·Article·

A multichannel human-swarm robot interaction system in augmented reality

Mingxuan CHEN, Ping ZHANG*, Zebo WU, Xiaodan CHEN

South China University of Technology, School of Computer Science & Engineering, Guangzhou 510006, *China*

*** Corresponding author,** pzhang@scut.edu.cn

Received: 25 February 2020 **Accepted:** 5 May 2020

Supported by Key-Area Research and Development Program of Guangdong Province (2019B090915002).

Citation: Mingxuan CHEN, Ping ZHANG, Zebo WU, Xiaodan CHEN. A multichannel human-swarm robot interaction system in augmented reality. Virtual Reality & Intelligent Hardware, 2020, 2(6): 518—533
DOI: 10.1016/j.vrih.2020.05.006

Abstract **Background** A large number of robots have put forward the new requirements for human-robot interaction. One of the problems in human-swarm robot interaction is how to naturally achieve an efficient and accurate interaction between humans and swarm robot systems. To address this, this paper proposes a new type of human-swarm natural interaction system. **Methods** Through the cooperation between three-dimensional (3D) gesture interaction channel and natural language instruction channel, a natural and efficient interaction between a human and swarm robots is achieved. **Results** First, A 3D lasso technology realizes a batch-picking interaction of swarm robots through oriented bounding boxes. Second, control instruction labels for swarm-oriented robots are defined. The instruction label is integrated with the 3D gesture and natural language through instruction label filling. Finally, the understanding of natural language instructions is realized through a text classifier based on the maximum entropy model. A head-mounted augmented reality display device is used as a visual feedback channel. **Conclusions** The experiments on selecting robots verify the feasibility and availability of the system.

Keywords Human-swarm interaction; Augmented reality; Multichannel integration

1 Introduction

With multi-robot systems being increasingly applied in different fields to solve practical problems, swarm robotics have become a hotspot and frontier of intelligent technology research. A swarm robotic system is a category of multi-robot systems. Unlike a multi-robot system, a swarm robotic system is inspired by the natural behavior of animal groups. Through the local interaction of a swarm of simple and low-cost robots, swarm robotics can help complete tasks that are difficult or impossible for a single robot. Distinct from the traditional central control, each robot in the swarm robot system obtains information about its own state and surrounding environment, including other robots around it, and interacts with other robots in the system through local control rules for self-organization and coordination control. Swarm robot systems accomplish the target work through self-organized collective behavior, such as foraging[1], covering[2], swarming[3], and other methods. Owing to the simple function and structure of each robot in a swarm robot system, it has distributed characteristics, enabling it to successfully complete a task when an individual

robot fails. It is high robust, scalable, and flexible. Replacement of a swarm robotic system can be accomplished by simply using the same simple and inexpensive robot.

Swarm robotic systems can be particularly useful for solving various problems[4], such as path finding, environment exploration, rescue or support in dynamic environments, and even space exploration[5]. With the use of swarm robotic systems, on the one hand, labor costs can be reduced, and workers can avoid injuries in hazardous working environments. On the other hand, they can complete tasks that cannot be completed by a robot, and provide new ideas for solving problems.

Faced with the current research situation and the wide application scope of swarm robotics, a natural and efficient interaction between humans and swarm robots has become imperative. Human-swarm interaction (HSI)[6] research focuses on the methods that can effectively transmit human control intent. At the same time, the self-organized behavior of a swarm robot can be improved. In fact, in some application scenarios, the introduction of human influence factors to a swarm robotic system can have beneficial, or even critical effects. For example, in the face of drastic environmental changes, it is difficult for a swarm of robots to provide a rapid response, or adapt in a self-organizing manner. Human experience can play a particularly important role in such a situation. However, HSI has not received enough attention, and the related research is still in its infancy[7]. There is an imperative need for an intuitive HSI interface to better communicate human interaction intentions to the swarm robots, and complete tasks faster and better. In this paper, we focus on the interface (batch selection and visual feedback) between a human operator and a swarm of robots.

One of the most studied problems in HSI is designing a suitable control input method for the swarm robots. Kolling et al.[6] pointed out to four modes of HSI: supervision mode[8], direct control mode[9], shared control mode[10], and environmental impact control mode[11]. Gromov et al. adopted wearable design and realized interaction with small-scale swarm robots through gesture interaction combined with voice, vision, and action channels[12]. Selective interactive operation on swarm robots can take time if the number of robots is large. In order to make the interaction time independent of the number of robots, we propose a batch selection interaction method.

Erat et al. utilized a head-mounted display (HMD) to deliver an exocentric perspective on a drone, letting the operator control the drone via gaze[13]. With human vision support, spatial understanding is augmented, and the user interface improves natural interaction with the drone. Our HSI system lets the operator control robots in an outdoor environment. The operator can overview the entire work scenario via a HMD.

Tsykunov et al. presented a vibrotactile glove for the interaction of a human with a swarm of aerial robots by intuitively mapping the formation state to human finger pads[14]. Tactile cues can supplement the visual channel, making the swarm control more immersive. However, the user has to memorize the meaning of each tactile pattern.

The focus of HSI research will still be on efficiently and accurately transferring human control intent to robot swarm(s) with self-organization through a natural and intuitive interaction.

The main contributions of this paper are the following: (1) A multichannel HSI system with augmented reality is designed. The interactive information is complemented through the collaboration of the augmented reality display channel, three-dimensional gesture interaction channel, and natural language instruction channel. The respective multichannel advantages effectively reduce the load on a single interaction channel and increase the overall interaction information bandwidth. (2) A coordinate system relationship between the interactive scene and the interactive object is established. A swarm robot selection method is presented, which uses a three-dimensional (3D) lasso based on space division, to select and interact with the virtual robots in the scene in batches. (3) A control instruction library and corresponding instruction labels for HSI are established, and natural language instructions corresponding

to control instructions are collected to build a corpus. Multiple channels are integrated through instruction label filling.

