中國機械工程學刊第四十一卷第六期第 755~762 頁(民國一百零九年)
Journal of the Chinese Society of Mechanical Engineers, Vol.41, No.6, pp.755~762 (2020)

Implementation of a Human-Robot Collaboration System Based On Smart Hand Gesture and Speech Recognition

Shang-Liang Chen *, Li-Wu Huang **, Chung-Chi Huang***, Feng-Chi Lee****, Ho-Chuan Huang***** and Chien-Yu Chen******

Keywords：HRC, deep learning, hand gesture recognition, RFID, ASR.

ABSTRACT

With the development of artificial intelligence technology, the robot arm will be applied in more fields, undertake more work, and become an important assistant of humans. Especially in the manufacturing industry, if a perfect human-robot cooperation (HRC) model can be developed, in which the engineer is responsible for complex operations and monitoring the production process and the robot arm is responsible for highly repetitive and tiring work, the flexibility of the manufacturing process can be improved, the risk of personnel injury can be reduced, and the production line productivity can be improved. To realize the human-robot collaboration system that can be applied in various fields, the technologies of robot arm control, human-robot interaction (HRI), Radio Frequency Identification, machine-to-machine communication, and automatic speech recognition (ASR), etc. are combined in this research. Besides, the HRC in intelligent manufacturing is taken as the demonstration result of this research. In this application example, the human-robot interaction function is realized through smart hand gesture and speech recognition technology, which enables the user to communicate commands to the robot arm most intuitively and easily. The robot arm can execute specific actions, provide processing prompts, or read corresponding motion control command macros according to different commands, to realize various processing contents required by users.

Paper Received September, 2020. Revised November, 2020. Accepted December, 2020. Author for Correspondence:Shang-Liang Chen.

* *Professor, Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan, 70101, R.O.C.*

** *Graduate Student, Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan, 70101, R.O.C.*

*** *Associate Professor, Department of Automation and Control Engineering, Far East University, Tainan, 74448, R.O.C.*

**** *Research Fellow, Department of Mechatronics Control, Industrial Technology Research Institute, Hsinchu, 310, R.O.C.*

***** *Associate Professor, Department of Intelligent Commerce, National Kaohsiung University of Applied Sciences, Kaohsiung, 824, R.O.C.*

****** *Graduate Student, Computer Science and Information Engineering, National University of Tainan, Tainan, 70005, R.O.C.*

INTRODCTION

Human-robot collaboration is a way for humans and machines to communicate to improve production efficiency of manufacturing processes. A well-designed human-machine collaboration mode can achieve advantages such as shortening the work time, improving work accuracy, saving labor costs, and improving product quality. Among them, if hand gesture and human speech recognition technologies are used in the HRC system, the benefits of the system in many different applications will be enhanced.

Although the manufacturing industry has been actively pursuing automation technology to improve production efficiency for a long time, the production model of human-robot collaboration is gaining attention because of the demand for various and small amounts of products. In the manufacturing industry, the existing number of industrial robot arms is still far greater than that of collaborative robot arms. Therefore, how to enable industrial robot arms to work safely around people has become an important issue for human-robot collaboration in the manufacturing industry.

M. Bdiwi et al. (2017) proposed four more specific classifications based on the four types of human-robot interaction defined by the International Federation of Robotics (IFR). Among them, Level two is defined as "robots and personnel share a workspace and tasks, but no physical contact". This level can correspond to the sequential

human-machine collaboration proposed by IFR and the Speed and Separation Monitoring (SSM) mode in ISO 10218-2 (2011). The system proposed in this research is designed concerning the above research to ensure the safety of personnel in the process of human-robot collaboration with a collision avoidance module.

