



Construction of Human-Robot Cooperation Assembly Simulation System Based on Augmented Reality

Qiang Wang¹, Xiumin Fan^{1,2}✉, Mingyu Luo¹, Xuyue Yin¹,
and Wenmin Zhu¹

¹ Institute of Intelligent Manufacturing and Information Engineering,
School of Mechanical Engineering, Shanghai Jiao Tong University,
Shanghai 200240, China
xmfan@sjtu.edu.cn

² Shanghai Key Lab of Advanced Manufacturing Environment,
Shanghai 200030, China

Abstract. Human-Robot cooperation (HRC) is the developing trend in the field of industrial assembly. Design and evaluation of the HRC assembly workstation considering the human factor is very important. In order to evaluate the transformational construction scenario of a manual assembly workstation to a HRC workstation fast and safely, a HRC assembly simulation system is constructed which is based on Augmented Reality (AR) with human-in-loop interaction. It enables a real operator to interact with virtual robot in a real scene, and the assembly steps of real workers can be restored and mapped to a virtual human model for further ergonomic analysis. Kinect and LeapMotion are used as the sensors for human-robot interaction decision and feedback. An automobile gearbox assembly is taken as an example for different assembly task verification, operators' data are collected and analyzed by RULA scores and NASA-TLX questionnaires. The result shows that the simulation system can be used for the human factor evaluation of different HRC task configuration schemes.

Keywords: Augmented Reality · Human-Robot Cooperation · Human factors

1 Introduction

Human-Robot Cooperation assembly is the developing trend of industrial assembly field in the future. How to design and evaluate HRC assembly station is a current research hotspot, but there hasn't been much research of human factors evaluation for HRC assembly work station planning. With the development of Virtual Reality (VR) and Augmented Reality, more and more researchers apply VR/AR technologies to industrial simulation and training. It can not only be used to train and guide operators, but also to collect operators' motion data for later analysis.

The HRC simulation system based on AR and VR can measure human factor data more realistically and securely. Matsas et al. [1] proposed a VR training system to solve human-robot contact and collision problems through visual and sound interaction. Qiyue Wang et al. [2] proved the safety and effectiveness of the VR welding system

with HRC through the VR simulation experiment. Besides, Neuhofer et al. [3] compared the simulation experiments of HRC based on VR and AR respectively, and proved that AR could bring more real simulation environment to participants than VR. D. Ni [4] designed an AR human-robot cooperation system based on tactile feedback. By collecting the depth image to reconstruct the model, and combining with the tactile sensor to define the welding path, the accuracy of welding operation was improved. Michalos et al. [5] developed an AR-based human-robot interaction system with a tablet PC, which improves the safety of operators and productivity by safety area division, production data visualization and voice alarm. However, it also points out that in the industrial scene, it is necessary to work with both hands, so wearing AR glasses will be a better choice, and the human factors are not considered enough in the system design process.

Based on the research above, this paper proposes to use AR to build a HRC assembly simulation system for real operators. The idea of human-in-the-loop is adopted for the human-robot interaction so as to build the system framework and lay the foundation for the subsequent exploration of human factors.

2 Simulation System Architecture

This paper builds the system architecture based on the communication of heterogeneous software platforms, respectively Unity in Window and ROS (Robot Operating System) in Ubuntu. The merit is that it can make full use of the advantages of each platform. It can not only use the development ability of ROS environment for quick, flexible and efficient debugging and compilation, but also use the excellent 3D engine rendering ability of Unity environment to present the virtual reality fusion scene. Moreover, the multi-platform architecture improves the adaptability and transportability of system. The modules between the two platforms are independent of each other, so that the verification experiment and workstation construction can be carried out with little changes in robot simulation.