2 Design of multichannel HSI system with augmented reality

In order to satisfy the portability, intuitive feedback, and natural interaction needs, the proposed interaction system uses an HMD as a visual feedback device. A gesture sensor fixed on the HMD captures the controller's gesture information. HMD's built-in microphone captures the controller's voice information, which is processed on a portable PC. The HMD is used as a visual feedback channel, realizing intuitive information feedback through augmented reality. Effective channels of the system include a 3D gesture interaction channel and a natural language instruction channel, allowing the controller to use 3D gestures and instructions that are closer to natural language. This ensures the naturalness of the interaction to a certain extent. Finally, the presented system integrates 3D gesture interactions and natural language instructions by filling instruction labels to achieve robot selection control and motion trajectory control, among others.

The software modules of the presented system is shown in Figure 1. The main five functional modules are: *augmented reality display* module, *three-dimensional gesture interaction* module, *speech recognition* module, *natural language instruction understanding and multichannel integration* module, and *message communication* module. The hardware components of the system include an augmented reality display device (Microsoft HoloLens) running *speech recognition* module and *augmented reality display* module. A gesture sensor (Leap Motion) acquires the position and posture data of the controller's hand. A portable PC (laptop) runs the *three-dimensional gesture interaction* module, *natural language understanding and multichannel integration* module, and *message communication* module. Finally, a wireless router is used for network transmission.

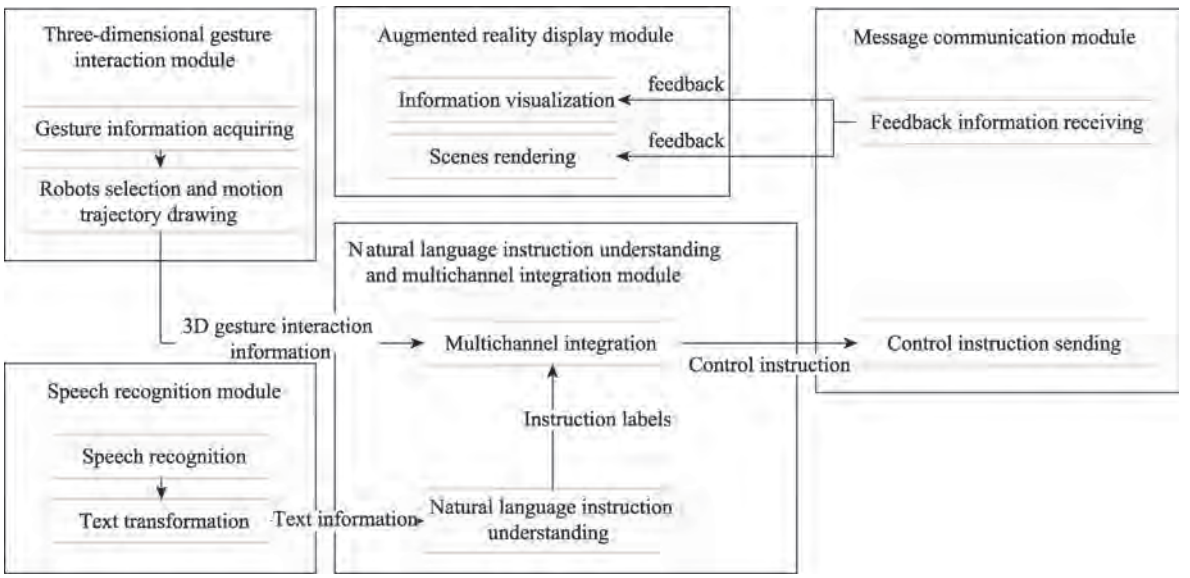

Figure 1 The software module of the human-swarm robot natural interaction system.

The *augmented reality display* module is responsible for the rendering of 3D virtual interactive scenes and the visualization of feedback information. 3D virtual interactive scenes include a digital map of structured environment and virtual robots. Digital map is used to monitor the current position and status of the swarm robots, and is invoked as interactive objects to designate coordinates. As chosen interactive

objects, virtual robots are the mappings of real robots in 3D virtual interaction scenes. The feedback information includes the system status and task status of robots.

After acquiring the controller's gesture information, the 3D gesture interaction module realizes the robot selection and draws the motion trajectory. 3D gesture interaction information generated by this module is transmitted to the *natural language instruction understanding and multichannel integration* module.

The *speech recognition* module recognizes the controller's voice signal through a speech recognition engine (Microsoft Speech Platform SDK), and transforms it into text. Then, the text is transmitted to the *natural language instruction understanding and multichannel integration* module.

After receiving the text, *natural language instruction understanding and multichannel integration* module understands the natural language instruction in the speech text through a pre-trained text classification model to obtain labels describing specific control instructions. Then, according to the content of the instruction label, and combined with the corresponding 3D gesture interaction information, an integrated control instruction is obtained.

The *message communication* module distributes control instructions from the *natural language instruction understanding and multichannel integration* module to the swarm robots, and sends the received task execution results and feedback information, such as robot status information, to the *augmented reality display* module for visualization.

3 3D gesture interaction technology with augmented reality

3.1 Acquisition of gesture data and establishment of 3D virtual interaction scene

Gesture data is obtained by Leap Motion, a body motion controller by Leap. When the controller's hand appears in the sensor's working space, it can identify and capture human hand information, and encapsulate the data captured at each moment in a data frame. Each data frame contains hand parameters captured at that moment, including palm position, normal vector, and movement speed, along with position, direction, and movement speed of fingertips.

