimized operation through human intelligence

Fig. 6 Visualization function of DT system for the robot arm. The virtual arm can be overlaid or separated from the real robot. As the real robot undergoes changes in its status, the virtual representation dynamically reflects these modifications through the adjustment of values, sliders, and the virtual robot itself



track objects, consequently allowing users to better navigate and manipulate their assets. Furthermore, the amalgamation of human and machine intelligence contributes to enhancing the overall functionality of the asset, as it fosters a more comprehensive and dynamic approach to problem-solving and decision-making. The integration of these cognitive technologies possesses the potential to transform asset optimization and elevate the overall user experience.

As the demand for adaptable and personalized smart manufacturing intensifies, human intelligence continues to play a vital role in the industry. Specifically, AR interfaces support the conveyance of human creativity and ideas to manufacturing systems, particularly in scenarios that require on-demand fabrication or smart agent control. In this research, human users contribute input and guidance to the DT, thereby optimizing its performance and capabilities. For example, during the pretest simulation, users can immerse themselves in the running condition. If an issue arises, such as collision, misplacement, or confused targets, as depicted in Fig. 7, users can employ their intelligence to address the problem and make informed decisions. Corresponding actions may encompass relocating the colliding item, adjusting the trajectory, and recognizing the true target, among other interventions. At the cognitive twin level, it demonstrates the capability to effectively navigate complex and unpredictable scenarios by leveraging human intelligence.

In summary, DT system offers a powerful approach to monitoring, simulating, and optimizing physical assets. Each level offers unique features and capabilities, enabling different degrees of interaction and optimization between the physical asset and its virtual twin.

5 User study

To gauge the effectiveness of the proposed AR-based DT system for a robotic arm, we designed and conducted a user study. The outcomes demonstrated that the proposed system considerably decreased the operation time and reduced errors compared to the conventional one. Furthermore, participants reported increased satisfaction and perceived ease of use with our system. The study involved a diverse group of 7 participants with varying levels of experience in the manufacturing sector, primarily consisting of engineering students aged between 20 and 28 years old. Three participants had prior familiarity with robotic arm operations, but none had any previous experience with AR or HoloLens 2 technologies.

In modern industry, the pick-and-place task is one of the most widespread scenarios, typically carried out by either automated machines or human workers. Due to the superior flexibility of human intelligence and physical abilities compared to machines, human workers are better equipped to adapt to changes in the manufacturing line. In contrast, machines necessitate human intervention for reprogramming, reconfiguration, or relocation when alterations are made to the factory layout. In a human-centric manufacturing environment, humans play a central role in the operation of the smart reconfigurable manufacturing

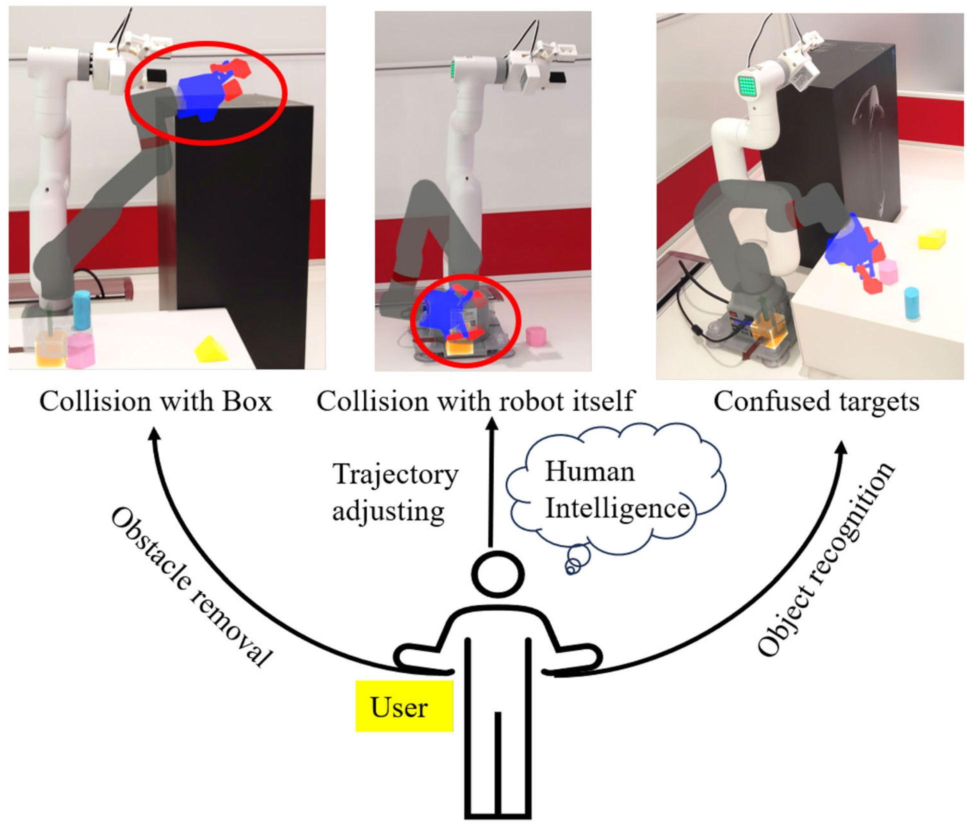


Fig. 7 In cognitive realms, DT scenarios pertain to the utilization of human intelligence to facilitate problem-solving for robots, particularly in instances involving collisions with the environment or the robot itself, as well as in situations where target identification becomes ambiguous

system. Therefore, it is imperative to enhance the performance of HRI in the pick-and-place scenario to drive the progress of smart manufacturing. This user study focused on a straightforward pick-and-place task, in which participants were required to pick up a cube from one location and transfer it to a designated goal position.