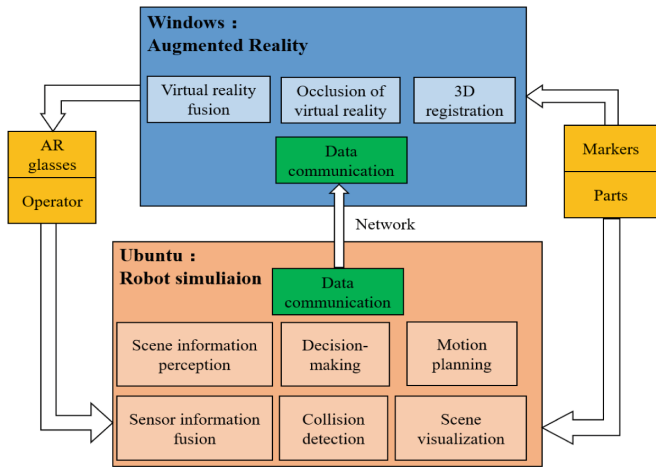


Fig. 1. Architecture of HRC simulation system

The architecture is shown in Fig. 1. The orange blocks on both sides of the diagram denote the real scene. The light red block in the diagram below shows the simulation system related to the virtual robot, reflecting the perception, decision-making and feedback of the robot in HRC. The blue block in the diagram above shows the augmented reality display system related to the real operator, reflecting the observation method of the virtual reality fusion scene when the human is in the simulation loop. The green block shows the communication between the two heterogeneous software platform.

In the construction of the robot simulation system based on AR, the robot simulation module judges the status of the robot and plan the motion path according to the interaction content, environment changes and predefined tasks in real time. The augmented reality display module is used to display the virtual robot and its motion in front of an operator through AR glasses.

In this paper, Rosbridge communication protocol [6] is used to realize the network communication between heterogeneous platforms. Its core idea is to use websocket protocol as a bridge to transmit data between ROS and non ROS environments. From the perspective of ROS, data is a message node in ROS. From the perspective of non ROS, data is in JSON data format. The Rosbridge protocol is equivalent to the translation between ROS and non ROS environments, and realizes full duplex communication.

Figure 2 shows the Rosbridge communication protocol integration under the ROS message publishing mode. The ROS side sets the `\rosbridge_server` as the message node, subscribes to the `\joint_states` message data, converts data to JSON format, and sends the subscribed `\joint_states` message to the network. On the Unity side, the Rosbridge Client receives data and drives the virtual robot to move in real time.

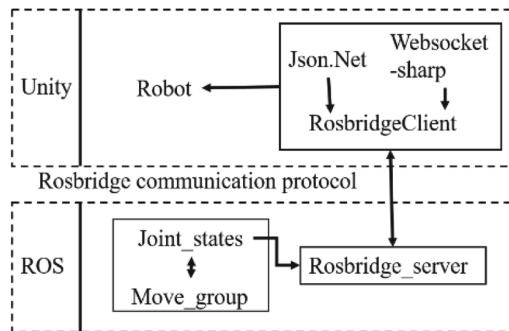


Fig. 2. Data communication based on Rosbridge

After the HRC simulation system is built, the robot driving display is shown in Fig. 3. The robot assembly simulation module based on ROS completes the motion planning and uses the heterogeneous platform communication mechanism to send the motion data. On the other hand, the augmented reality display module based on Unity receives the motion data and drives the AR robot to move in real time.

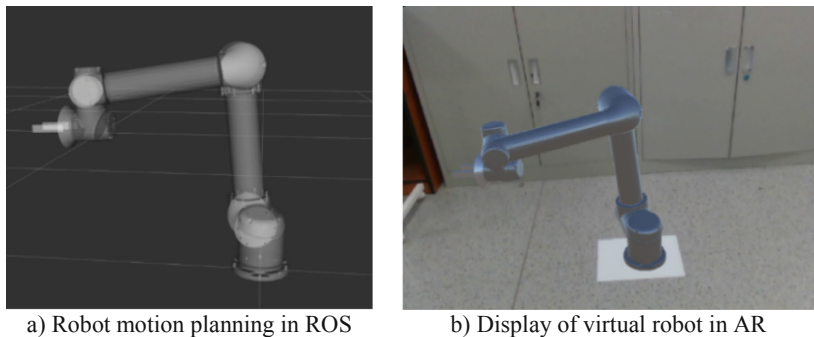


Fig. 3. Robot motion simulation with data driven from operator

3 Selection of AR Display Scheme

The display device determines the final virtual-real scene fusion effect. Generally, according to different display equipment, it can be divided into head mounted displays which are also called AR glasses, mobile displays such as mobile phones, flat or fixed displays such as computer screens or projectors. In the assembly operation, operators need both hands to work while constantly changing the viewing angle. Using AR glasses as display devices can make virtual-real scene fusion on the premise of freeing hands.