Hand gesture and speech are common communication methods in human-computer interaction. The voice command is suitable to control call-and-come service (Oh, Yoo Rhee, et al., 2008), while hand gesture command is more suitable for continuous control (Jain, M., et al., 2019). Among them, hand gestures recognition refers to the computer's interpretation of human hand movements. In order to cooperate with people, robots in the plant need to understand human gestures correctly and perform the specified actions effectively according to the gestures. Therefore, providing a convenient form of hand gesture communication between humans and robots is an important work to realize HRC in the manufacturing industry. There are many ways to collect hand gesture data that the computer can understand, such as through a camera or a glove with devices such as accelerometer or gyroscope. To allow users to communicate with this human-machine collaboration system without wearing additional equipment, computer vision technology combined with deep learning is used in this research to realize hand gesture recognition.

It is a difficult task to implement automatic speech recognition in a manufacturing environment. However, researchers are still committed to giving robots the ability to understand voice commands (Gomez, R, et al., 2015). Speech is not only an indispensable means of communication between people but also an accurate, hands-free, and natural way for people to interact with applications. By using automatic speech recognition technology, spoken commands can complement or even replace mice, keyboards, controllers, and other human-computer interaction devices. Common uses of ASR today include voice dialing, voice navigation, indoor device control, and speech input. When combined with other natural language processing (NLP) technologies such as machine translation and speech synthesizing, ASR can be used to build more complex applications. For example, Peter X. Liu (2005), Subhash P. Rasal (2014) et al. pointed out that it is a better human-computer interaction way for users to control the robot by voice. As a result, the system proposed in this paper combines ASR technology with hand gesture recognition function to enhance the intuitiveness and convenience of the human-robot interaction module. The user of the system can quickly send control commands to the robot through hand gestures, and use speech commands to convey more complete instruction details.

The experiment in this research takes the human-robot collaboration in the manufacturing environment as a case. To make the system functions more complete and practical, it is necessary to add the function of workpiece management and the communication function between various devices and modules. Meng et al. (2018) use RFID technology to present an IoT sensing and networking framework for data integration and ubiquitous access in smart manufacturing, focusing on product identification, data modeling, interphase data integration, and ubiquitous data access. The RFID data collection module of this paper refers to the above-mentioned research and designs a factory workpiece management method based on RFID technology and cloud database.

OPC Unified Architecture (OPC UA) is the transmission protocol for the M2M (Machine to Machine) network developed by OPC Foundation and applied to automation technology. It is the only recommended communication protocol for the communication layer of RAMI4.0 (Reference Architectural Model Industry 4.0) in 2015 (Adolph, et al., 2016; Ghazivakili, et al., 2018). The characteristics of OPC UA include its open-source standards, high scalability, cross-platform use, and standard security models. Therefore, the communication module of this system is designed based on the OPC UA transmission protocol and is responsible for the communication function between the modules of the system.

MOTIVATION AND PURPOSE

In order to make the applications of robot arm more efficient in various fields, the main purpose of this research is to take the advantages of hand gesture recognition and human speech recognition, and to propose a human-robot cooperation system that can be imported into various application environments.

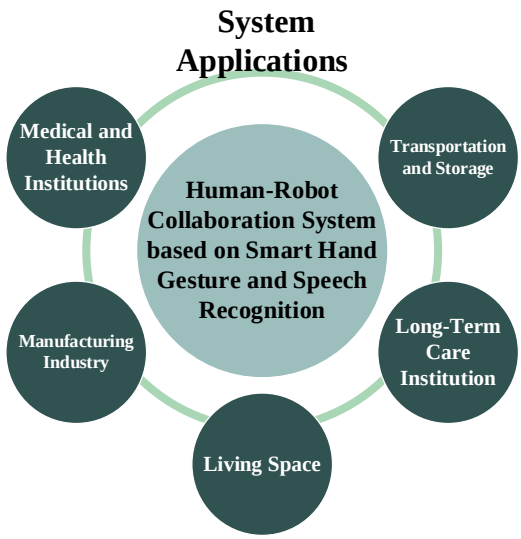

Fig. 1. Potential applications of the system proposed in this research.