Conventional methods of HRI often involve the use of joystick-based, screen-based, and tangibility-based kinesthetic teaching techniques. However, industrial robots can be bulky and hazardous in practical situations, thereby posing a risk to workers employing tangibility-based approaches. Considering the similarities in functionality between joystick-based and screen-based interactions, this study chose to utilize screen-based interaction as the control group. To support this task, an augmented reality (AR)-based DT system was developed for the robotic arm. The study was equipped with all necessary components for both AR and conventional interactions, including HoloLens 2, a display screen, a mouse, and a computer. For safety reasons, a desktop robot arm, Elephant robot myCobot 280-Pi, was employed in this research.

5.1 Procedure

The procedure section describes the step-by-step process of the user study, comprising a training session, a controlled experiment, and a post-experiment survey. The framework of the user study is shown in Fig. 8.

5.1.1 Training session

The participants underwent a brief yet comprehensive training session on the usage of both the AR-based DT system and the conventional 2D screen control. The training session for the AR-based DT system included instructions on the AR interface, gesture controls, and visualization features of the DT. Conversely, the training session for the conventional 2D screen control encompassed instructions on the 2D screen GUI, mouse control, and keyboard input for the joint angle. Each participant will undergo a 30-minute training session in order to familiarize themselves with the operation of the robot in both HRI modes. This training session allows users to practice the task multiple times in both interaction modes, thereby ensuring that they are proficient in operating the robot. Once participants demonstrate proficiency in both interaction modes, the experiment will commence.

5.1.2 Controlled experiment

In this rigorously controlled experiment, the participants were assigned to complete two distinct sets of tasks. The

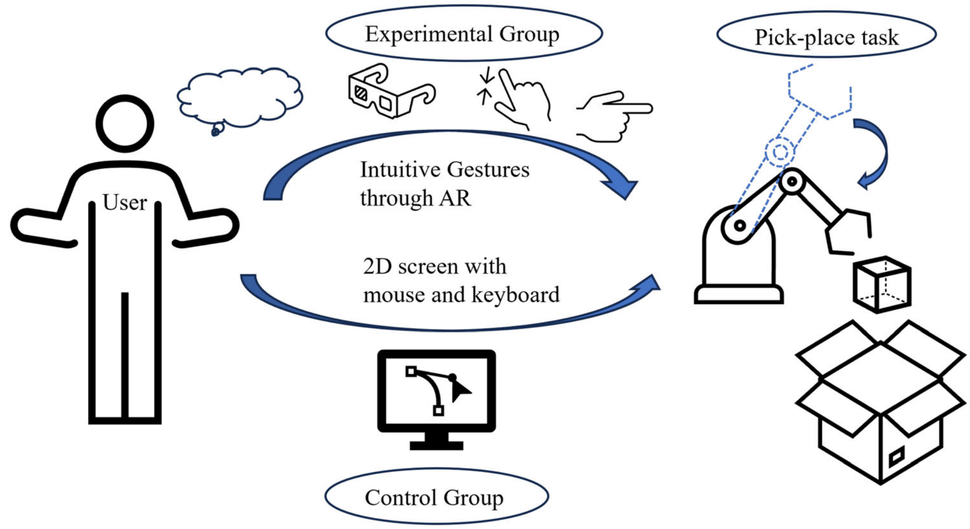


Fig. 8 The framework of the user study. The user needs to finish the pick-place task through the AR and 2D system

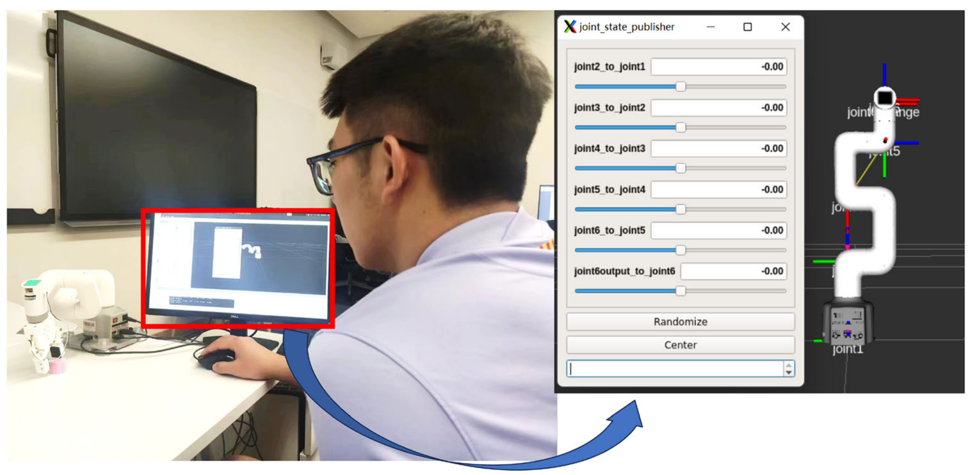


Fig. 9 The setup of control group using 2D screen interaction. Left: The user adjusted the joint angle using a mouse or keyboard while facing the 2D screen. Right: The GUI of the robot control was shown on 2D screen