The display scheme with AR glasses can be divided into video see through (VST) mode and optical see through (OST) mode according to the display principle [7]. The biggest difference between the two schemes is the way to obtain the real scene image, in which VST is obtained by camera, while OST is obtained by eyes. Figure 4 shows the display effect under the two schemes.

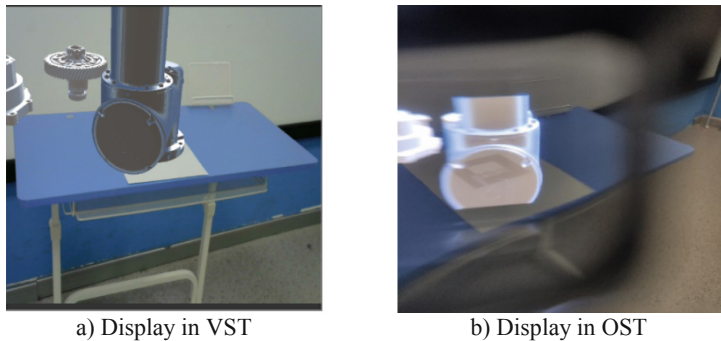


Fig. 4. Effect comparison of different AR scheme

In OST mode, the real scene is observed by eyes. However, the 3D registration position obtained by a camera is the position of the camera relative to the real scene matrix P_{co} rather than the position of eyes relative to the real scene matrix P_{eo} . Therefore, in the OST mode, the camera needs to calibrate the pose of eyes and itself one more time than the VST mode, so as to obtain the relative matrix P_{ec} . The relationship among the three matrix is as follow:

$$P_{ec} = P_{eo} * P_{co} \quad (1)$$

From the perspective of principle and display effect, compared with the VST mode, the OST mode has the following problems:

- 1) Limitation of the accuracy and universality of camera-eye calibration. The calibration method of P_{ec} is monocular, which can't completely calibrate the binocular coordinates, and the subjective cognition will affect the accuracy of the calibration matrixes. In addition, due to the different physiological structure, observation ability and cognition of each person, the calibration need to be conducted for different users.
- 2) Registration delay error between the virtual scene and the real world. The 3D registration needs additional calculation time, while the real scene directly enters eyes through light reflection almost without time delay. It leads to the inconsistency between the superposition of the virtual scene and the real world, resulting in a sense of violation.
- 3) Field of view (FOV) limitation of AR glasses at present time. Due to the FOV limitation of display screen, virtual objects can't be fully presented within the scope of eyes, resulting in the cognitive discordance.

As a result, the VST mode is superior to the OST mode in terms of registration accuracy, display effect and real-time performance. By measuring the data of 16 users wearing AR glasses and manually adjusting the coincidence degree of virtual object and real scene, Plopski [8] also proved that through VST, the operator could better judge the location of virtual model in real world. Therefore, this paper chooses the VST scheme to implement the AR module.

4 Design the Human-Robot Interaction Scheme

The interaction scheme with safe and natural way is the premise and advantage of human-robot collaboration [9]. This paper presents a human-robot interaction model based on sensors and AR technology, as shown in Fig. 5.

In the process of interaction, the operator has autonomy without additional design, while the robot needs to perceive the environmental information and the operator's intention to judge and make decisions according to the information processing, so as to realize the real-time interaction with the operator. In this section, Kinect and Leap-Motion are used as sensors to acquire the environmental information and hands interaction information of an operator, then establish a decision mechanism to integrate multiple interaction information, so that the robot can make decision independently.

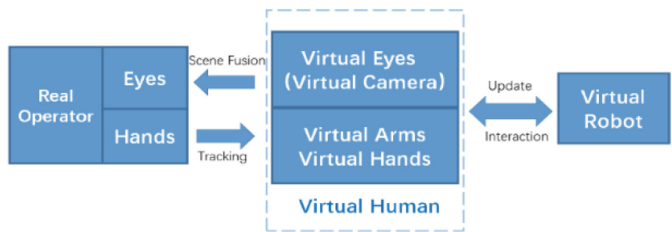


Fig. 5. Human robot interaction model

4.1 Interaction Information Tracking Based on Kinect

Microsoft Kinect is used to track the motion of operators’ limbs and real environment. A Kinect structure is shown in Fig. 6. The camera can obtain color image with a resolution of 1920×1080 , and the depth image can be obtained by time of flight technology to generate point cloud information with a resolution of 512×424 . Therefore, Kinect can be used to collect the 3D information of real environment and map this information to the simulation space of the virtual robot, so as to realize the robot’s perception of the external world.