For example, in the medical and healthcare field, the use of a human-robot collaboration system to take care of patients will reduce the risk of infection by medical personnel; in the plant of heavy industry, it can be used to assist in handling huge equipment; in the military, it can be used to handle explosives and dangerous items; in the household, it can be used for automatic cleaning and other household duties; and for the long-term care, it can also be used to assist the elderly or the physically and mentally disabled in their daily life.

The hand gesture recognition in this research is realized by computer vision technology combined with image processing and deep learning technology. This method uses a common web camera to collect the user's gesture information, which is more convenient and easier to implement than using gloves with sensors. To overcome the common problems of current hand gesture recognition methods, such as low recognition accuracy, slow identification speed, and poor adaptability to the complex background, the YOLOv4 is adopted by this system as the gesture recognition algorithm. The YOLOv4 algorithm has the advantages of fast recognition speed, outstanding recognition accuracy, light system architecture, high-efficiency algorithm, etc., and can be calculated by using general graphics processor (GPU). These advantages not only reduce the hardware cost but also greatly reduce the difficulty of introducing the system into the actual application environment.

Besides, the speech recognition function of this system is implemented using the System.Speech managed-code namespace provided by the .NET Framework. Users can easily access Microsoft's advanced speech recognition and speech synthesis technologies in Windows operating system by using the System.Speech namespaces. Because speech engines are built into Windows operating system, programmer won’t need to create their own engines for speech recognition, and the tools of the System.Speech can be used to create sophisticated speech applications. In addition, its advantages include that it can recognize voice offline, and it is quite easy to add new voice commands, which can reduce development costs and difficulties.

METHODOLOGY

The system proposed in this study adopts a modular design. Each important function is divided into different modules, and the communication module is responsible for the information exchange between all functional modules so that the system is easy to manage and has the extensibility of functions. The system is designed with the.NET Framework developed by Microsoft and has a human-machine interface for displaying information from the function modules. The methods implemented by each functional module are described below.

Human-robot interaction module

The input of control commands in this research comes from a wired directional microphone and a webcam directed toward the user. For the hand gesture recognition, the processing speed is a very important factor to realize the near real-time image detection in the processing field. Compared with object detection technologies such as Single Shot Detector (SDD), region-based CNN (R-CNN), Fast R-CNN, and Faster R-CNN, YOLOv4 not only had similar detection accuracy but also had better image processing speed. Therefore, this paper chooses YOLOv4, an object detection model based on CNN, to detect the hand gestures.

In this paper, the YOLOv4 model is trained with a customized image data set, in which Tensorflow and Keras framework are used for model training. The data set is sorted out regarding the VOC-2007 standard format proposed by Pascal organization, while the image labeling tool LabelImg is used for image labeling. Considering that depth camera is not often used in the robot control interface in the real manufacturing environment, the hand gesture data collection in this paper is only carried out through a 2D webcam.

During the process of hand gesture image collection, the researcher stood in a controlled laboratory environment and recorded videos about one minute and 40 seconds for each of the eight gestures designed for the system. The researcher changed the position, orientation, and distance from the camera by moving or rotating the wrist from time to time, to obtain a more generalized data set. After recording all the gesture videos, the videos were converted into images at the frequency of 30 frames per second to generate an image data set of about 3,000 images for each category and a total of 24,315 images, which were randomly assigned to test sets and training sets at a ratio of two to eight.

The hand gesture recognition program of this module is shown in Figure 2. The user's real-time video is obtained by the system through a webcam set up in the manufacturing environment. Then, the video will be divided into the frame by frame and inputted to the model for object detection. The system will responsible for analyzing the confidence levels of all gesture categories, returning the position of the bounding box and the category with the highest confidence, and finally displaying the bounding boxes with labels containing category and confidence information in real-time images.

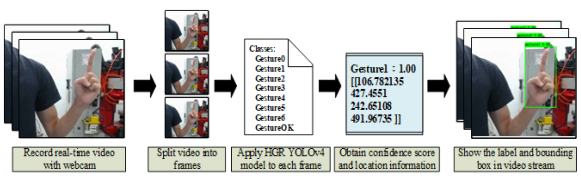

Fig. 2. Hand gesture recognition program of the human-robot interaction module.