The proposed system uses augmented reality as a visual feedback channel. The controller interacts with digital maps and virtual robots in a 3D virtual interactive environment rendered by an augmented reality device. As shown in Figure 2, in a 3D virtual interactive scene, the digital map contains the environmental information of the robot's actual working scene. The virtual robots are mappings of the real robots in the virtual scene. The virtual robots' coordinates in the interactive scene are determined by the real robots' positions.

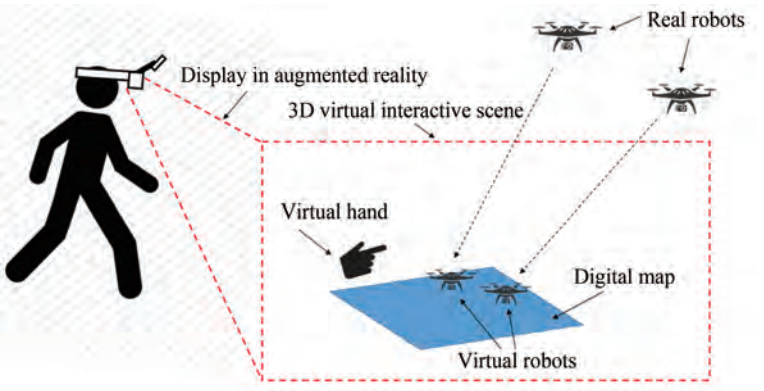

Figure 2 3D virtual interaction scene.

The coordinate system of the 3D virtual interactive scene is shown in Figure 3. $X_lY_lZ_l$ is used to represent the Leap Motion sensor coordinate system, which describes the gesture data. $X_hY_hZ_h$ is used to represent the HoloLens observation coordinate system, which represents the coordinate system where the graphics rendering results are located, and is also the coordinate system observed by the controller's eyes. The origin of $X_hY_hZ_h$ changes with the controller's head movement. $X_wY_wZ_w$ represents the world coordinates of the 3D virtual interactive scene. As the base coordinate system of the entire scene, its origin is determined internally by HoloLens, and remains unchanged during the entire interaction process. $X_{d1}Y_{d1}Z_{d1}$, $X_{d2}Y_{d2}Z_{d2}$, and $X_{d3}Y_{d3}Z_{d3}$ describe the coordinates systems of each virtual robot. Virtual robots' positions correspond to the real robots' GPS information. $X_mY_mZ_m$ represents the coordinate system of the digital map.

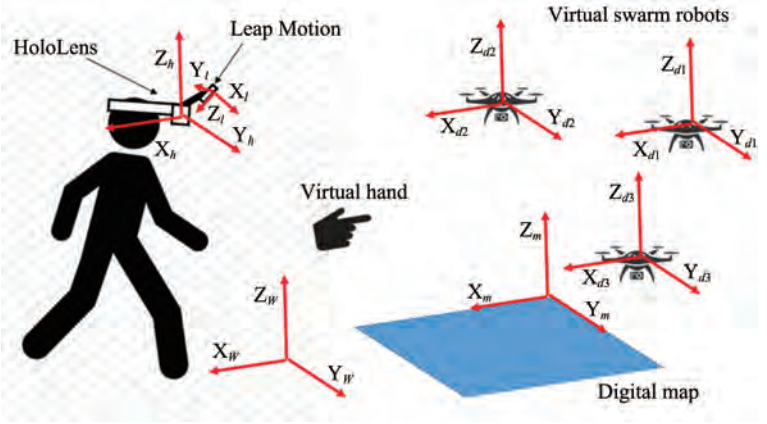

Figure 3 Relationship between coordinate systems of the 3D virtual interaction scene.

We subtract the origin O_w of the world coordinate system from the origin O_h of the HoloLens observation coordinate system to obtain the translation matrix T_h^w between the two coordinate systems.

$$T_h^w = O_w - O_h \tag{1}$$

The three Euler angles α, β, and γ can be obtained from HoloLens's pose information, and the transformation matrices around the world coordinate X, Y, and Z axes are defined as R_w^x,R_w^y, andR_w^z.

In order to facilitate the matrix transformation, we transform T_h^w, R_w^x, R_w^y, andR_w^z into a homogenously represented matrix, and multiply them to obtain the transformation matrix of the world coordinate system $X_wY_wZ_w$ to the HoloLens observation coordinate system $X_hY_hZ_h$. Let M_h^w represent the transformation matrix. Then:

$$M_h^w = R_w^z \cdot R_w^y \cdot R_w^x \cdot T_h^w \tag{2}$$

From any point p_w of $X_wY_wZ_w$ to point p_h of $X_hY_hZ_h$,

$$p_h = M_h^w \cdot p_w \tag{3}$$

The controller's movements and operations in the Leap Motion workspace will be identified and transmitted to the 3D virtual interactive scene to drive the virtual hand for the interaction operation. Therefore, we need to obtain the transformation matrix of the Leap Motion coordinate system $X_lY_lZ_l$ to the HoloLens observation coordinate system $X_hY_hZ_h$.

Similarly, the transformation matrix M_h^l from $X_lY_lZ_l$ to $X_hY_hZ_h$ can be obtained according to the Leap Motion coordinate system's Euler angle relative to the HoloLens observation coordinate system. The transformation of the coordinates of a point in $X_lY_lZ_l$ to $X_hY_hZ_h$ is calculated as follows:

$$p_h = M_h^l \cdot p_l \tag{4}$$

where p_h describes a point in $X_hY_hZ_h$ and p_l describes a point in $X_lY_lZ_l$.