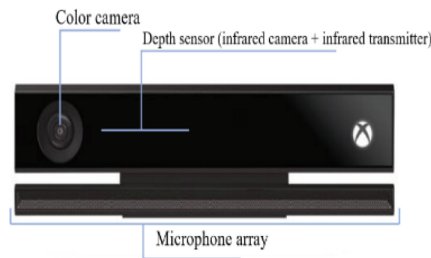


Fig. 6. Kinect structure

However, the size of point cloud information obtained by Kinect is too big, and a frame of image has more than 200000 point cloud. Figure 7 shows the comparison between original image and point cloud image, which can’t be imported into the simulation environment for real-time calculation. So Octomap [10] method is used to reconstruct point cloud information, and the octree structure is used to store the depth image. By sacrificing the accuracy, the calculation efficiency is greatly improved. Figure 8 shows the point cloud transforming into octomap model.

For the octomap models, the higher the resolution is, the more real and rich the 3D information is, and the slower the calculation speed is, and vice versa. In addition, the size of data can be controlled by setting the range of the octomap model, then the calculation speed can be changed at the same time. Considering that the HRC assembly is a close operation between human and robot, the activity range of the robot is within 1 m, so the final parameters in octomap model are determined as the resolution of 0.05 m and the sensing distance of 1 m.

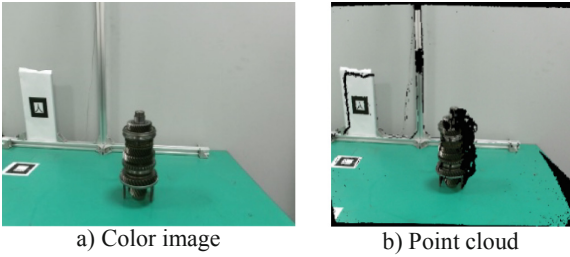


Fig. 7. Comparison between image and point

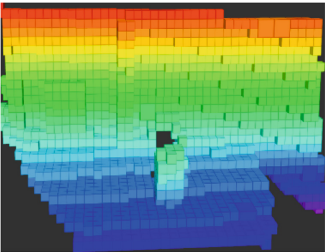


Fig. 8. Octomap model

4.2 Gesture Recognition for Interaction Based on LeapMotion

Interaction by hand gesture is one of the basic ways of communication between human beings and the outside world, and it is an intuitive and natural non-contact interaction method [11]. Through image recognition, specific gestures can be designed and used as means of human-robot interaction to realize the robot’s judgment on human action. LeapMotion sensor is used for hand gesture recognition. As shown in Fig. 9, LeapMotion can capture 215 frames image per second, and calculate the position of hands in cartesian coordinates of right hand. What’s more, it can detect more than 20 joints of hand with an accuracy of 0.01 mm. Figure 10 shows the results of hand detection using LeapMotion.

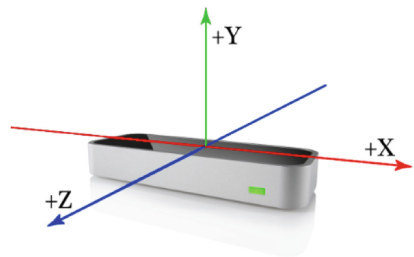


Fig. 9. LeapMotion



Fig. 10. Hand detection with LeapMotion




Gesture pictures			
Gesture name	OK	PREV	NEXT
Gesture semantics	The operator is ready, robot starts to work	Command robot to continue Previous action	Command the robot to continue next action

Fig. 11. Hand gesture definition

For the easiness of human-robot interaction, the naturalness (whether it is too difficult to make gestures), discrimination (whether it has obvious feature differentiation), intuitiveness (whether gestures conform to human behavior logic and associative content) of gestures are considered. Three gestures are selected and the semantics of each one are defined, as shown in Fig. 11. They are OK (ready to begin), PREV (the previous step) and NEXT (the next step) respectively. They are consistent with the corresponding semantics of the human behavior logic, and are commonly used in interpersonal communication. They don't affect the human behavior logic, so as to ensure the natural and intuitive of human behavior.

Experiments on above three gestures are carried out, as shown in Table 1. The average recognition rate is 99.46%, close to 100%, which means these hand gestures can be recognized effectively and accurately.