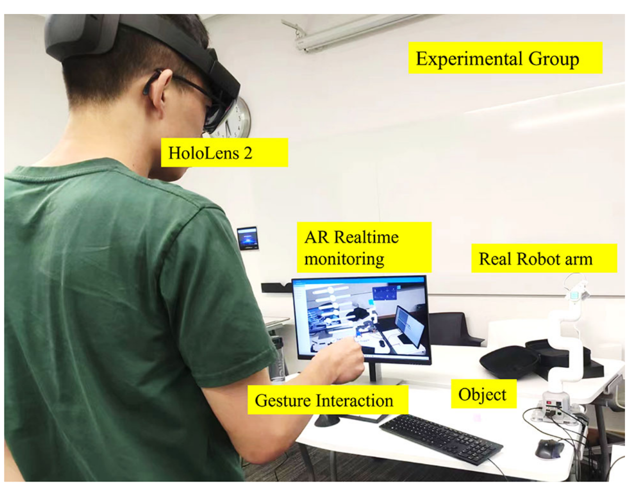
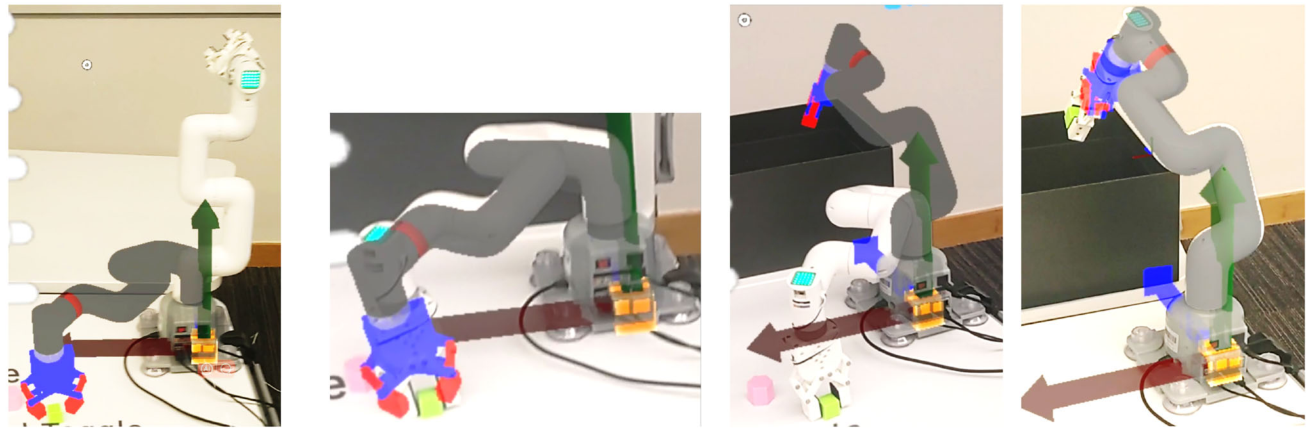


Fig. 10 The setup of the experimental group. The participant wore the AR glasses and utilized natural gestures to control the virtual robot. The real-time monitoring of AR environment was presented on screen to facilitate the experimenters to observe the users' actions and offer guidance

first group, known as the control group, employed a conventional 2D screen with a mouse and keyboard to perform the tasks. In this group, the GUI is shown as Fig. 9. The GUI contains the visualization of the robot status, real-time represents the gesture of a real robot; an interface of the sliders, which the participants used to control the joint angles.

The second group, denoted as the experimental group, employed an AR-based DT system, as illustrated in Fig. 10. The participants utilized natural gestures, specifically finger pinch to confirm and release to deselect, to control the virtual robot. To enable experimenters to comprehend the participants' actions, a screen was utilized for real-time monitoring of the AR environment. Experimenters could observe the users' actions and offer guidance to the participants to manipulate the robot to the appropriate position.

The experimental group using the AR-based DT system followed the procedure depicted in Fig. 11. Participants in



(a) Simulation (b) Execution (c) Simulation (d) Execution

Fig. 11 The procedure of the experimental group using DT: **a** First the user moved the virtual robot arm to the object and adjusted the position and orientation, subsequently closing the virtual gripper to grasp the object; **b** then activated the real robot to mirror the same pose as the virtual robot; **c** Upon successfully gripping the object, the virtual robot arm was moved to the desired location, and the virtual gripper was opened: **d** the real robot was activated to replicate the same pose

this group could manipulate the virtual robot arm as a simulation before operating the real robot. Once they identified the appropriate position, they activated the real robot to reach the same position as the virtual robot. Key performance metrics, including the time taken and the number of errors, were meticulously recorded for each task. The task involved pick-and-place assignments, which are widely prevalent in the manufacturing industry and require users to regulate the joint angle of a robotic arm to lift a cube and transport it to specific target positions. The experimental group utilized an AR-based intuitive control, while the control group relied on a 2D screen with sliders. Throughout the experiment, objective factors such as the time required to complete the task and the number of user errors were recorded. The collision number, which included collisions with the ground and the robot itself during the operation of the robotic arm, was used to measure the number of errors.

