Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 93 (2020) 1188–1193

www.elsevier.com/locate/procedia

53rd CIRP Conference on Manufacturing Systems

Symbiotic human-robot collaboration: multimodal control using function blocks

Sichao Liu, Lihui Wang*, Xi Vincent Wang

Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

* Corresponding author. *E-mail address:* lihui.wang@iip.kth.se

Abstract

Complex assembly tasks require increased flexibility and adaptability, as well as higher effort on the conventional (re)programing of robots. To solve such challenges, this paper presents a function block-enabled multimodal control scheme for symbiotic human-robot collaborative assembly. Data/event-driven function blocks with smart decision algorithms are used for human-centred robot control with multimodal fusion. Then, multimodal control commands in the form of haptics, gesture and voice are defined as the inputs of the function blocks to trigger task execution. This novel scheme facilitates the implementation of the multimodal symbiotic human-robot collaborative assembly with enhanced stability and flexibility.

Keywords: Robot, Assembly, Multimodal control, Function block, Human-robot collaboration

1. Introduction

Human-robot collaboration (HRC) in the context of manufacturing allows a human operator work side by side with a robot in close proximity, in which the accuracy, strength and repeatability of the robot are combined with the flexibility and adaptability of the human operator to realise a symbiotic HRC manufacturing for higher overall productivity [1]. Compared with the traditional manufacturing paradigms, HRC empowers a manufacturing paradigm towards a high level of flexibility, adaptability and controllability. In recent years, research efforts on HRC have been numerous, and application-oriented HRC has been reported in many industrial fields [2][3]. The research of utilising robots working as collaborating partners in assembly lines in an unstructured environment has been active recently [4][5]. However, the implementation of complex or customised assembly tasks requires increased flexibility and adaptability for much-improved assembly performance and better product quality, as compared with fully automated assembly and purely manual assembly [6]. Today's robots used in HRC assembly are mostly controlled by rigid native codes and programme-based pre-planned tasks are defined. Once an assembly process/operation is changed, the pre-planned tasks need to be re-programmed due to the dynamic environment of HRC. This not only mostly requires much higher effort on the offline/online programming but lacks the adaptability in response to any uncertainty, which can no longer support symbiotic HRC.

Multimodal intuitive programming offers opportunities for robot programming and control without specialised expert knowledge in which diverse communication channels, such as voice, gesture, haptic interaction, and even brainwave, between a human and a robot are used [1]. In the last decade, much attention was focused on multimodal HRC, especially on multimodal robot control. It allows human operators to control robots intuitively and naturally by using voice and gesture commands, haptic interaction, and human thoughts in the form of human brainwaves [7]. Due to the limitation of robots' interface design, combination of these multimodal control commands used for intuitive control of the robot turned out to

10.1016/j.procir.2020.03.022

be a bottleneck, where typical difficulty exists in integrating these diverse communication and control commands with adaptive responses. Additionally, HRC assembly is in a dynamic environment, adaptive responses to commands and correct trigger of the relevant control blocks is also a typical challenge for symbiotic HRC assembly.

Within the context, this paper proposed a framework of function block-based multimodal intuitive control for HRC assembly, which is a novel solution to the need of the flexibility and adaptability. Function blocks (FBs) are used for encapsulation of tasks/processes and transferring them to the controller level of machines for execution. Then, multimodal intuitive robot control can be facilitated by the use of event/data-driven function blocks with smart algorithms embedded for adaptive robot control.

The rest of the paper is organised as follows. Section 2 presents related work of human-robot collaboration, function blocks and multimodal robot control. A system design of FB-based multimodal control for symbiotic HRC assembly is presented in Section 3. Section 4 introduces multimodal intuitive control driven by haptic interaction instructions, voice, body and hand motion commands. Task-oriented function blocks for HRC assembly are presented in Section 5, followed by the conceptual implementation architecture in Section 6. Finally, conclusions and future work are given in Section 7.