The speech recognition function in this research is implemented with the System.Speech namespace provided by the.NET Framework. To ensure the identification accuracy of the speech recognition function meets the application requirements, the program is designed with a pre-established list of all commands to be received. Although the user's flexibility in sending instructions is reduced by this method, the accuracy of identification can indeed be improved. The pointing microphone is used to continuously monitor the processing environment. When the voice information issued by the operator is received by the microphone, the content of the speech will be analyzed and compared with the commands list. If the voice command exists in the list, the system will control the robot arm according to the motion design in the command list.

The flow diagram of human-robot interaction process based on speech recognition is shown in Figure 3. After the recognition module is activated, the processing environment will be monitored continuously by the module through the microphone. If the user's instructions are received, the voice data will be converted into text format through the speech recognition engine, and the text content will be compared with the commands in the preset list. If the instruction is not within the range of the list or the instruction recognition is wrong due to human error in pronunciation, the speech recognition module will display a prompt message asking the user to issue a new instruction. If an instruction that conforms to the list content is received, the module will send the text data of the instruction to the robot arm control module. After the content has been analyzed by the control module, the corresponding robot arm motion control can be performed.

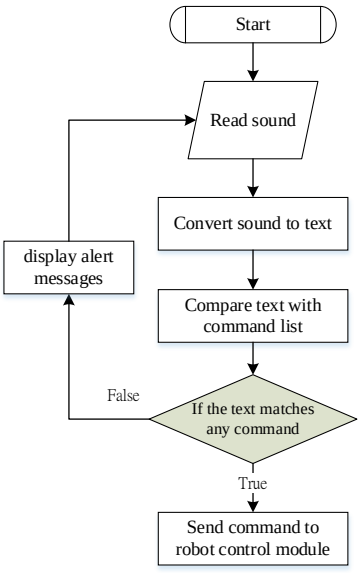

Fig. 3. Flow diagram of human-robot interaction process based on speech recognition.

Machine control module

This module is responsible for logical analysis of the operator's gesture or speech instructions and making various responses to the user's commands. This module contains two control modes: direct control mode and teaching control mode. The user can directly issue a single control command to the robot arm, CNC or AGV, or call a pre-established motion macro file to make the hardware device to perform a complete moving process.

The process flow diagram of this module is shown in Figure 4. When using this module, the user must open the desired control mode with commands at first. The system will analyze whether the user has followed the design procedure to give instructions and turn on the corresponding control mode. In the direct control mode, the user can call various control functions in the motion control library through hand gestures or speech commands, to carry out direct remote control of the robot arm, CNC or AGV. In the process of direct control, the user can also record the coordinate data during the motion process of the devices and save it as commands or output it as a motion macro file. The output macro file can be read by the user in the teaching control mode. In the teaching control mode, the user can call each motion command description of the system which taught and recorded in the list or the macro files through simple hand gesture or speech instructions, so that the devices can execute the complete processing tasks that have been taught automatically.

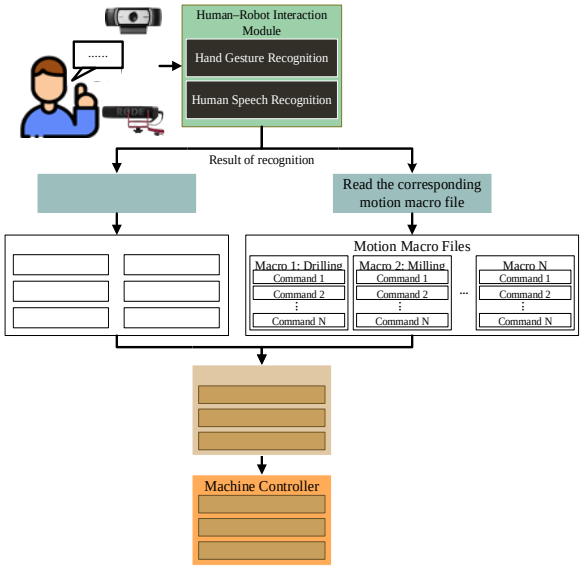

Fig. 4. Process flow diagram of the machine control module.