5.1.3 Post-experiment survey

After the completion of the tasks, a survey was conducted to gather feedback from participants regarding their experience using the AR-based DT system. The purpose of this survey was to obtain insights into user satisfaction, perceived ease of use, and potential areas for improvement.

The post-experiment survey is referred to several standardized questionnaires, including the System Usability Scale (Brooke 1996), AttrakDiff (Hassenzahl et al. 2003), and NASA TLX (Hart et al. 1988). The questionnaire utilized in this study is presented in “Appendix B”. The questionnaire consisted of three sections. Section 1 focused on the user experience with regard to accessibility, as well as the mental and physical demands of the system. Section 2, on the other hand, focused on the system’s performance, including its accuracy, efficiency, and user willingness. In this context, accuracy refers to the precision of manipulation, specifically the user’s ability to utilize the interface to accurately grasp an object and place it in the intended position. Both sections utilized a 7-Likert scale, with users being asked to score each question from 1 to 7, with 1 being the worst and 7 being the best. Section 3 explored the system’s imperfect aspects, difficulties in using it, and suggestions for future development.

5.2 Data analysis and discussion

This section presents a comprehensive analysis and discussion of the controlled experiment’s data. The performance of the control group and the experimental group was evaluated using statistical methods, specifically single-factor analysis, which is employed to compare differences between the two groups.

5.2.1 Objective metrics

Figure 12 illustrates the recorded data during the experiment, comparing the number of errors and the time taken for successful completion in both AR and 2D screen interaction. The respective p values between AR and 2D screen interaction are 0.0026 and 0.0556. The p value for the number of errors is significantly smaller than 0.05, indicating a substantial difference between the two groups. However, the difference in successful completion time is not as significant, as its p value is slightly larger than 0.05. It is important to note that the average number of errors and

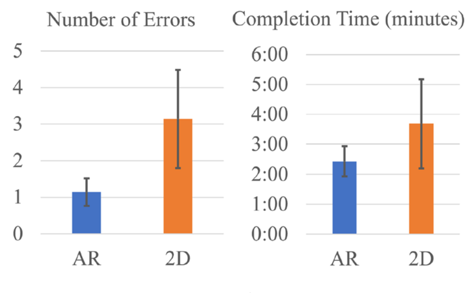


Fig. 12 The subjective data in the user study (Error-bars show ±1 SD). The number of errors denotes the collision number, including collisions with the ground and the robot itself during the operation of the robotic arm. Completion time is the time required to complete the task

successful completion time for the AR method are considerably smaller than those of the conventional 2D screen interaction.

In terms of the number of errors, most participants experienced only one collision with the table, as the AR system allows for simulation before actual execution. Participants can maneuver the virtual robot arm to the target and adjust its position around the object. After finding the ideal position, they can send a command to the real robot arm, which then replicates the virtual arm's position. Consequently, the majority of participants encountered only one collision with the object. In contrast, conventional 2D screen interaction requires users to adjust the joint angle using a mouse or keyboard without prior simulation, resulting in potential collisions with the ground, object, or even the robot itself before successfully completing the task.

As for the successful completion time, almost all participants spent less time on AR operation, primarily because the virtual robot exhibits a faster response rate, while the real robot arm experiences execution delays. This higher efficiency in moving the virtual robot arm contributed to the participant's preference for AR operation, as they felt more at ease using the virtual arm without worrying about collisions, leading to more decisive and efficient actions.

5.2.2 Subjective metrics

The analysis of subjective data, i.e., the values gathered from the questionnaire, reveals interesting insights into the comparison between AR and 2D screen interaction. The p values of questions Q1–Q7 in sections 1 and 2 are 0.0021, 0.0018, 0.0306, 0.2572, 0.0793, 0.0013, and 0.0054, respectively. Only the p values of Q4 and Q5 exceed 0.05, while the others are significantly smaller, indicating that the levels of physical demand and accuracy are not significantly different between the two interaction methods. However, the other aspects exhibit significant variations. The overall p value for all questionnaires comparing AR and 2D screen interaction is 2.441e−11. The average value for each question can be found in Fig. 13. The data clearly shows that the AR method outperforms conventional 2D screen interaction in all aspects. Questions Q6 and Q7, in particular, reveal that efficiency and willingness to adopt the technology are the most significant advantages of AR interaction. The high efficiency of the AR system, as evidenced by previous objective metrics, is further supported by the users' perception, emphasizing the effectiveness of the AR system. The participants also demonstrated a high level of enthusiasm for the new interaction method, expressing greater interest and willingness to adopt AR interaction in future manufacturing scenarios compared to 2D screen interaction.