2. Related work

2.1. Human-robot collaboration

Based on the classification of human-robot relationship, HRC is defined as a *'state in which a purposely designed robot system and an operator work on simultaneous tasks within a collaborative environment'* [8]. The goal is to enable close collaboration between humans and robots, in all service and industrial tasks that require the adaptability of humans to be merged with the high performance of robots. Many academic and industrial studies on HRC have been successfully carried out in recent years. A number of factors placed HRC in the spotlight and they are summarised as the combined strength of the robots and humans, collaborative and shared working conditions, increased robustness in the control performances and higher degree of resilience, and improved ergonomics [1].

HRC in manufacturing has been widely discussed in a number of applications (i.e. in automotive manufacturing [9], in cellular manufacturing [10], and in hybrid manufacturing cell [11]). In any HRC system, human safety is of paramount importance [12] and the related approaches and safe strategies were reported in [13]. HRC is also applied as the solution to meet the needs in advanced collaborative manufacturing systems of the future industry. In such a manufacturing context, topics related to HRC in manufacturing covered from communication and data fusion, task planning, adaptive robot control to reconfigurable HRC framework design have also been actively discussed [1][14]. In addition, the modelling of human operators working as a collaborating partner in HRC was widely investigated, e.g., the assessment of worker skills [15] and psychophysiological human studies [16].

2.2. Function blocks

Function block [17], as an International Electrotechnical Commission (IEC) 61499 standard, is used for distributed industrial process and control systems. Each FB is a control software unit containing complex algorithms and an execution control chart (ECC), with inputs and outputs of data and events [18]. When triggered by event inputs, algorithms embedded in FBs are invoked to perform the tasks and control commands. An ECC would be used for representing the EC states, transitions and actions of that function block instance. An internal finite state machine can be used for controlling the different states and transition of subtasks within the function blocks. After the completion of a task, a function block's output event is generated as a notification so that a subsequent task or function block can be triggered.

Research of the use and implementation of FBs in manufacturing systems has been active in recent decades, and the use of event/data-driven FBs can be found in adaptive process planning, execution control of manufacturing devices, and dynamic process decision making [19]. In this context, FBs can be viewed as an enabler to encapsulate control and task algorithms, monitor a process plan during execution, and control machining jobs and equipment with enhanced adaptability [20].

Current HRC assembly is characterised by complex assembly demands and dynamic changes of assembly tasks and operations; applying FBs to the control of robots and assembly plans can give them a higher level of flexibility and autonomy to better handle the changes due to their abilities in reuse and modularity [18].

2.3. Multimodal intuitive control

The higher effort of conventional programming of robots in the context of manufacturing motivates robotic engineers and researchers to develop robotic programming and control tools with a higher degree of adaptability [1]. Diverse programming methods have been developed to efficiently and effectively control industrial robots [21].

Despite the advancement of the programming approaches developed, these programming and control approaches cannot satisfy the high demands of HRC in which adaptive and intuitive robot control, dynamic task planning and execution, smart decision-making to unpredicted changes [22], and high overall productivity are expected to be performed. These factors enable HRC to be controlled by a new means that is both symbiotic and multimodal to facilitate any changes during collaboration [23].

Voice commands, as an effective tool for human-human communication, have been extensively applied to robot intuitive robot control [24]. However, the noisy background in the context of manufacturing can affect the accuracy of recognising voice commands. Therefore, many researchers have explored the possibility to adopt nonverbal/context-dependent communication channels used for robot intuitive control and programming [25][26]. Additionally, the emerge and use of lightweight collaborative robots makes it possible for a human operator to haptically control the robots [1].

However, a typical challenge of intuitive control in HRC is to fuse these diverse control modes in an integrated system.

3. System design

The developed system provides the ability for a human operator to control an industrial robot using multimodal intuitive programming and function blocks with embedded smart algorithms. Such a collaboration aims to combine the accuracy and repeatability of the robot with the flexibility of the human operator. Meanwhile, function blocks and multimodal robot control are adopted to enhance the control performance with less effort and handle changes to the pre-planned tasks in the HRC assembly.