Table 1. Experimental data of gestures recognition

Hand gesture	Total test number	Correct recognized number	Recognition rate
OK	1445	1434	99.24%
PREV	1065	1062	99.72%
NEXT	850	845	99.41%
Average	1120	1114	99.46%

4.3 Decision Mechanism During Interaction

In the process of assembly, robots need to constantly make feedback according to the information input from sensors. Therefore, a decision mechanism is established to help robots make decisions. The first principle that a robot needs to abide is safety, which ensures the safety of humans. The second principle is efficiency, which requires a robot to make decisions quickly in a short time and ensure the efficient completion of assembly tasks. The third principle is response to humans' orders, which requires a robot to accurately execute the orders given by humans. The decision mechanism is shown in Fig. 12.

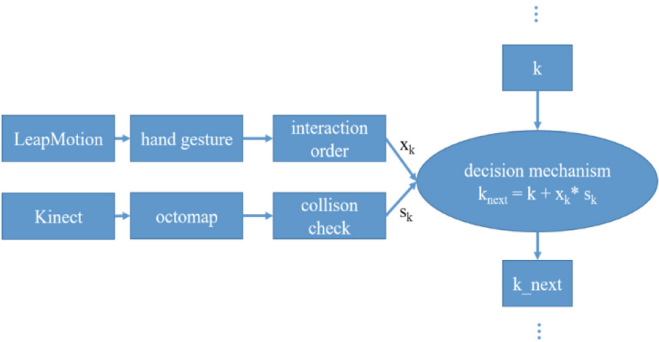


Fig. 12. Decision mechanism

The gesture recognition is obtained from LeapMotion and converted into interaction commands. In Fig. 12, the x_k represents the k steps of assembly operations when the interaction command from the outside is obtained. When operators give the NEXT operation command, then $x_k = 1$. When operators give the PREV operation command, then $x_k = -1$.

The octomap model is obtained from Kinect and used for collision detection with the virtual robot. The s_k represents the collision detection and motion planning results in the current scene. When the robot occurs collision interference or fails to complete the motion planning, then $s_k = 0$. When the robot is not interfered with the octomap model, then $s_k = 1$.

If the robot is currently in the k^{th} step of assembly operations, and after integrating the interaction command and octomap model information, the robot can make the decision for the next assembly operation k_{next} :

$$k_{\text{next}} = k + x_k * s_k \quad (2)$$

By this means, the robot decision-making process can comprehensively consider the external scene information and the operator's instructions to carry out an automatic obstacle avoidance planning, which ensures a natural, safety and real-time interaction.

5 Experiments and Analysis

5.1 Prototype System and Experiments Design

A HRC assembly simulation prototype system is constructed which includes hardware and software. The hardware configuration is shown in Fig. 13, two computers are needed. The one is installed with Ubuntu system for robot motion planning, and the other is installed with Windows system for augmented reality display. Kinect and LeapMotion are connected to computer in Ubuntu system as robot sensors, and also used for driving digital human model for human factor analysis. RealSense and NDI AR glasses are connected to computer in windows system as AR camera and display screen. Router is used to build network to complete communication between two computers.

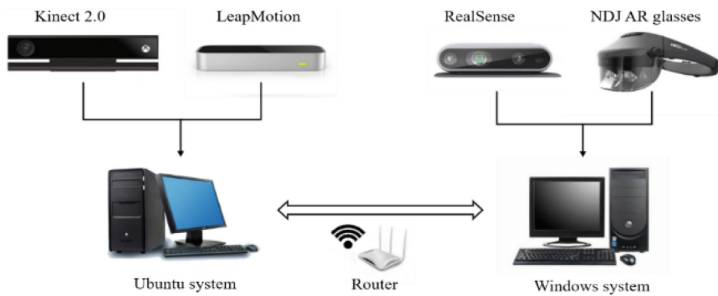


Fig. 13. Hardware device configuration

The software architecture makes full use of advantages of the module independence of ROS platform. Each module directly uses ROS to build message nodes for data distribution and transmission. The advantage is that it improves the independence and integrity of each module, and can easily and quickly add new functions to each module, as shown in Fig. 14.