Regarding Q4, which assesses the level of physical demand, the AR method only slightly surpasses the 2D method, as well as in Q5, which examines accuracy. In terms of physical demand, the AR device is somewhat heavy and requires some degree of gesture interaction. However, the interaction method is intuitive, and users do not need to use a mouse or keyboard, conserving physical strength. Consequently, the survey results show that the AR method holds a slight advantage. The accuracy of the AR system relies primarily on two factors: visualization accuracy, such as the location registration between AR and real robotic arms, and control accuracy. Participants reported that the former was not a significant issue. However, the latter could be affected by hand tremors during gesture control, leading to misplacement. Fortunately, the DT's simulation functionality mitigates the impact of hand tremors on the real robot.

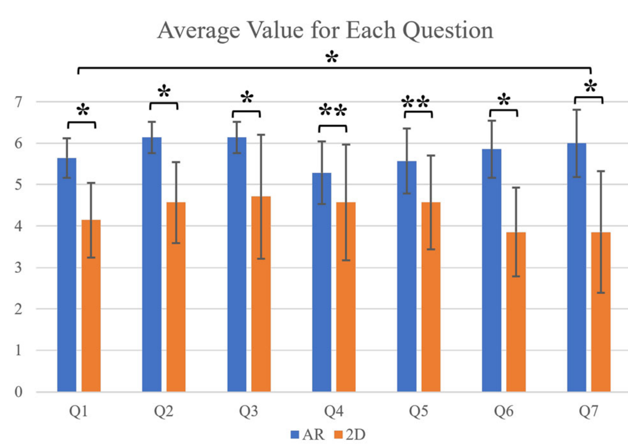


Fig. 13 The average value for collected questionnaire in "Appendix B" (Error-bars show ±1 SD; $*p<0.05$, $**p>0.05$.). In the questionnaire, the 7-Likert scale is adopted, with 1 being the worst and 7 being the best

In Q1, all participants assigned higher values to the AR system, indicating a superior user experience compared to 2D screen interaction. In Q2, participants unanimously considered the AR system more comfortable for learning and controlling the robot. However, some users initially struggled with the AR gesture control, necessitating a training session before the experiment. Q3 addresses mental demand, which is primarily due to confusion regarding the corresponding joint angle and concerns about collisions. AR offers an intuitive control method by directly linking to the real joint, eliminating the need for users to memorize which slider controls which joint and accelerating the learning process. Additionally, the AR system's simulation functionality alleviates users' concerns about collisions.

The feedback from participants on section 3 has been collected in Table 1. The most frequently requested features for Q8 include collision detection, an integrated AR demonstration tutorial, and trajectory replay. For Q9, participants reported difficulties with gesture recognition, glasses affecting their view, challenges in identifying collisions with the ground, and difficulties in selecting objects in AR using the finger pinch gesture. In Q10, they suggested providing hints to users on operating the robot arm, improving location registration between the AR and real robotic arms, and providing additional instruction on AR gesture interaction.

The controlled experiment conducted in this study compared the performance of AR and conventional 2D screen interaction in a manufacturing scenario. The objective metrics, evaluated using statistical methods, indicated that the AR method outperformed the 2D screen interaction in terms of number of errors and successful completion time. The subjective metrics, gathered from the questionnaire, revealed that the AR method provided a superior user experience, with higher efficiency and willingness to adopt the technology. However, the physical demand and accuracy of the AR system only slightly surpassed those of the 2D method. The participants also provided valuable feedback on the AR system's strengths and weaknesses, suggesting improvements such as collision detection, an integrated AR demonstration tutorial, and improved location registration. Overall, the study demonstrated the potential of AR technology in enhancing manufacturing processes, providing a more intuitive and efficient interaction method.

The user study also corroborates that AR-facilitated interactions confer a significant benefit by leveraging human cognitive prowess to improve DT's performance in the context of a pick-and-place task. The notion of a 'Cognitive Digital Twin' embodies the synthesis of human experiential knowledge within the digital twin's process of enhancement. The infusion of human insight provides a layer of contextual acumen and strategic oversight, which fosters a level of sophisticated decision-making that transcends the limitations of static computational algorithms. In assessing the efficacy of AR against conventional 2D user interfaces in terms of engaging human operators, it is evident that the AR platform, with its immersive representation of robotic mechanisms, empowers operators to manage robotic functions with an intuitive and expedited approach. This form of direct interaction engenders a dynamic educational exchange and refines operational processes via a synergistic human–robot interface. The results of the study indicate that the AR-enabled DT system surpasses its 2D interface counterpart in accuracy and efficiency, as evidenced by a reduction in operational errors and a decrease in the time required to complete tasks. Furthermore, the respondents expressed a greater level of positivity towards AR manipulation and a willingness to use the AR interface in the questionnaire. These findings underscore the AR's advantages of integrating human intelligence into the DT framework, thus yielding a more reactive and adaptable HRI scenario.