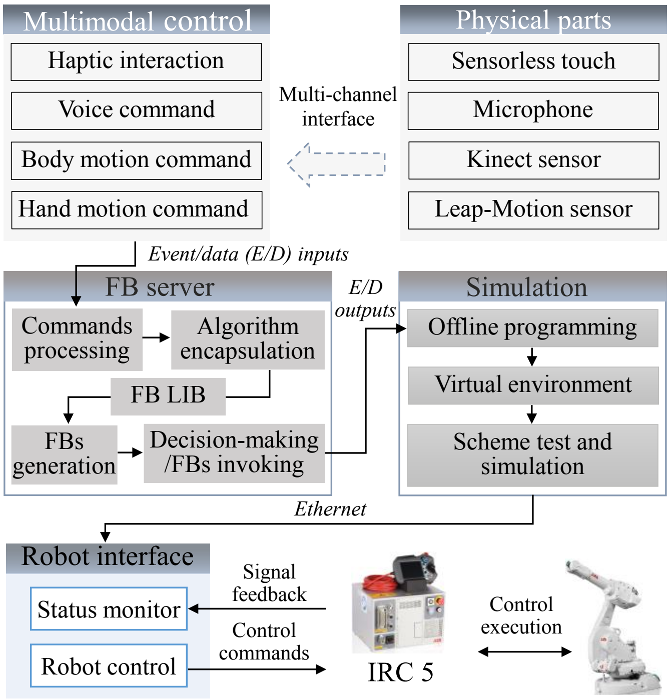

Fig. 1. System design of FB-based multimodal control for symbiotic HRC

The proposed multimodal control architecture for symbiotic HRC assembly is driven by microphone-based voice instructions, Leap Motion-based hand motion recognition, depth sensor-based body motion commands, and sensorless haptic interaction. As shown in Fig. 1, human instructions used for robot control and task execution are collected from diverse communication channels and these commands are defined as input events of function blocks embedded with algorithms for processing and handling these commands. When triggered by event inputs, the function block will start handling and recognising these control commands. Feature/process-based assembly planning and execution control are encapsulated into function blocks. For a given assembly operation, different function blocks can be chosen and grouped together according to the control commands. These function blocks are invoked in a specified sequence once triggered by the coming signals, and the commands are sent via a robot interface to control the robot and perform the assembly tasks. In the case of a change to the pre-planned tasks, related function blocks will be triggered to respond the change for adaptive assembly task execution. A user interface is designed to allow a human operator to enable the control commands, observe the robot state, and simulate and test the schemes in a virtual environment. Here, the communication between the human-machine interface and a robot controller is built based on Ethernet, while the interface among the built function blocks is developed. In addition, depth sensor-based body motion recognition is used to monitor the human and ensure the human safety. Moreover, an operator can give a voice command to instruct a robot to move, then haptic interaction is responsible for accurate motion control of the robot. Finally, Leap Motion-based human motion is used to control '*open/close*' of the robot's gripper.

4. Multimodal intuitive control

The framework of intuitive robot control for symbiotic HRC is devised, aiming for more natural HRC and intuitive control using multimodal communication channels between a human and a robot, as shown in Fig. 2. Within the framework, four communication modalities are used to control the robot and execute the operations of assembly tasks, which are haptic interaction, voice instructions, body motion commands and hand motion commands.

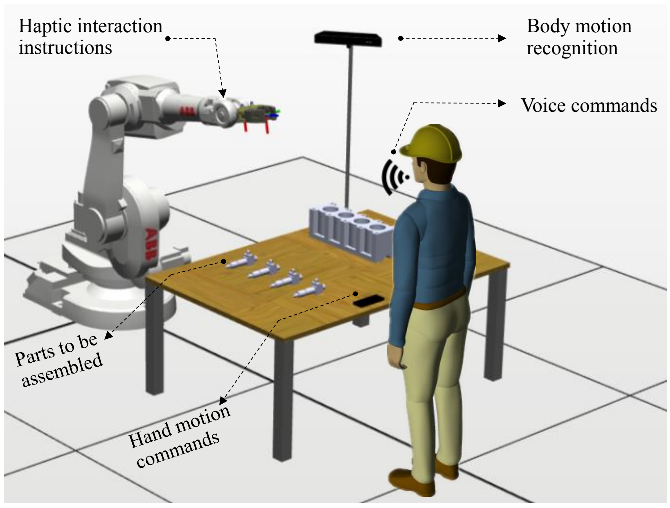

Fig. 2. Multimodal intuitive robot control for symbiotic HRC

A brief description of multi-modality recognition is introduced as follows.