This machine control module shares several kinds of information with other functional modules. For example, it will provide the joint coordinates of the robot arm to the collision avoidance module so that it can calculate the appropriate operating speed of the robot arm. The calculated result will be read by

the machine control module, and the corresponding robot speed control can be performed. Besides, the module will also read the workpiece information from the RFID data collection module, so that it can perform specified motion commands on the workpiece, and display processing prompt information to the engineer. After the processing is completed, the module will also provide the processing information to the RFID data collection module so that the status of the workpiece can be updated into the database.

Communication module

Since each functional module runs in different industrial computers or embedded devices, to connect the information of each functional module in the system, the system introduces concepts of the Internet of things such as fog computing and edge computing. The system established the factory information communication module through OPC UA. OPC UA Server was built in the Raspberry Pi of the communication module, and the other functional modules were set as OPC UA Client. This module is built by wireless or wired network master-slave architecture, each functional module can easily read and write information by IP address.

The communication module shares factory information by reading data nodes. The module transmits various client messages downwards, and can also transmit information upwards to the cloud database, which is integrated with other plant information and provided to users for viewing. The system design of information exchange between communication modules is shown in Figure 5.

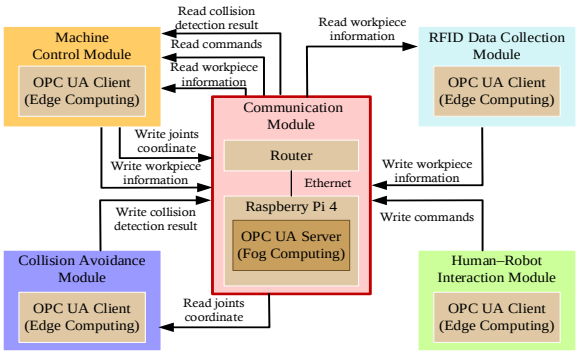

Fig. 5. System design of information exchange between communication modules.

Collision avoidance module

In this research, the collision avoidance module is designed to ensure the safety of personnel in human-robot collaboration. This module is designed according to the Speed and Separation Monitoring mode in ISO 10218-2. This module captures images of the factory through webcams and analyzes the images to determine whether there are people in the working area. If yes, and then calculate their locations in the working area. In addition, the module will continuously read the joint coordinates of the robot arm written by the robot arm control module. After the coordinates of the personnel and the robot arm are obtained, the dynamical safety space will be calculated by the module. The dynamical safety space enables users to use as much workspace as possible because the smallest safety space is calculated based on the location of the robot arm.

When a person is in an area that the robot arm cannot reach, since there is no possibility of collision, the robot arm will run the processing task at full speed. If a person continues to approach the robot arm, although he or she is still outside the moving range of the robot arm but may accidentally be injured by the falling workpiece, the robot arm will decelerate to the preset safe speed. When a person gets too close and has entered the processing range of the robot arm, the robot arm will stop immediately to avoid any potential collision accidents.

Finally, the module writes the information needed to prevent conflicts through the OPC UA server and provides it to the manipulator control module for reading.

RFID data collection module

The RFID data collection module in this study is implemented using the RFID Reader FX7500 produced by Zebra. The frequency of the RFID reader is Ultra High Frequency (UHF), which has the advantages of small tag size, low power consumption, fast data conversion rate, long reading distance, and low price. The researchers use C# and the API provided by the RFID reader to write a Windows Form application to listen to the RFID status nodes in the OPC UA Server. Besides, the processing information of the workpiece is also transmitted to the machine control module through the communication module.