Table 1 The questionnaire feedback for potential improvements

Section 3	Potential improvements
Q8	What features or functionalities would you like to see added to the system?
Feedback	Collision detection, an integrated AR demonstration tutorial, and trajectory replay
Q9	Were there any aspects of the system that were confusing or difficult to use?
Feedback	Gesture recognition, glasses affecting eyesight, identifying collisions with the ground, and selecting objects in AR using the finger pinch gesture
Q10	Do you have any other feedback or suggestions for improving the system?
Feedback	Providing hints to users on operating the robot arm, improving location registration between the AR and real robotic arms, and providing additional instruction on AR gesture interaction

6 Conclusion

In this study, we present a cutting-edge, AR-based digital twin system aimed at augmenting HRI in the manufacturing sector. Our proposed system artfully combines AR technology and DT principles to generate a virtual counterpart of a robot arm, achieving three distinct DT levels: virtual twin, hybrid twin, and cognitive twin. This innovative approach offers unique functionalities for real-time monitoring, intuitive interaction and control, and optimized operation through human intelligence. At the virtual twin level, the system facilitates in-situ monitoring through the transmission of physical to virtual data. In the hybrid twin level, bidirectional message transmission between virtual and physical entities is permitted, and an innovative AR-based interface is developed to enable users to engage with the DT using natural gestures. This streamlines the processes of programming and control, fostering a more intuitive interaction. At the cognitive twin level, human intelligence is incorporated into the interaction, resulting in optimized operation. A human-centric user study substantiates the system's effectiveness in reducing operation time, minimizing errors, and enhancing overall productivity within manufacturing environments. As per participant feedback, the proposed system significantly elevates their experience, increasing their inclination to utilize AR technology in future manufacturing tasks. Our research contributes to the progression of advanced manufacturing solutions by capitalizing on the potential of AR and DT technologies in HRI.

This study makes three key contributions. Firstly, it integrates a three-level DT system comprising the virtual twin for in-situ monitoring, the hybrid twin for intuitive interaction, and the cognitive twin for optimized operation. This reveals a comprehensive and embedded structure from virtual to cognitive levels. Secondly, it introduces intuitive interactive schemes specifically tailored for AR interfaces, which enable more nuanced and direct control over the DT, thereby enhancing human–robot interaction. Lastly, the study emphasizes the empirical validation of the framework through user studies, providing insights into its practical benefits and effectiveness in real-world scenarios, and offering evidence of its advantages over traditional approaches.The study's limitations encompass various aspects, including system performance, high-level DT implementation, and practical manufacturing scenarios. Given that the DT system is still in its early stages, future research will concentrate on refining the proposed system based on feedback garnered from user studies. This will involve enhancing collision detection capabilities with real-world objects, integrating an augmented reality demonstration tutorial, and implementing trajectory replay functionality. Additionally, the integration of machine intelligence with human intelligence in a harmonious manner for the human-centric industry has been overlooked, despite the current cognitive level primarily encompassing human intelligence. Attaining a more advanced level of DT will necessitate the integration of advanced artificial intelligence algorithms for optimized operation and autonomous decision-making. Finally, while this research only applies a simple yet widely used scenario at the application level, exploring the suitability of our system in more complex manufacturing scenarios will be pursued.

Appendix A: DH parameters

See Fig. 14.

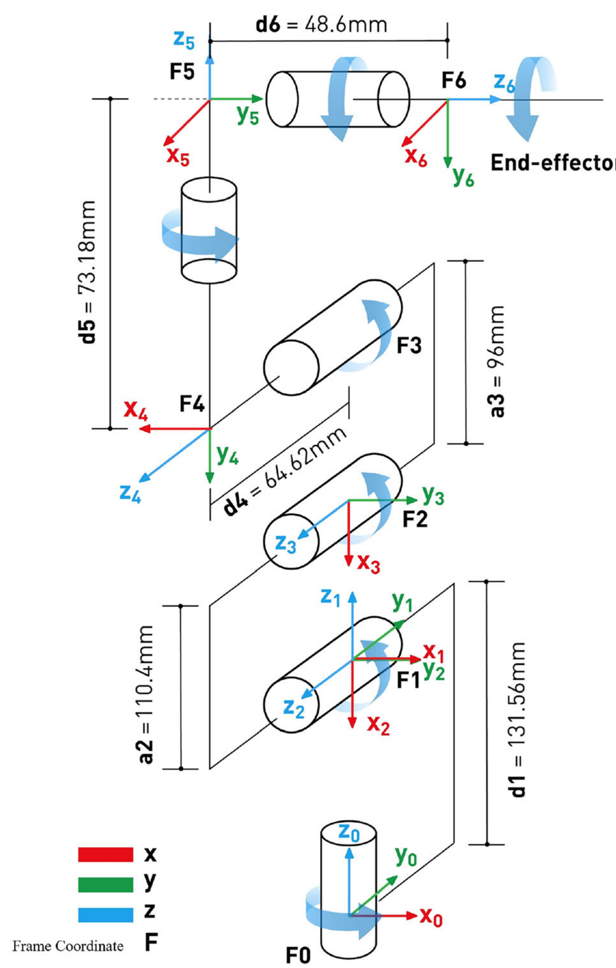


Fig. 14 The DH parameters for the robot arm (ER myCobot 280 Pi)

Appendix B: Questionnaire

See Table 2.