Convolutional Neural Network (CNN) for voice recognition: CNN is a popular deep learning model that can be adopted for sequential data recognition tasks such as speech command recognition. Before using CNN to process the voice commands, the voice command dataset is transformed into 2-dimensional spectrograms by Fast Fourier transform [27]. The convolution operation is formulated as a function of input map and feature map in which a *ReLU* activation function is adopted [28]. In the formulated convolution function, the weights of the input map are shared among each convolution neuron, and Max-pooling outputs the maximum value of each of the local neighbour. Max-pooling makes each feature map invariant to local translations in the input map, which is also proven to be useful in CNN [29]. In model training, the categorical cross-entropy is defined as the cost function. The voice commands used in the multimodal HRC are classified and labelled as *left*, *right*, *on*, *off*, *up*, and *down*. The recorded audio that is essentially a 1-D vector of strength signals is transformed into a 2-D matrix that can be treated as a single-channel image.

Leap-Motion-based hand motion recognition: The Leap Motion can be connected with the mobile device (i.e. personal computer) for capturing hand motion and the data output of Leap motion offers a real-time representation of human hands with a series of timestamps, the finger positions, and hand position. To implement hand motion recognition in HRC, the Leap motion dataset built based on a number of hand-waving are labelled and multi-category classification is developed, while other random motion datasets collected from the sensors are labelled as negative. The algorithm that is to identify the hand-waving motion is used to train the labelled dataset. The hand motion dataset consists of a number of sequences of hand motions with six different categorical labels. The Leap Motion Controller [30] captures the direction and orientation of key hand joints and bones frequency of 100 Hz. The hand skeleton models can be built by tracking and capturing hand bones and joints. In each time stamp of the hand motion, 64 hand motion features which are defined by the parameters from the skeletons are captured.

Human body posture recognition: transfer learning can be defined by a domain, a task, a learning source, and a target source where a domain consists of a feature space and a marginal probability distribution. When the domain is given, the task can be represented by a label space and a predictive function which can be learned from the training dataset. In the specific case of body motion recognition, a function is trained from the source domain and the source task with a large amount of labelled images. Since the source domain is in image format and the target domain is in video format, the data of the target domain can be sampled as sequences of images where the pre-trained network can act as a generic feature extractor to transfer the knowledge represented by parameters learned from the source domain. After the transfer learning-enabled feature extraction, the representation of the training dataset can be denoted as a function of parameters transferred from the target predictive function and the training samples from the body motion dataset. The network is further trained by minimising the cross-entropy loss, before being fed into a softmax layer for nomalisation. To facilitate the training of the transfer learning, the collected video clips are sampled into sequences of images where the features are extracted from the image sequences by Inception-v3 pre-trained model [31]. The processed image sequences are stored and prepared as the input data for training model.

Sensorless haptic interaction: A sensorless haptic control strategy is developed to control human-robot interaction, which incorporates HRC module, kinematics and dynamics module, and robotic control module. In HRC module, a sensorless haptic command enables a human operator to control an industrial robot by hands in a dynamic, and unconstructed environment through a human-machine interface. Then, a kinematics module is used for the transfer between the joint positions in the joint space and the end-effector's positions in the Cartesian space, and a dynamics module is built to obtain the control output. The external force/torque for a robotic system is defined as the difference of output force/torque of the dynamic model of the robot and the feedback data of the joint torque acquired from the robot controller. To haptically control the robot to move along a given direction smoothly, an admittance controller with parameter adaptation is designed to transfer the detected external force into the reference velocity and the reference position, and the damping and the moment of inertia of the controller can be adjusted to adapt the dynamic change of the force applied by a human operator and control the robot move smoothly.