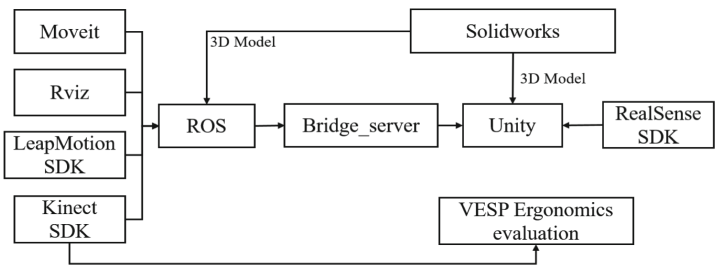


Fig. 14. Modules structure of software system

An automobile gearbox is taken as an assembly object for testing. In this HRC assembly experiment, the layout configuration of the assembly workstation is as shown in Fig. 15. Some parts for assembly process steps are selected, and these assembly parts are assembled to the lower gearbox house which is placed in the center of the workbench and fixed. A real operator and a virtual robot complete the assembly task together. According to the characteristics of strong load capacity and low flexibility of robots, and weak load capacity and high flexibility of human, heavier parts are assigned to the robot, while parts with higher assembly flexibility requirements are assigned to the operator. The assembly task assignment schemes are shown in Fig. 16.

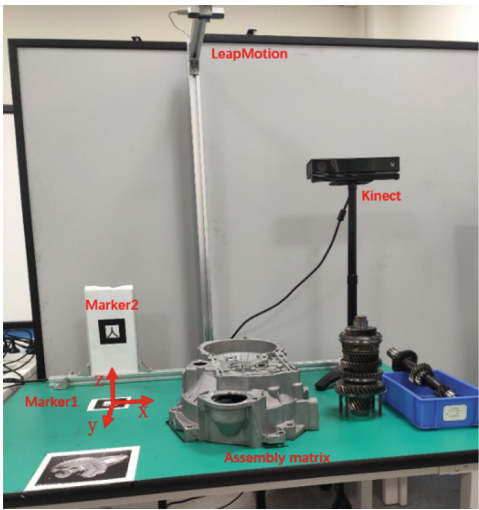












Fig. 15. Assembly workstation for HRC experiment

Assembly steps	① Differential installation	② Output shaft installation	③ Input shaft installation	④ Reverse gear installation	⑤ Reverse gear shaft installation
Assembly scheme 1					
Assembly scheme 2					



 Human assembly Robot assembly

Fig. 16. Assembly task assignment scheme

Six testers are invited to repeatedly assemble and disassemble task for 10 times under each experimental scheme, so as to ensure the stability of experimental results. The starting of the robot’s movement is controlled by hand gesture. When each tester completes assembly operation, he use the “next” gesture to send command to the robot. It should be noted that only when the robot needs to be stopped and restarted, the interaction gesture would be used. Otherwise, the robot will continue to finish its operation steps. For example, in assembly scheme 1, the robot will maintain a waiting state after completing the first step. When each tester finishes the second step, the third step needs to be started with a hand gesture. On the contrary, in assembly scheme 2, the robot continuously carries out the first and second assembly step without waiting for gesture commands.

Figure 17 shows the process of assembly experiment. The picture a) is the third-person perspective of HRC assembly experiment, and the picture b) is the first-person perspective of the operator in HRC assembly experiment. During the experiment, Kinect is used to obtain the testers’ motion data, and Rapid Upper Limb Assessment (RULA) is used to analyze the testers’ physiological fatigue. After experiment, all testers are invited to fill in the National Aeronautics and Space Administration-Task Loads Index (NASA-TLX) [12] questionnaires, and the data are used to analyze the testers’ psychological fatigue.

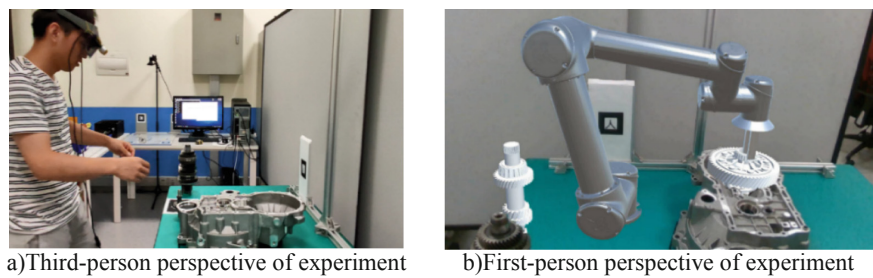


Fig. 17. HRC assembly experiment

5.2 Results Analysis

Virtual Engineering Simulation Platform (VESP) [13] is a virtual simulation system developed on the basis of OpenSceneGraph. In this paper, Kinect is used to collect the real testers' data, and these data are used to drive the virtual human movement. The posture of the virtual human is evaluated by RULA through the human factor evaluation module in VESP. As shown in Fig. 18, the picture a) shows a virtual human with corresponding posture after collecting a real tester's data and mapping to this virtual human, and the picture b) shows the RULA results obtained from VESP.