Table 2 The questionnaire for post-experiment survey

	Questions
Section 1	User experience
Q1	How would you rate your overall experience using the AR-based digital twin system during the experiment?
Q2	How easy was it to navigate the interface?
Q3	How mentally demanding was the task? (7 means no mentally demanding)
Q4	How physically demanding was the task? (7 mean no physically demanding)
Section 2	System performance
Q5	How would you rate the overall accuracy of the AR-based digital twin system compared to the 2D screen used by the control group?
Q6	How would you rate the overall effectiveness of the AR-based digital twin system compared to the 2D screen used by the control group?
Q7	How would you rate the willingness to use the AR-based digital twin system in future manufacturing tasks?
Section 3	Potential improvements
Q8	What features or functionalities would you like to see added to the system?
Q9	Were there any aspects of the system that were confusing or difficult to use?
Q10	Do you have any other feedback or suggestions for improving the system?

Acknowledgements This research is supported by The Hong Kong University of Science and Technology (Guangzhou) and the Department of Science and Technology of Guangdong Province (2021QN02Z112).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Amtsberg F, Yang X, Skoury L et al (2021) iHRC: an AR-based interface for intuitive, interactive and coordinated task sharing between humans and robots in building construction. In: ISARC. Proceedings of the international symposium on automation and robotics in construction. IAARC Publications, pp 25–32

Arevalo Arboleda S, Rücker F, Dierks T et al (2021) Assisting manipulation and grasping in robot teleoperation with augmented reality visual cues. In: Proceedings of the 2021 CHI conference on human factors in computing systems, pp 1–14

Baicun W, Yuan X, Jianlin Y et al (2020) Human-centered intelligent manufacturing: overview and perspectives. Strateg Study CAE 22(04):139

Baroroh DK, Chu CH, Wang L (2021) Systematic literature review on augmented reality in smart manufacturing: collaboration between human and computational intelligence. J Manuf Syst 61:696–711

Bhatt PM, Malhan RK, Shembekar AV et al (2020) Expanding capabilities of additive manufacturing through use of robotics technologies: a survey. Addit Manuf 31:100933

Breque M, De Nul L, Petridis A (2021) Industry 5.0: towards a sustainable, human-centric and resilient European industry. European Commission, Directorate-General for Research and Innovation, Luxembourg

Brooke J (1996) SUS—a quick and dirty usability scale. Usab Eval Ind 189(194):4–7

Cai Y, Wang Y, Burnett M (2020) Using augmented reality to build digital twin for reconfigurable additive manufacturing system. J Manuf Syst 56:598–604

Cao Y, Xu Z, Li F et al (2019) V. ra: an in-situ visual authoring system for robot-IoT task planning with augmented reality. In: Proceedings of the 2019 on designing interactive systems conference, pp 1059–1070

Cauchard JR, Tamkin A, Wang CY et al (2019) Drone. io: a gestural and visual interface for human-drone interaction. In: 2019 14th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 153–162

Chen L, Takashima K, Fujita K et al (2021) Pinpointfly: an egocentric position-control drone interface using mobile ar. In: Proceedings of the 2021 CHI conference on human factors in computing systems. Association for Computing Machinery, New York. https://doi.org/10.1145/3411764.3445110

Choi S, Cai Y (2014) A virtual prototyping system with reconfigurable actuators for multi-material layered manufacturing. Comput Ind 65(1):37–49

Dimitropoulos N, Togias T, Zacharaki N et al (2021) Seamless human–robot collaborative assembly using artificial intelligence and wearable devices. Appl Sci 11(12):5699

Eswaran M, Bahubalendruni MR (2022) Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: a state of the art review. J Manuf Syst 65:260–278

Fuste A, Reynolds B, Hobin J et al (2020) Kinetic ar: a framework for robotic motion systems in spatial computing. In: Extended abstracts of the 2020 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, pp. 1–8. https://doi.org/10.1145/3334480.3382814,

Grieves M, Vickers J (2017) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. Transdisciplinary perspectives on complex systems: new findings and approaches, pp 85–113

Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol 52:139–183

Hassenzahl M, Burmester M, Koller F (2003) Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. Mensch & Computer 2003: Interaktion in Bewegung, pp 187–196

Hedayati H, Walker M, Szafir D (2018) Improving collocated robot teleoperation with augmented reality. In: Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction, pp 78–86

Hoang KC, Chan WP, Lay S et al (2022) Virtual barriers in augmented reality for safe and effective human-robot cooperation in manufacturing. In: 2022 31st IEEE international conference on robot and human interactive communication (RO-MAN). IEEE, pp 1174–1180

Huang S, Wang B, Li X et al (2022) Industry 5.0 and society 5.0-comparison, complementation and co-evolution. J Manuf Syst 64:424–428

ISO I (2020) Dis 23247-1 automation systems and integration-digital twin framework for manufacturing. International Organization for Standardization, Geneva

Kuts V, Otto T, Tähemaa T et al (2018) Adaptive industrial robots using machine vision. In: ASME international mechanical engineering congress and exposition. American Society of Mechanical Engineers, p V002T02A093

Leng J, Zhu X, Huang Z et al (2023) Manuchain II: blockchained smart contract system as the digital twin of decentralized autonomous manufacturing toward resilience in industry 5.0. IEEE Trans Syst Man Cybern: Syst

Li C, Zheng P, Yin Y et al (2023) An AR-assisted Deep Reinforcement Learning-based approach towards mutual-cognitive safe human-robot interaction. Robot Comput-Integr Manuf 80:102471. https://doi.org/10.1016/j.rcim.2022.102471