5. Task-oriented function blocks for HRC assembly

As mentioned earlier, function blocks are a programming specification that applies an event/data-driven model for distributed adaptive control. One potential application of function blocks in the HRC assembly is multimodal robot control. Fig. 3 shows a basic FB and a composite FB together with the external interface and internal behaviour of each type of the FB. A basic FB is responsible for defining the fundamental functional relationship of the event input and output, and can code APIs using programming language. The algorithms used for control and task planning can be integrated into a basic FB and can only be accessed by the FB itself. A composite FB consisting of several basic FBs and/or other composite FBs does not encapsulate any algorithms inside itself but invoke the internal basic FBs for control execution when triggered by events and data. For a basic FB labelled by a type name and instance, the input events/data trigger the function call and an internal event/data connection, and then the embedded algorithms in the FB can be controlled through its execution control chart. The internal variables and algorithms can be tuned to match the environmental conditions. Finally, the running results are notified by the event output and expressed by the data output.

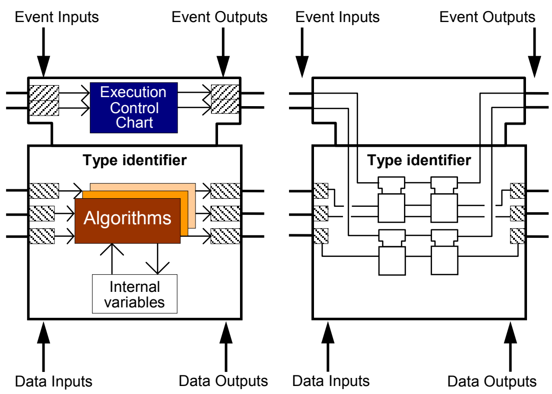

Fig. 3. Structure of function block (basic: left; and composite: right)

In this section, FBs are used to encapsulate multimodal robot control algorithms proposed in Section 4 and enable a human operator to activate and handle control instructions. Modular and reusable FBs can achieve adaptive control for these commands such as a free switch of different control modes and immediate termination of control execution. These functional characters ensure the individual encapsulation and flexible invocation of multimodal commands. Additionally, the assembly tasks that are comprised of a series of assembly operations can be decomposed into a set of assembly features and handled by FBs for the ease of robotic control. The mating relationships in assembly can be defined as basic assembly features (e.g., picking up, placing, inserting, and screwing) that can be mapped to basic FBs.

Thus, FBs used for symbiotic HRC assembly can be categorised into two types, one for multimodal robot control, and another for assembly operations, as shown in Fig. 4. For FB-based multimodal robot control, control algorithms for sensorless haptic interaction, hand motion recognition, voice processing, and body motion recognition that are introduced in Section 4 are encapsulated into composite FBs, and defined as *Com_HI*_FB, *Com_HM*_FB, *Com_VP*_FB, and *Com_BM*_FB, respectively. These four types of composite FBs are explained as follows. *Com_VP*_FB is used to instruct the robot to move along a given direction according to a subset of voice commands; *Com_HI*_FB is responsible for haptically controlling the robot to perform an accurate motion; *Com_BM*_FB is used to monitor the collaborative workspace of a human and a robot in which active collision detection and avoidance is adopted to ensure human safety. *Com_HM*_FB is for tracking, and then recognising human hand motion, which is defined as a signal for gripper control. In each composite FB, a *Sig_Proc*_FB (a basic FB) is built for identifying human operator's control instructions, and an *Init_M*_FB for initialising the function block.

To facilitate assembly planning, varying types of basic FBs for HRC assembly are defined: *Start*_FB for triggering the start of an assembly task; *Init*_FB for initialising events/data of function blocks and the preparation of HRC assembly; and *Place/Pick*_AF_FBs for placing/picking up a part. Based on the above basic FBs, *Com_Assemble*_AF_FB (a composite FB) is designed to handle the collaborative operations of the assembly (i.e. inserting and holding) while performing screwing by a human operator.