The transformation matrix M_w^{d1},..., M_w^{dn} of each robot's local coordinate system $X_{d1}Y_{d1}Z_{d1}$,..., $X_{dn}Y_{dn}Z_{dn}$ to the

world coordinate system can be obtained by the same method according to the positions and Euler angles of each robot in the world coordinate system $X_wY_wZ_w$, thereby obtaining the coordinates of each robot in the virtual interaction scene. Likewise, the transformation matrix M_w^m of the coordinate system $X_mY_mZ_m$ of the digital map to the world coordinate system can be obtained.

3.2 3D lasso selection interaction of swarm robots

The 3D lasso technology allows the operator's hand to draw a closed irregular curve in the Leap Motion workspace. The drawn curve will be posted in augmented reality. The robots inside the irregular lasso curve, closed from the operator's perspective, are selected, and robots outside the lasso curve are unselected to achieve batch swarm robot selection interaction.

The 3D lasso technology uses curve interpolation to transform the controller's discrete positions in the Leap Motion workspace into a smooth lasso curve in a 3D virtual interactive scene. Differing from a curve composed of piecewise low-order polynomials, a spline curve connects points with smooth curve segments. The cubic spline is a widely used curve interpolation method. The obtained curve is smooth and of second order, and the convergence is guaranteed. The cubic spline curve[15] is used to model the lasso curve. All virtual robots and the lasso curve plane are converted to HoloLens screen coordinate system to determine the selected robots. The screen coordinate system is the plane coordinate system to which the virtual environment transforms when the controller is viewing the scene. This is a 2D coordinate system. The algorithm for determining the relationship between points and curves on the plane can be used to select the robots (Figure 4).

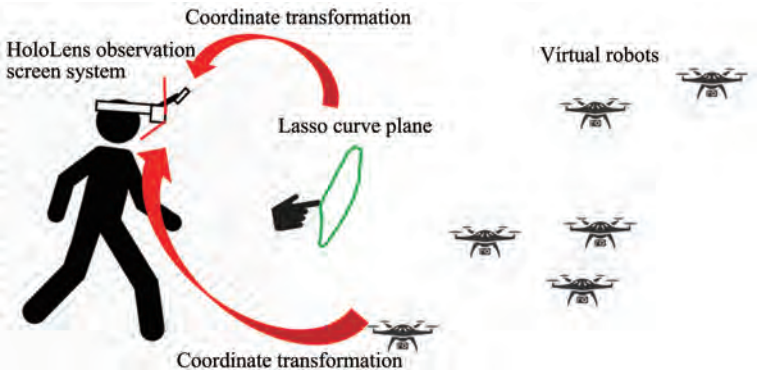

Figure 4 The coordinate transformation to HoloLens screen coordinate system.

In order to transform the figure from the observation coordinate system to the screen coordinate system, a projection transformation matrix should be obtained. The projection transformation matrix is determined by the parameters of the frustum. The frustum of the perspective transformation is shown in Figure 5.

The frustum of the perspective projection is shaped as a quadrangular pyramid. Objects inside the frustum will enter the next workflow of the rendering pipeline. Objects outside will be discarded and not

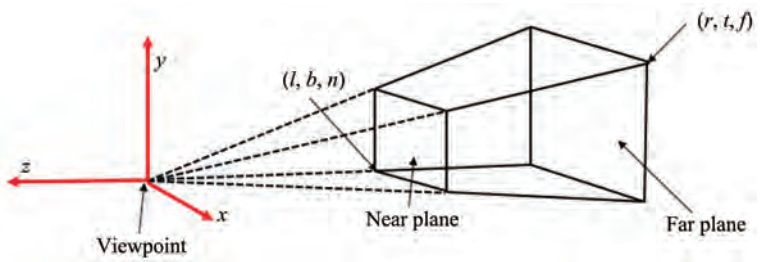

Figure 5 The frustum of the perspective transformation.

processed. The lower left corner point of the near plane is (l, b, n), and the upper right corner point of the far plane is (r, t, f). The perspective projection transformation matrix M_{proj} is as follows:

$$M_{proj} = \begin{bmatrix} \frac{2n}{r-l} & 0 & \frac{r+l}{r-l} & 0 \\ 0 & \frac{2n}{t-b} & \frac{t+b}{t-b} & 0 \\ 0 & 0 & \frac{n+f}{n-f} & \frac{2nf}{n-f} \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{5}$$

The transformation matrices M_w^d, M_h^w, and M_h^l can be obtained from the coordinate system described in Section 3.1, realizing the transformation from the robot local coordinate system $X_dY_dZ_d$ to the world coordinate system $X_wY_wZ_w$, the world coordinate system $X_wY_wZ_w$ to the HoloLens observation coordinate system $X_hY_hZ_h$, and the Leap Motion coordinate system $X_lY_lZ_l$ to HoloLens observation coordinate system $X_hY_hZ_h$, respectively.

The oriented bounding box (OBB) is used instead of the virtual robot. The robot's OBB has 8 vertices, 12 midpoints, and 21 center points as the reference points, as shown in Figure 6.

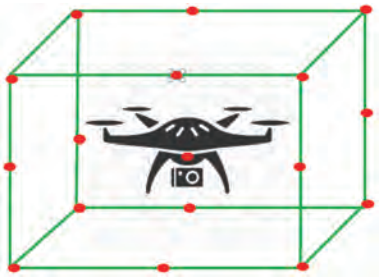

Figure 6 The OBB of swarm robot.