Lindlbauer D, Grønbæk JE, Birk M et al (2016) Combining shape-changing interfaces and spatial augmented reality enables extended object appearance. In: Proceedings of the 2016 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, pp 791–802. https://doi.org/10.1145/2858036.2858457

Liu S, Wang XV, Wang L (2022a) Digital twin-enabled advance execution for human–robot collaborative assembly. CIRP Ann 71(1):25–28

Liu X, Zheng L, Wang Y et al (2022b) Human-centric collaborative assembly system for large-scale space deployable mechanism driven by digital twins and wearable AR devices. J Manuf Syst 65:720–742

Liu YK, Ong SK, Nee AYC (2022c) State-of-the-art survey on digital twin implementations. Adv Manuf 10(1):1–23. https://doi.org/10.1007/s40436-021-00375-w

Lu Y, Wang H, Feng N et al (2022) Online interaction method of mobile robot based on single-channel EEG signal and end-to-end CNN with residual block model. Adv Eng Inform 52:101595

Mukherjee D, Gupta K, Chang LH et al (2022) A survey of robot learning strategies for human–robot collaboration in industrial settings. Robot Comput-Integr Manuf 73:102231

Müller F, Deuerlein C, Koch M (2021) Cyber-physical-system for representing a robot end effector. Procedia CIRP 100:307–312

Okegbile SD, Cai J, Yi C et al (2022) Human digital twin for personalized healthcare: vision, architecture and future directions. IEEE Netw 37:262–269

Ostanin M, Mikhel S, Evlampiev A, et al (2020) Human-robot interaction for robotic manipulator programming in mixed reality. In: 2020 IEEE international conference on robotics and automation (ICRA), pp 2805–2811. IEEE

Özgür A, Lemaignan S, Johal W et al (2017) Cellulo: versatile handheld robots for education. In: Proceedings of the 2017 ACM/IEEE international conference on human–robot interaction, pp 119–127

Paripooranan CS, Abishek R, Vivek D et al (2020) An implementation of AR enabled digital twins for 3-d printing. In: 2020 IEEE international symposium on smart electronic systems (iSES) (Formerly iNiS). IEEE, pp 155–160

Pedersen EW, Hornbæk K (2011) Tangible bots: interaction with active tangibles in tabletop interfaces. In: Proceedings of the SIGCHI conference on human factors in computing systems. Association for Computing Machinery, New York, pp 2975–2984. https://doi.org/10.1145/1978942.1979384

Peng H, Briggs J, Wang CY et al (2018) Roma: interactive fabrication with augmented reality and a robotic 3d printer. In: Proceedings of the 2018 CHI conference on human factors in computing systems, pp 1–12

Quintero CP, Li S, Pan MK et al (2018) Robot programming through augmented trajectories in augmented reality. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 1838–1844

Sifat MMH, Choudhury SM, Das SK et al (2022) Towards electric digital twin grid: technology and framework review. Energy AI 11:100213

Sisinni E, Saifullah A, Han S et al (2018) Industrial internet of things: challenges, opportunities, and directions. IEEE Trans Ind Inf 14(11):4724–4734

Siu AF, Yuan S, Pham H et al (2018) Investigating tangible collaboration for design towards augmented physical telepresence. In: Design thinking research: making distinctions: collaboration versus cooperation, pp 131–145

Suzuki R, Karim A, Xia T et al (2022) Augmented reality and robotics: a survey and taxonomy for AR-enhanced human–robot interaction and robotic interfaces. In: CHI conference on human factors in computing systems. ACM, New Orleans, pp 1–33. https://doi.org/10.1145/3491102.3517719

Takashima K, Oyama T, Asari Y et al (2016) Study and design of a shape-shifting wall display. In: Proceedings of the 2016 ACM conference on designing interactive systems, pp 796–806

Tao F, Sui F, Liu A et al (2019) Digital twin-driven product design framework. Int J Prod Res 57(12):3935–3953

Tuegel EJ, Ingraffea AR, Eason TG et al (2011) Reengineering aircraft structural life prediction using a digital twin. Int J Aerosp Eng 2011:1687–5966

Villani V, Pini F, Leali F et al (2018) Survey on human-robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatronics 55:248–266

Wan J, Li X, Dai HN et al (2020) Artificial-intelligence-driven customized manufacturing factory: key technologies, applications, and challenges. Proc IEEE 109(4):377–398

Wang B, Zheng P, Yin Y et al (2022) Toward human-centric smart manufacturing: a human-cyber-physical systems (hcps) perspective. J Manuf Syst 63:471–490

Watanabe A, Ikeda T, Morales Y et al (2015) Communicating robotic navigational intentions. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 5763–5769. https://doi.org/10.1109/IROS.2015.7354195

Yin Y, Zheng P, Li C et al (2023) A state-of-the-art survey on Augmented Reality-assisted Digital Twin for futuristic human-centric industry transformation. Robot Comput-Integr Manuf 81:102515. https://doi.org/10.1016/j.rcim.2022.102515

Yuan L, Reardon C, Warnell G et al (2019) Human gaze-driven spatial tasking of an autonomous MAV. IEEE Robot Autom Lett 4(2):1343–1350