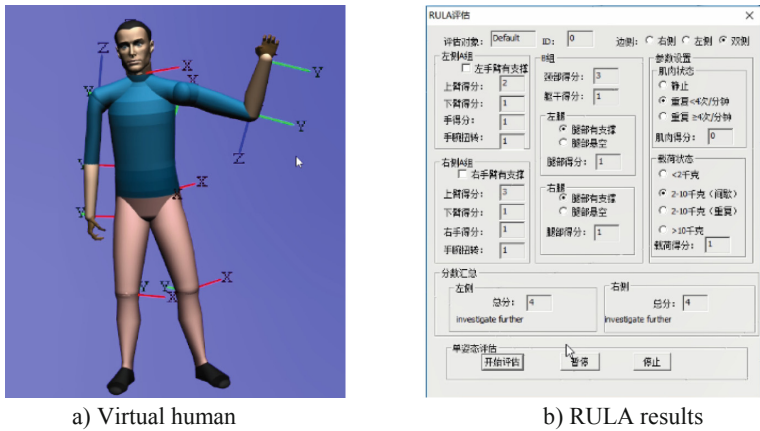


Fig. 18. Analysis of psychological fatigue in VESP

Figure 19 shows the distribution quantity of the two assembly schemes in each scoring interval of RULA. It can be seen that scheme 1 has a larger number of scores in the high score interval, which indicates that the HRC assembly scheme 1 is not reasonable enough and needs to be changed. In contrast, scheme 2 is more reasonable. The average RULA score of scheme 1 is 2.7, close to 3, which belongs to evaluation level II. It means further investigation and research are needed and possible modification may need. The average RULA score of scheme 2 is 2.08, close to 2, which belongs to evaluation level I and is in the acceptable range.

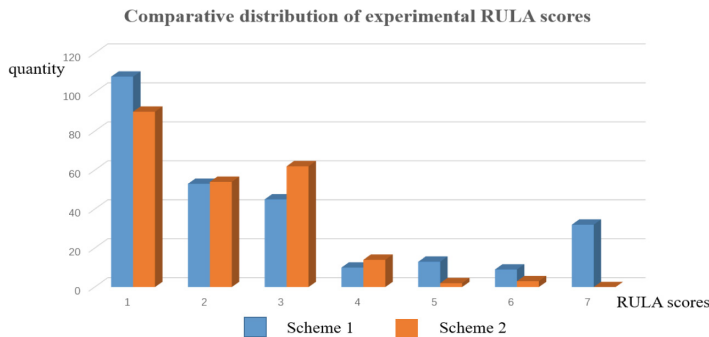


Fig. 19. RULA score comparison

Figure 20 shows the scores of each dimension in NASA-TLX of the two schemes. It can be seen that the score of scheme 2 is lower than that of scheme 1 in all dimensions, which indicates that the task allocation of scheme 2 enables testers to complete tasks more comfortably, reduces the psychological load level in all dimensions. The result of NASA-TLX is consistent with the result of RULA score.

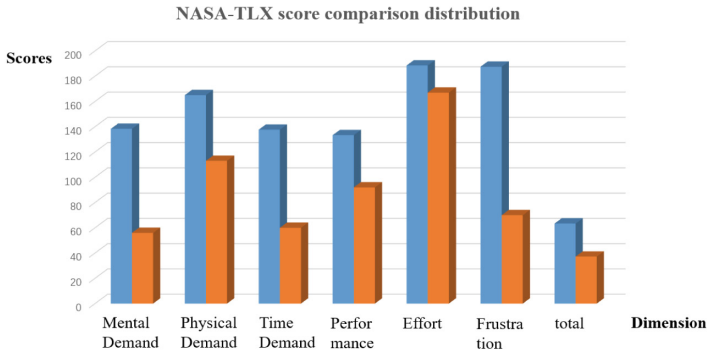


Fig. 20. NASA-TLX score comparison

Compared with the difference between two schemes, scheme 2 isolates the tasks of human and robot. There are no frequent tasks switching between human and robot, which reduces the memory demand of human and frequent contacts with robot, so it can get a better evaluation score. Therefore, it is necessary to avoid frequent switching between human and robot assembly tasks, otherwise it will cause unnecessary psychological burden to the operator.