For the i-th reference point on the OBB of robot d, the coordinate point p_{proj}^i in the HoloLens screen coordinate system is obtained by the following transformation:

$$p_{proj}^i = M_{proj} \cdot M_h^w \cdot M_w^d \cdot p_d^i \tag{6}$$

Similarly, the following formula is used to transform each data point p' of the lasso curve in the Leap Motion coordinate system to the coordinate point p'_{proj} in the HoloLens screen coordinate system:

$$p'_{proj} = M_{proj} \cdot M_h^l \cdot p' \tag{7}$$

The lasso curve and swarm robots are transformed from an observation space to a screen space through perspective projection matrix transformation, thereby obtaining the projection information. As shown in Figure 7, the screen space is a 2D coordinate system, and the vertical axis is stored as a depth value in a depth buffer for the rendering process. Therefore, judging whether the robot is inside the lasso curve in the screen coordinate system can realize the selection interaction of the swarm robots.

Lasso curve data points and reference points of the robot are transformed into the HoloLens screen space by the perspective projection matrix, which transforms the problem from 3D to 2D space. After obtaining

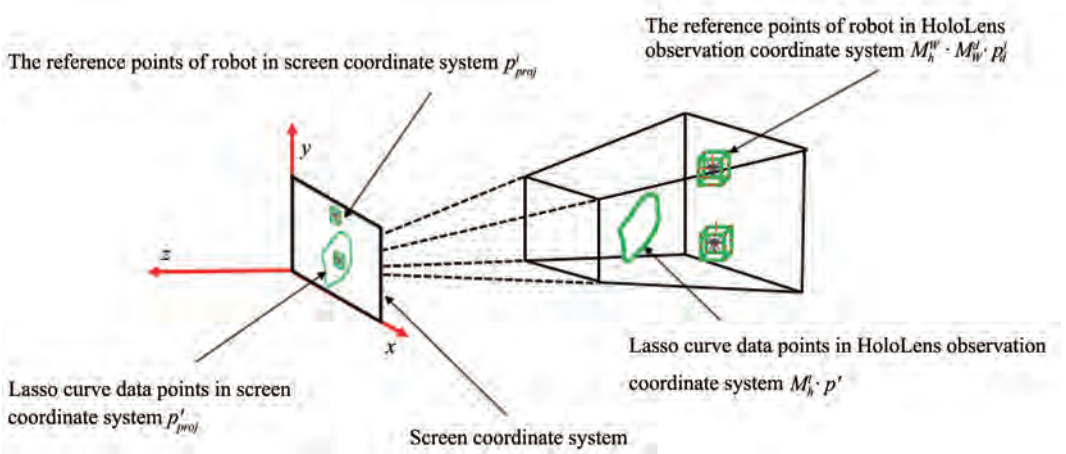

Figure 7 The transformation from observation space to screen space.

the projection of the lasso curve data points and reference points of the robot on the screen space, we use the ray method[16] to determine whether the robot reference points are surrounded by the lasso curve.

For the 21 reference points of robot k, if the number of points inside the curve is 11 or more, the robot is selected. On the other hand, if it is less than 11, the robot is not in the lasso.

3.3 Swarm robot selection interaction based on space division

The purpose of the robot selection interaction is to obtain a set of robots that contain the target robots and exclude non-target robots. In a practical application, the positions of various robots in the swarm robot system will change as the task progresses, showing a certain degree of mixing in space. Therefore, certain accuracy requirements of the robot selection interaction are put forward. As shown in Figure 8a, there is a spatial mixing of robots in the scene. When the target robots are all type 1, the 3D lasso drawn by the controller is required to include them, while excluding all type 2 and type 3 robots. As shown in Figure 8b, 3D lasso technology described in Section 3.2 imposes high accuracy requirements on the 3D lasso drawn by the operator when implementing specific robot selection interactions, resulting in a reduced interaction efficiency and declining success rate.

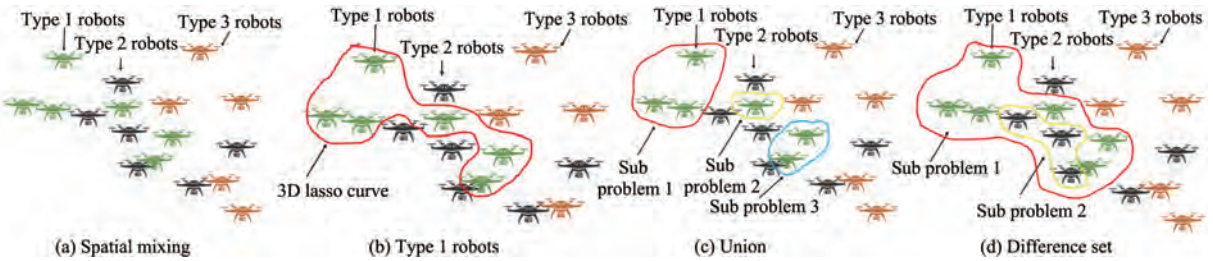

Figure 8 The spatial mixing scenario of selection interaction.

In order to solve the problem of reduced interaction efficiency and accuracy of 3D lasso technology when selecting specific robots, the problem is decomposed into smaller sub-problems. The solutions to the sub-problems are then aggregated to solve the entire problem. As shown in Figure 8c, the original problem is decomposed into three sub-problems related to selecting robots in different spaces. After solving the three sub-problems, the solution to the original problem can be obtained by an union operation on the robot sets. Figure 8d shows that the original problem is decomposed into sub-problems that select two types of robots in different spaces. The solution to the original problem can also be obtained through a difference set operation on the robot sets.

During the interaction, the controller interacts with the virtual 3D scene through a viewing angle. The robots can easily block each other in the direction of the viewing angle, making it impossible to accurately select specific robots through the 3D lasso. Owing to the head-mounted augmented reality device, the controller can change the viewing angle by adjusting his/her position and attitude, and interact with the robots based on a space division method to solve the problem of robot selection with obstruction in the line of sight. As shown in Figure 9, the robot selection problem is decomposed into two sub-problems using space division. The sub-problems are obtained from the 3D lasso curves from different perspectives. Their solutions are then combined to solve the original problem.

$$T = A_1 * A_2 * \cdots * A_n \tag{8}$$

where T is the target selection robot set, A_i is the solution of the robot selection sub-problem, and n is the number of sub-problems. Symbols $*, * \in \{\cap, \cup, \setminus\}$ are set operators, denoting the intersection, union, and difference of two sets. That is, the set of selected robots is generated by the sub-problems.