This module is responsible for continuously detecting the feed state of the workpiece through the RFID antenna installed in the processing and storage area, and reading the corresponding workpiece information in the cloud database. When the workpiece is deliver to the working area, the RFID antenna will scan the tag on the workpiece and read the ID information of the workpiece stored in the tag. Then, the module will search the database for the type, processing status, and other information of the workpiece corresponding to the ID. Through this information, the machine control module can drive the CNC to carry out the corresponding processing action or display the processing prompt information to the engineer on the human-machine interface. Besides, the module also updates the status information of the workpiece stored in the label through the antenna after the workpiece is processed by the personnel or CNC, to achieve intelligent manufacturing management.

EXPERIMENTS AND DISCUSSIONS

To verify the functional practicability of the human-robot collaboration system proposed by this research, the researchers conducted an experiment in a simulation environment for smart manufacturing. The physical facilities of this simulation application environment are shown in Figure 6., and the application environment simulated in this experiment is shown in Figure 7.

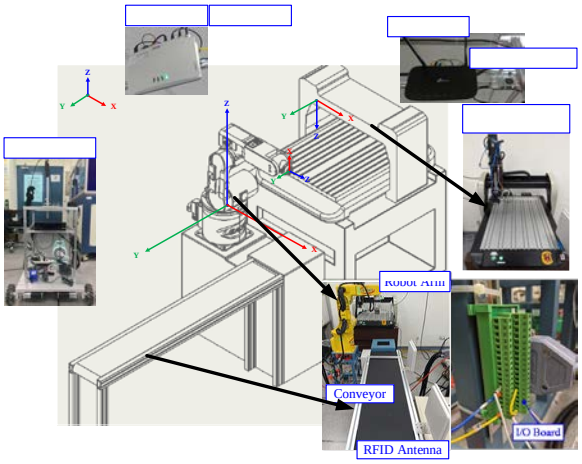

Fig. 6. Physical facilities of the simulation application environment.

The automatic loading and unloading of the production line and human-robot collaborative processing are used as a simulated application scenario. The transportation and processing of workpieces are executed by the AGV, conveyor belt, robot arm, and CNC milling machine in the production line. RFID tags are attached to the surface of each workpiece, and the RFID reader and antennas around the production line are responsible for scanning, reading, and editing the information in the tags. Besides, a webcam and microphone for receiving hand gestures and voice commands from personnel, as well as a Wi-Fi router and Raspberry Pi for information communication are also set up above the production line. Before the start of the experiment, the AGV and CNC milling machine are set to automatic processing mode, and the controllers of the above devices will remain stationary and monitor specific communication ports. The motion macro file that AGV and CNC will not be executed until the notification signal for executing processing is sent to the communication port.

The experimental process designed in this study is shown in Figure 8. In this process, the system proposed by this research is applied in the manufacturing environment, and the automatic loading and human-robot collaborative processing in the actual production line is simulated. The blue block in the flow chart is the task of the system, while

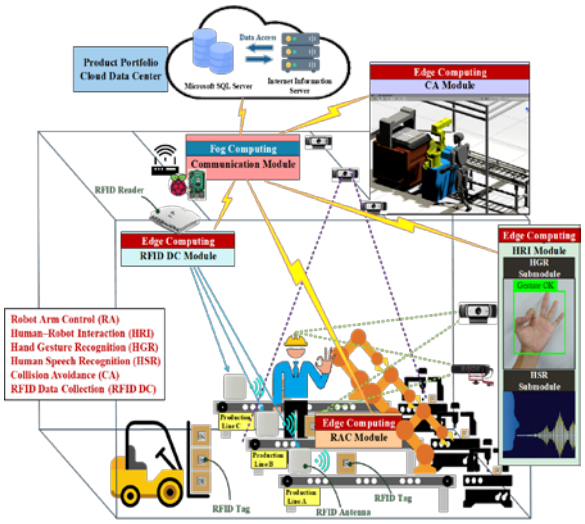

Fig. 7. Application environment simulated in this experiment.

the green block is the step of human-robot collaboration. During the experiment, the collision avoidance module will constantly monitor the position of the personnel and the robot arm. When a person is too close to the robot arm, this module will send a stop or deceleration command to the machine control module to avoid potential accidents.