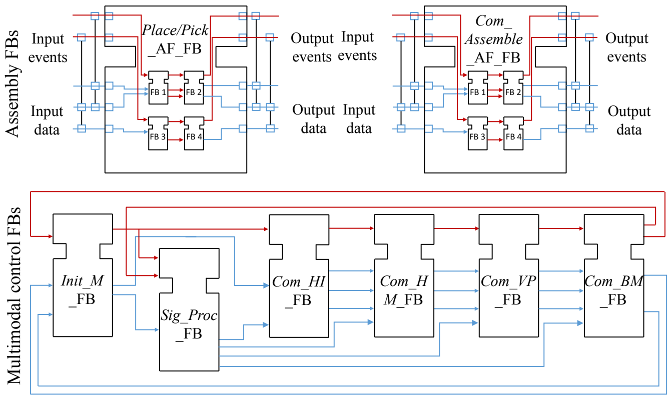

Fig. 4. FBs used for symbiotic HRC assembly

6. Implementation procedures

In this section, the implementation procedure of FB-based multimodal control for HRC assembly is presented. ROS® (Robot Operating System) is selected as a middleware between control algorithms developed for multimodal interaction and assembly task execution and the robot controller. Through the use of ROS, the robot controller can communicate and access the embedded decision algorithms that are developed in a Java environment, and control commands can be sent to the controller for robot control and task execution. For the design and configuration of FBs, an in-house web-based FB IDE (integrated development environment) is used, which allows to define and design basic assembly feature function blocks. To demonstrate the process of HRC assembly, RobotStudio® is selected and used as a simulation tool in which a virtual experimental setup is designed and the path planning of the robot in HRC assembly can be performed. The functionalities embedded in the function blocks for HRC assembly can be then simulated to verify the control performance.

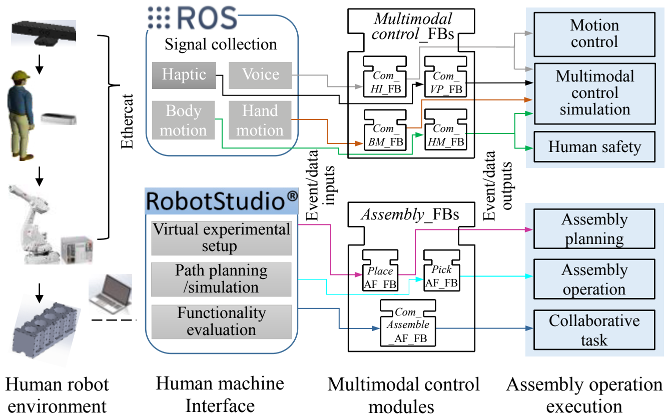

Fig. 5. A conceptual implementation architecture of symbiotic HRC assembly driven by FB-based multimodal commands