4 Natural language instruction understanding and multichannel integration

4.1 Establishment of interactive control instructions

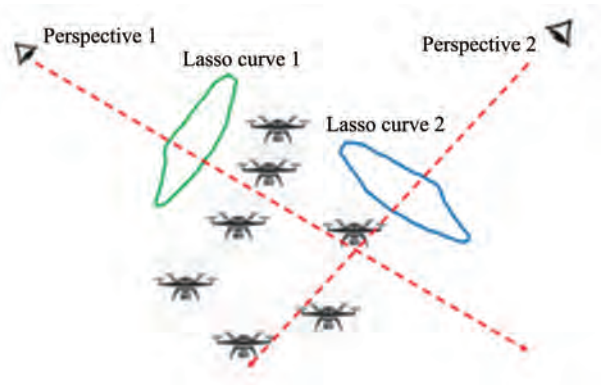

Figure 9 Spatial division method combined with perspective transformation.

Based on the need for interactive control, four types of instructions are established. (1) *Robot selection instructions*: Select any number of specific robots in the swarm robotic system. It is necessary to integrate the robot set information provided by the 3D gesture interaction technology to obtain complete control instructions. (2) *Coordinate position control instructions*: Control instructions associated with specific coordinate points, such as going to a certain place for gathering, standing by, or formation, need to integrate the coordinate information provided by the 3D gesture interaction technology to obtain a complete control instruction. (3) *Motion trajectory control instructions*: Control instructions associated with a specific trajectory, such as moving along a trajectory, patrolling, reconnaissance, searching, etc., also need to integrate the trajectory information provided by 3D gesture interaction technology to obtain complete control instruction. (4) *State control instructions*: These complete the most basic system state control operations of swarm robots, such as take-off, landing, stand-by, and report status.

To describe the control instruction of the robot, the instruction label and the label classified in the text classification model are introduced. The basic instruction label is a triplet $<T_{id}, T_{key}, T_{val}>$. T_{id} is the category of the control instruction, including the above four types: *robot selection instruction*, *coordinate position control instruction*, *motion trajectory control instruction*, and *state control instruction*. T_{key} is the label keyword, which usually indicates the task action of the control instruction. T_{val} is the label value, which is usually a parameter that controls the action of a command task.

In order to describe some special robot selection instructions, this article introduces an extended instruction label $<T_{id}, T_{key}, T_{val1}, T_{val2}, T_{link}>$ which has two label values, and adds T_{link} to describe the relationship between the two label values. The extended instruction label is used to describe the extended robot selection instruction.

Control instruction corpus is a natural language text form of the control instructions. Each control instruction has several corpora corresponding to it. As shown in Table 1, the left side is the labels of the control instructions, while the right side is the natural language text corpus. Natural language text corpus is used as the training data set. The instruction label is used as the classification label of the corpus to train the text classification model.

Table 1 Example of control instruction and corpus

Instruction labels	Natural language test corpus
<0, select collection>	Select these robots
<0, select, collection 1, collection 2, and>	Select these robots in this range
<1, reach, coordinate point>	Go here
<2, move, trajectory>	Follow this trajectory
<3, report, battery>	Report battery

4.2 Multichannel integration

The proposed human-swarm robot natural interaction system involves cooperation and information complementation between the 3D gesture interaction channel and the natural language instruction channel. The multichannel integration integrates the complementary interaction information of the natural language instruction channel and the 3D gesture interaction channel into complete control instructions.

The integration of speech and gesture channels is achieved through the filling of instruction labels[17], which is divided into two phases. First, in the instruction label generation and speech parameter filling phase, the instruction label is generated after understanding the controller's speech text information by natural language instructions, and the speech interactive parameter is filled in the component corresponding to the instruction label. Then, in the gesture parameter filling phase, the gesture interaction parameter from the 3D gesture interaction system is filled in the corresponding instruction label component.

In the instruction label generation and speech parameter filling phase, the information of the speech channel is translated into specific instruction labels, such as basic instruction labels $<T_{id}, T_{key}, T_{val}>$ and extended instruction labels $<T_{id}, T_{key}, T_{val1}, T_{val2}, T_{link}>$, by understanding text instructions. The instruction label is in a partially filled state after filling information of the speech channel to the corresponding component. In the gesture parameter filling phase, the 3D gesture interaction parameters are filled according to the type and quantity of instructions. After properly filling each component of the instruction label, a complete control instruction that the robot can execute can be obtained, as shown in Figure 10.

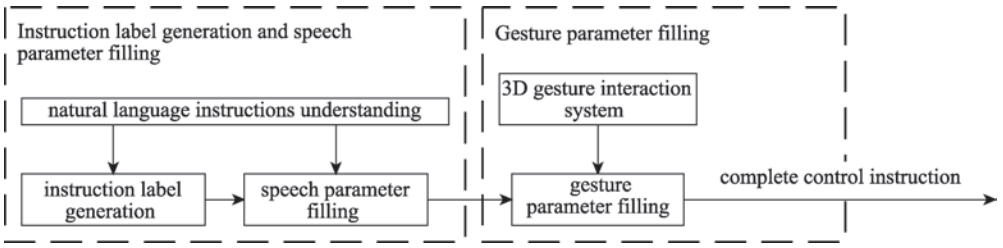

Figure 10 The process of multichannel integration.