6 Conclusions

Aiming at the lack of simulation links in the design and transformation of HRC workstation, a method of building a simulation system for HRC workstation based on AR is proposed. A hand-eye interaction model of virtual human is presented, which is realized through AR glasses, LeapMotion and Kinect sensors. These sensors are used for environmental information perception and gesture recognition to complete the natural interaction between human and robot.

Take the assembly workstation transformation of an automobile gearbox as an example, an HRC assembly simulation prototype system including software and hardware is built. RULA by digital human model and NASA-TLX questionnaire are used for testers' data analysis. With human factors analysis, the advantages and disadvantages of each HRC assembly schemes are easily obtained. The results show that the simulation system can be used for the evaluation of human factors of different HRC tasks configuration planning.

Acknowledgement. The work is partially supported by the NSFC project (51475291) and MIIT project (19GC04252), China. The authors are grateful to the editors and anonymous reviewers for their valuable comments.

References

1. Matsas, E., Vosniakos, G.C.: Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *Int. J. Interact. Des. Manuf. (IJIDeM)* **11**(2), 139–153 (2017). <https://doi.org/10.1007/s12008-015-0259-2>
2. Wang, Q., Cheng, Y., Jiao, W., Johnson, M.T., Zhang, Y.: Virtual reality human-robot collaborative welding: a case study of weaving gas tungsten arc welding. *J. Manuf. Process.* **48**, 210–217 (2019)
3. Neuhofer, J.A., Kausch, B., Schlick, C.M.: Embedded augmented reality training system for dynamic human-robot cooperation. *Nato Research and Technology Organization Neuilly-Sur-Seine, France* (2009)
4. Ni, D., Yew, A.W.W., Ong, S.K., Nee, A.Y.C.: Haptic and visual augmented reality interface for programming welding robots. *Adv. Manuf.* **5**(3), 191–198 (2017). <https://doi.org/10.1007/s40436-017-0184-7>
5. Michalos, G., Karagiannis, P., Makris, S., Tokçalar, Ö., Chryssolouris, G.: Augmented reality (AR) applications for supporting human-robot interactive cooperation. *Procedia CIRP* **41**, 370–375 (2016)
6. Crick, C., Jay, G., Osentoski, S., Pitzer, B., Jenkins, O.C.: Rosbridge: ROS for Non-ROS users. In: Christensen, H.I., Khatib, O. (eds.) *Robotics Research. STAR*, vol. 100, pp. 493–504. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-29363-9_28
7. Azuma, R.T.: A survey of augmented reality. *Presence: Teleoper. Virtual Environ.* **6**(4), 355–385 (1997)
8. Plopski, A., Moser, K.R., Kiyokawa, K., Swan, J.E., Takemura, H.: Spatial consistency perception in optical and video see-through head-mounted augmentations. In: *2016 IEEE Virtual Reality (VR)*, pp. 265–266. IEEE (2016)
9. Zhu, W., Fan, X., Zhang, Y.: Applications and research trends of digital human models in the manufacturing industry. *Virtual Real. Intell. Hardw.* **1**(6), 558–579 (2019)
10. Hornung, A., Wurm, K.M., Bennewitz, M., Stachniss, C., Burgard, W.: OctoMap: an efficient probabilistic 3D mapping framework based on octrees. *Auton. Robots* **34**(3), 189–206 (2013). <https://doi.org/10.1007/s10514-012-9321-0>
11. Nguyen, V.T.: Enhancing touchless interaction with the leap motion using a haptic glove. *University of Eastern Finland, Joensuu* (2014)
12. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: *Advances in Psychology*, North-Holland, vol. 52, pp. 139–183 (1988)
13. Qiu, S., Fan, X., Wu, D., He, Q., Zhou, D.: Virtual human modeling for interactive assembly and disassembly operation in virtual reality environment. *Int. J. Adv. Manuf. Technol.* **69**(9–12), 2355–2372 (2013). <https://doi.org/10.1007/s00170-013-5207-3>