A conceptual implementation architecture for symbiotic HRC assembly is developed to illustrate FB-based multimodal control, as shown in Fig. 5, in which an ABB robot is used to perform the assembly operations. The assembly scenario is to assemble a fuel injector tube into a hole of an engine. The functionalities and invocation logic of the function blocks and sequences of assembly operations are demonstrated as follows. Assembly begins when the *Start*_FB in a composite FB *Com_BM* _FB receives a triggering event which can be initiated by a human operator via a user interface. The *Init*_FB is then triggered by the output event from the *Sig_Proc*_FB. The ECC of the *Init*_FB coordinates the execution of the *Com_BM*_FB for monitoring the human operator and the robot in the workspace. Once a possible collision is detected, an active collision avoidance strategy is applied by a function block in the *Com_BM*_FB. Then, the *Com_VP*_FB is activated to give the instruction '*down*', and the robot controller executes '*down*' motion at a safe speed from the '*Home*' position. When the end-effector approaches the target position, the *Com_HI*_FB is triggered to enable a human operator to guide the robot's end-effector for accurate positioning. Then, *Place/Pick*_AF_FB is responsible for performing part placement in which the *Com_HM*_FB is used to recognise the human hand and control the griper '*open*' and then grasp the fuel injector tube with a '*close*' hand motion. When moving towards the hole of the engine after picking up the part, the *Com_VP*_FB and *Com_HI*_FB are invoked to process voice commands and handle haptic interaction for motion control, respectively. The *Com_Assemble*_AF_FB is invoked to instruct the robot to hold the part while a human operator fastens screws. Finally, the *Com_HM*_FB and *Com_HI*_FB are used to open the gripper and perform '*Home*' operation followed by the termination of the safety monitoring (*Com_BM*_FB). In this research, the respondence time of the proposed assembly scenario in this research can be up to a level of seconds on average, which is affected by the programmes for robot control and computation, algorithms embedded in function blocks, and hardware (i.e. robot controllers and communication protocol). The visualisation of the sequences of the operations and the performance analysis of the proposed scheme in terms of the enhanced flexibility and adaptability are out of the scope. In our future work, a benchmark test will be performed to analyse the performance of the system.

7. Conclusions

This paper proposed a systematic framework for intuitive robot control in HRC assembly. Within the context, a multimodal intuitive control strategy driven by voice and gesture commands, haptic interaction, and hand motion instructions are developed to perform programme-free adaptive robot control. The runtime decision-making and adaptive capacity of the HRC assembly can be realised by integrating the event/data-driven function blocks into the robot control and assembly process plan for task execution, where the embedded knowledge and algorithms are triggered by the corresponding input event/data. This paper focuses on a conceptual design of the multimodal robot control for HRC assembly using the combined strength of the multimodal command recognition and function blocks, which can be applied as a novel solution for multimodal robot control in industrial practices. The current system demonstrated using a simple part assembly in this paper is for proof-of-concept only.

Our future work will focus on the performance analysis of the proposed multimodal control approach and experimental demonstrations of a real robot assembling complex products with humans. The dynamic assembly planning driven by assembly feature and the modelling of human operators in symbiotic HRC assembly will also be investigated.

References

[1] Wang L, Gao R, Váncza J, Krüger J, Wang X V., Makris S, Chryssolouris G (2019) Symbiotic human-robot collaborative assembly. *CIRP Ann* 68(2):701–726.

[2] Vincent X, Kemény Z, Váncza J, Wang L (2017) CIRP Annals - Manufacturing Technology Human – robot collaborative assembly in cyber-physical production : Classi fi cation framework and implementation. 66:5–8.

[3] Vanderborght B (2019) Unlocking the potential of industrial human – robot collaboration for economy and society. doi: 10.2777/568116

[4] Krüger J, Lien TK, Verl A (2009) Cooperation of human and machines in assembly lines. *CIRP Ann* 58(2):628–646.

[5] EU project: SYMBIO-TIC. http://www.symbio-tic.eu/.

[6] Ji W, Yin S, Wang L (2018) A Virtual Training Based Programming-Free Automatic Assembly Approach for Future Industry. *IEEE Access* 6:43865–43873.

[7] Liu H, Fang T, Zhou T, Wang L (2018) Towards Robust Human-Robot Collaborative Manufacturing: Multimodal Fusion. *IEEE Access* 6:74762–74771.

[8] ISO 10218-1:2011 Robots and robotic devices — Safety requirements for industrial robots — Part 1: Robots.

[9] Elena R, Brian A (2008) Levels of Human and Robot Collaboration for Automotive Manufacturing. *Perform. Metrics Intell. Syst. Work.*

[10] Tan JTC, Duan F, Zhang Y, Watanabe K, Kato R, Arai T (2009) Human-robot collaboration in cellular manufacturing: Design and development. *2009 IEEE/RSJ Int Conf Intell Robot Syst IROS 2009* 29–34.

[11] Sadrfaridpour B, Saeidi H, Wang Y (2016) An integrated framework for human-robot collaborative assembly in hybrid manufacturing cells. *IEEE Int Conf Autom Sci Eng* 2016-Novem:462–467.