The text classification natural language instruction understanding based on the maximum entropy model implements the instruction label generation and speech parameter filling. During the gesture parameter filling, the gesture parameter number matching rules and parameter type matching rules are used to fill the gesture parameters according to the control instruction. When the gesture parameters obtained from the 3D gesture interaction system cannot meet the above rules, a conflict occurs between the interaction channels. At this time, all the interaction information of the gesture channel and the speech channel are cleared, and the controller is prompted to restart the interaction through information feedback. Otherwise, the instruction label is completed to obtain a complete control instruction.

For instructions described by the basic instruction tag $<T_{id}, T_{key}, T_{val}>$, the control instruction type T_{id} and task action T_{key} are filled in during the natural language instruction understanding process. The task parameter T_{val} needs to be obtained from the 3D gesture interaction system, whose number is 1.

The extended instruction tag $<T_{id}, T_{key}, T_{val1}, T_{val2}, T_{link}>$ with 5 components is used to describe the extended robot selection control instructions. T_{id} and T_{key} are determined during the natural language instruction understanding process. The second task parameter component T_{val2}, based on differences between control instructions, may be filled in during natural language instruction understanding, or obtained from 3D gesture interaction system.

After the instruction label generation and speech parameter filling stages, the control instruction's T_{id}∈(0, 1, 2, 3), task action T_{key}∈(reach, select, standby, formation, ……), and task parameter T_{val}∈ (COLLECTION, with-camera, battery-high, …), where COLLECTION information is provided by the gesture channel and the rest are provided by the speech channel. The task parameter relationship is T_{link}∈(and, union, except). As shown in Figure 11, the extended robot selection instruction obtains the semantic information of the target robot through the speech channel, and sets T_{val2} to with-camera. Then, the robot collection of the camera performs set operations with COLLECTION1 of the gesture channel. Through the cooperation of the 3D gesture channel and the natural language instruction channel, it generates more accurate collection information, which improves the efficiency and accuracy of the interaction.

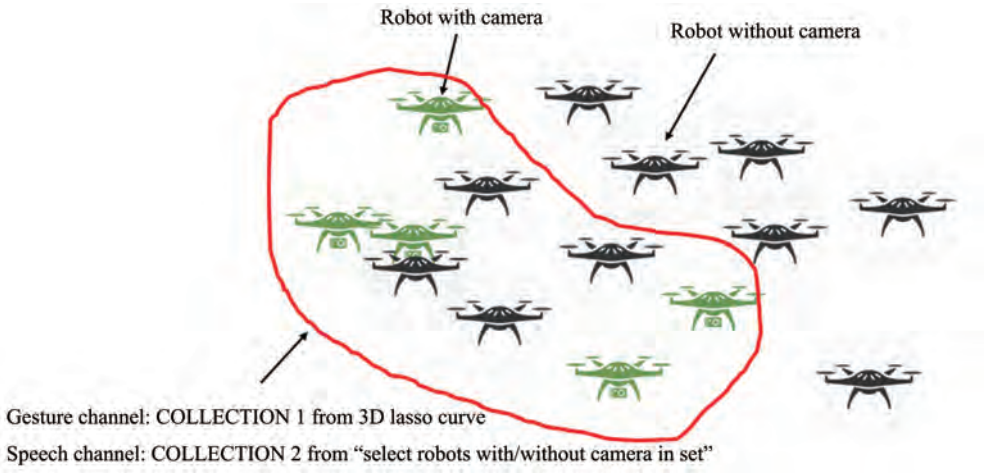

Figure 11 Extend multichannel selection instruction.

5 Experiment

5.1 Experimental design and experimental environment

Based on the proposed multichannel human-group robot natural interaction system with augmented reality, a prototype system is built, which realizes the functions of robot selection, coordinate position control, and motion trajectory control. In order to verify the feasibility and practicability of the robot selection interaction, this paper designs three virtual scene swarm robot selection experiments and one real scene swarm robot interaction experiment. As shown in Figure 12, the virtual scene swarm robot selection experiment places a number of virtual robots in the interactive scene, and sets selection targets. The controller is required to select the target virtual robots through the method presented above. This experiment verifies the efficiency based on the interaction time, and verifies the accuracy rate based on the selection of robot. Five participants (no females, mean age 24) volunteered for our experiment. All of them are postgraduate students without prior experience in interacting with the system. Twenty minutes of familiarization time was given before the experiments.

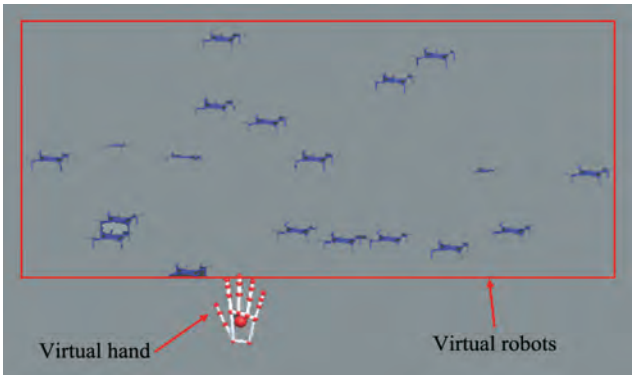

Figure 12 The diagram of swarm robot selection experiment.

Considering the diversity of practical applications, this article sets up three scenarios for the

swarm robot selection experiment. Scenario 1: Selecting a small number of robots in the scene; Scenario 2: Selecting a large number of robots in the scene; and Scenario 3: Selecting a specific robot from a large number of robots in the scene. In order to approach an actual interaction scenario, the target and non-target robots in scene 3 are mixed in space.