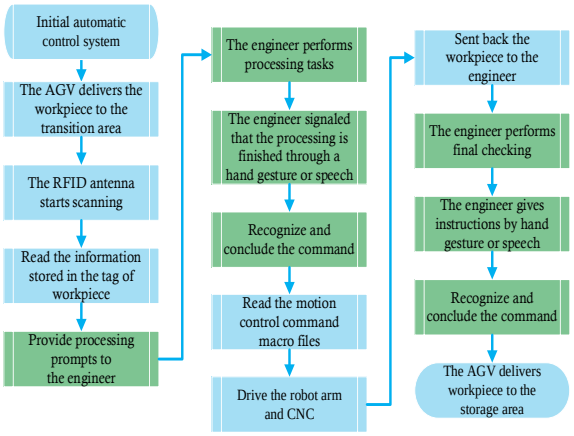

Fig. 8. Human-robot interaction process flow diagram based on speech recognition.

At the beginning of the experiment, the machine control module instructs the AGV to move the workpiece to the transition area. After the RFID tag on the workpiece is scanned by the antenna, the RFID data collection module will read the ID stored in the tag and search the data of the object in the workpiece management database. The RFID data collection module will then transmit the machining information of the workpiece to the machine control module through the OPC UA communication module. According to this information, the machine control module can display the machining prompt information through the human-machine interface and read the motion macro file corresponding to the workpiece.

After the engineer completes the processing tasks with the prompts provided by the human-machine interface, the completed information can be conveyed to the system through simple hand gestures or oral speech. The system will execute the motion macro file, move the workpiece through the robot arm to the processing platform of a CNC machine. After the workpiece is sent to the CNC machining platform, the machine control module will send a signal to the CNC through the I/O board to trigger the CNC to start the machining action.

Finally, after CNC machining is completed, the machine control module will also be notified through the I/O board. The module can execute subsequent motion macros to send the finished workpiece back to the engineer for the final checking. After the inspection is completed, the engineer can communicate the inspection result to the system by hand gesture or speech. The RFID data collection module will write the result into the tag of the workpiece through the antenna and update the data in the cloud database at the same time. If the inspection result is normal, the AGV will send the finished product to the storage area; and if the inspection result is bad, the AGV will send the defective product to the defective product stacking area.

To test the recognition rate of the custom YOLOv4 model, this research also designed a simple experiment. During the experiment, the subject stood in a controlled laboratory environment, made each control gesture toward the webcam, and recorded videos of each gesture of varying length. After recording the videos, one hundred frames of each video will be randomly captured and input into the program. The frames will be detected by the custom YOLOv4 model, and the recognition results and degrees of confidence will be recorded. The results of this experiment are shown in Table 1. It can be seen from the table that the hand gesture in each frame can be correctly recognized, and high degrees of confidence are obtained.

Table 1. Results of recognition rate and average confidence experiment.

Types of gesture / Results	0	1	2	3	4	5	6	OK
Recognition rate (%)	100	100	100	100	100	100	100	100
Average confidence (%)	97.2	97.7	97.1	98.9	95.2	99.5	99.9	82.7

CONCLUSIONS

Several technologies have been integrated into this research to realize a human-robot collaboration system that can be applied to a variety of application environments, and the manufacturing simulation field is used as an experimental environment to prove its practicality. Because the system proposed in this study adopts a modular method to design functions, it is easy to edit and has good scalability. In the future, if the system needs to be integrated into different application environments, developers only need to adjust the functional modules according to their needs or add new customized functional modules, and the benefits of human-robot collaboration can be exerted in the application.