[12] Michalos G, Makris S, Tsarouchi P, Guasch T, Kontovrakis D, Chryssolouris G (2015) Design considerations for safe human-robot collaborative workplaces. *Procedia CIRP* 37:248–253.

[13] Michalos G, Kousi N, Karagiannis P, Gkournelos C, Dimoulas K, Koukas S, Mparis K, Papavasileiou A, Makris S (2018) Seamless human robot collaborative assembly – An automotive case study. *Mechatronics* 55:194–211.

[14] Kim W, Lorenzini M, Balatti P, et al (2019) Adaptable Workstations for Human-Robot Collaboration: A Reconfigurable Framework for Improving Worker Ergonomics and Productivity. *IEEE Robot Autom Mag* 26(3):14–26.

[15] Abujelala M, Gupta S, Makedon F (2018) A collaborative assembly task to assess worker skills in robot manufacturing environments. *ACM Int Conf Proceeding Ser* 118–119.

[16] Arai T, Kato R, Fujita M (2010) Assessment of operator stress induced by robot collaboration in assembly. *CIRP Ann* 59(1):5–8.

[17] International Electrotechnical Commission, 2005. International Standard of Function Blocks – Part 1: Architecture, IEC 61499, p. 1-111.

[18] Wang L, Haghighi A (2016) Combined strength of holons, agents and function blocks in cyber-physical systems. *J Manuf Syst* 40:25–34.

[19] Wang L, Keshavarzmanesh S, Feng H-Y (2008) Design of adaptive function blocks for dynamic assembly planning and control. *J Manuf Syst* 27(1):45–51.

[20] Wang L, Adamson G, Holm M, Moore P (2012) A review of function blocks for process planning and control of manufacturing equipment. *J Manuf Syst* 31(3):269–279.

[21] Pan Z, Polden J, Larkin N, Van Duin S, Norrish J (2010) Recent progress on programming methods for industrial robots. *Jt 41st Int Symp Robot 6th Ger Conf Robot 2010, ISR/ROBOTIK 2010* 1:619–626.

[22] Iba S, Paredis CJJ, Khosla PK (2005) Interactive multimodal robot programming. *Int J Rob Res* 24(1):83–104.

[23] Perzanowski D, Schultz AC, Adams W, Marsh E, Bugajska M (2001) Building a multimodal human-robot interface. *IEEE Intell Syst Their Appl* 16(1):16–21.

[24] Zhang H, McLoughlin I, Song Y (2015) Robust sound event recognition using convolutional neural networks. *ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc* 2015-Augus:559–563.

[25] Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y (2009) What is the best multi-stage architecture for object recognition? *Proc IEEE Int Conf Comput Vis* 2146–2153.

[26] Kardos C, Kemény Z, Kovács A, Pataki BE, Váncza J (2018) Context-dependent multimodal communication in human-robot collaboration. *Procedia CIRP* 72:15–20.

[27] Akan B, Ameri A, Cürüklü B, Asplund L (2011) Intuitive industrial robot programming through incremental multimodal language and augmented reality. *Proc - IEEE Int Conf Robot Autom* 3934–3939.

[28] Pavlovic VI, Sharma R, Huang TS (1997) Visual interpretation of hand gestures for human-computer interaction A review. *IEEE Trans Pattern Anal Mach Intell* 19(7):677–695.

[29] Gonzalez TF (2007) ImageNet Classification with Deep Convolutional Neural Networks. *Handb Approx Algorithms Metaheuristics* 1–1432.

[30] Weichert F, Bachmann D, Rudak B, Fisseler D (2013) Analysis of the accuracy and robustness of the Leap Motion Controller. *Sensors (Switzerland)* 13(5):6380–6393.

[31] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) DeCAF: A deep convolutional activation feature for generic visual recognition. 31st Int. Conf. Mach. Learn. ICML 2014. , pp 988–996.

This is a resupply of March 2023 as the template used in the publication of the original article contained errors. The content of the article has remained unaffected.