In the human-computer interaction functional module of this system, the method of interacting with the robot with intuitive hand gestures and speech is realized. Compared with the method of using wearable devices to recognize hand gestures, the computer vision combined with deep learning technology used in this research is more convenient to use. The YOLO4 algorithm is used to realize the hand gesture recognition function and has good recognition accuracy and speed, which are important factors affecting practicality in various applications. For human speech recognition, the implementation method used by this research allows the system to recognize the personnel voice offline, and it is simple to expand the types of speech commands. The above advantages make the developed system easier to be imported into different real industry applications.

ACKNOWLEDGMENT

The authors would like to extend our sincerest thanks for the support by Industrial Technology Research Institute under the contract B109-F007 and Ministry of Science and Technology under the contract MOST 107-2221-E-006-230-MY2.

REFERENCES

Adolph, Lars, T. Anlahr, and H. Bedenbender. "German standardization roadmap: Industry 4.0." Version 2. Berlin: DIN eV (2016).

Bdiwi, Mohamad, Marko Pfeifer, and Andreas Sterzing. "A new strategy for ensuring human safety during various levels of interaction with industrial robots." CIRP Annals 66.1 (2017): 453-456.

Ghazivakili, Mohammad, Claudio Demartini, and Claudio Zunino. "Industrial data-collector by enabling OPC-UA standard for Industry 4.0." 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS). IEEE, 2018.

Gomez, R., Nakamura, K., Mizumoto, T., & Nakadai, K. (2015, November). Compensating changes in speaker position for improved voice-based human-robot communication. In 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids) (pp. 977-982). IEEE.

ISO, ISO. "10218-2: 2011: Robots and robotic devices-Safety requirements for industrial robots-Part 2: Robot systems and

integration." Geneva, Switzerland: International Organization for Standardization 3 (2011).
Jain, M., et al. "Object Detection and Gesture Control of Four-Wheel Mobile Robot." 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, 2019.
Liu, Hongyi, and Lihui Wang. "Gesture recognition for human-robot collaboration: A review." International Journal of Industrial Ergonomics 68 (2018): 355-367.
Meng, Zhaozong, Zhipeng Wu, and John Gray. "RFID-based object-centric data management framework for smart manufacturing applications." IEEE Internet of Things Journal 6.2 (2018): 2706-2716.
Peter X. Liu, A.D.C. Chan, R. Chen, K. Wang, Y. Zhu, "Voice Based Robot Control", Proceedings of the 2005 IEEE International Conference on Information Acquisition, June 27 - July 3, 2005.
Subhash P. Rasal, "Voice Controlled Robotic Vehicle", International Journal of New Trends in Electronics and Communication, vol. 2, no. 1, pp. 28-30, 2014, ISSN 2347 - 7334.

基於智慧手勢和語音辨識實現之人機協作系統

陳響亮 黃立武
國立成功大學製造資訊與系統研究所

黃仲麒
遠東科技大學自動化控制系

李峰吉
工業技術研究院機電控制整合部

黃河銓
國立高雄科技大學智慧商務系

陳建佑
國立臺南大學資訊工程學系

摘 要

隨著人工智慧技術的發展，機械手臂將在更多樣的應用領域中承擔不同的工作，成為人類重要的助手。尤其在製造業中，若能開發出完善的人機協作模式，由工程師負責複雜的操作並監控生產過程，而機械手臂則負責重複性高的勞力工作，將可以提高製造過程的彈性、減少工作人員受傷的風險，並且提高產線的生產率。為了實現能應用在各個領域中的人機協作系統，本研究結合了機械手臂控制、人機互動、無線射頻辨識、機器對機器通訊和自動語音識別等多項技術，並以智慧製造中的人機協作作為研究成果的展現。在此應用範例中，人機互動功能通過智慧手勢和語音辨識實現，讓使用者能以最為直觀、簡易的方式將指令傳達給機械手臂。根據不同的指令內容，機械手臂可以執行特定的動作、提供加工提示資訊或讀取指定的運動控制巨集，以實現加工過程中使用者的各項需求。