Figure 13a is the 3D virtual interactive scene of scenario 1, with three robots to be selected. Figure 13b is the 3D virtual interactive scene of scenario 2, with 20 robots to be selected. Figure 13c is the 3D interactive scene of scenario 3, and the targets to be selected are the red robots among 20 given robots. The initial color of the swarm robots in the scene is blue or red. When selected by the interaction, the color changes to green.

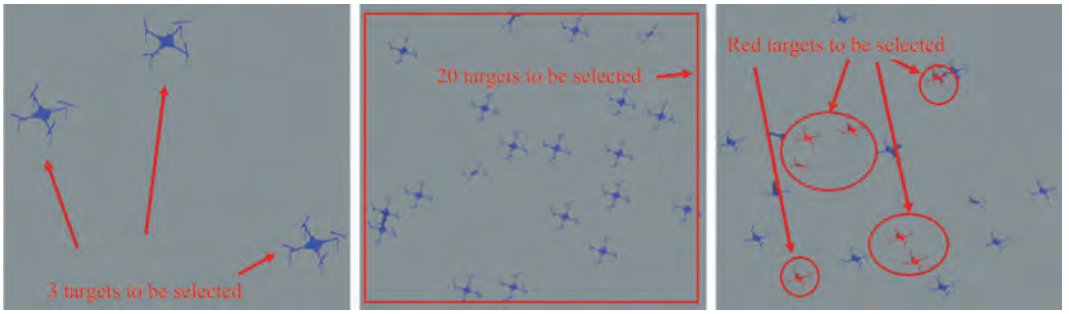

Figure 13 Experimental scenarios 1, 2, and 3.

The real experimental scene of the proposed system is shown in Figure 14a. The participant wore HoloLens on the head, observing the 3D virtual interactive scene in augmented reality. The Leap Motion gesture sensor installed on HoloLens captured the controller's hand information for interaction. The controller's view is shown in Figure 14b. The 3D virtual interactive scene contains virtual hands, a digital map, and virtual swarm robots. The virtual hand is used to map the controller's hand for realizing gesture interaction. A digital map is used to model and display the real scene onto a 3D virtual space. The virtual robots are used as interactive objects, and their positions are updated according to the coordinate information corresponding to the real robots.

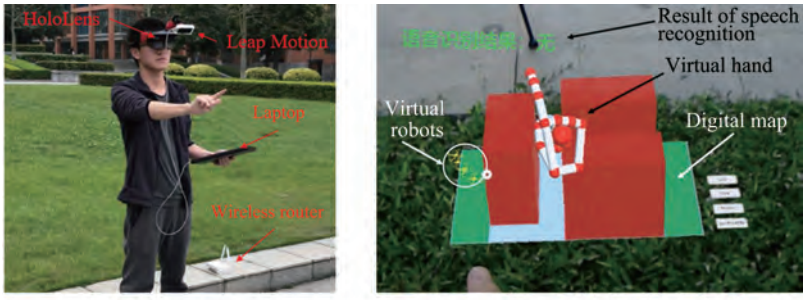

Figure 14 Real experimental scene.

The digital map in the interactive scene is shown in Figure 15a. The size of the digital map is 50cm× 30cm, scaled to a ratio of 1:200 in the practical environment. The upper three red objects are obstacles with the length and width dimensions of 20m×40m and height of 30m. The green rectangular area numbered 4 is the starting area of the group robot, and the green rectangular area numbered 5 is the target reach area. The actual working environment of the swarm robotic system is of size 100m×60m, as shown in Figure 15b. Four DJI M100 quad-rotor drones are used as real robots, measuring 88.5cm in length and 35cm in height (Figure 15c).

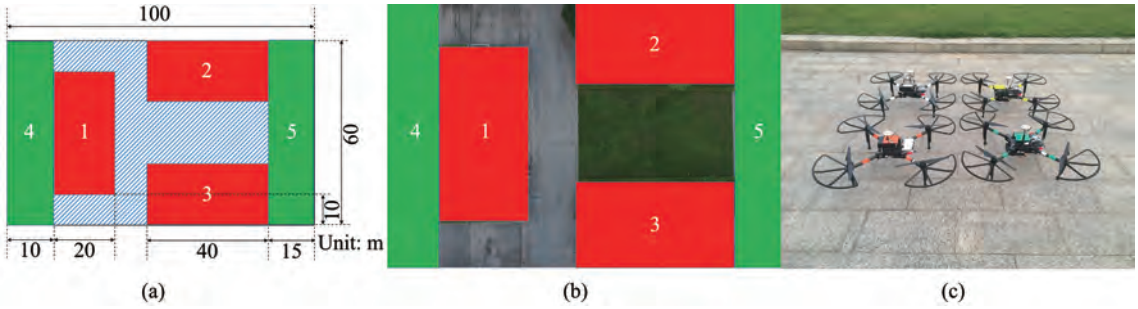

Figure 15 The environment of swarm robots in real experiment.

5.2 Experimental results and analysis

In the virtual experiment of swarm robot selection, participants observed the 3D interactive scene through the HoloLens HMD, and changed the viewing angle through walking and head movement. After moving to a suitable angle, the 3D lasso curve was drawn by hand to include the target robots. Further, the result of the selection was fed back in augmented reality.

In scenarios 1 and 2, participants adjusted the viewing angle by moving their position and turning their head, issuing a speech command "select robots in this range," while surrounding all virtual robots with a 3D lasso. Figures 16a and 16d are the participants in scenarios 1 and 2, respectively, observing the interactive scene and sending selection instructions via speech. Green characters are the result of speech recognition in Chinese (display language, can also choose other languages). Figures 16b and 16e show the controller drawing 3D lasso curves for respectively surrounding all robots in scenarios 1 and 2. Figures 16c and 16f show the feedback of the results after the selection control is executed in scenarios 1 and 2, respectively. The red text is a comment that is added later.

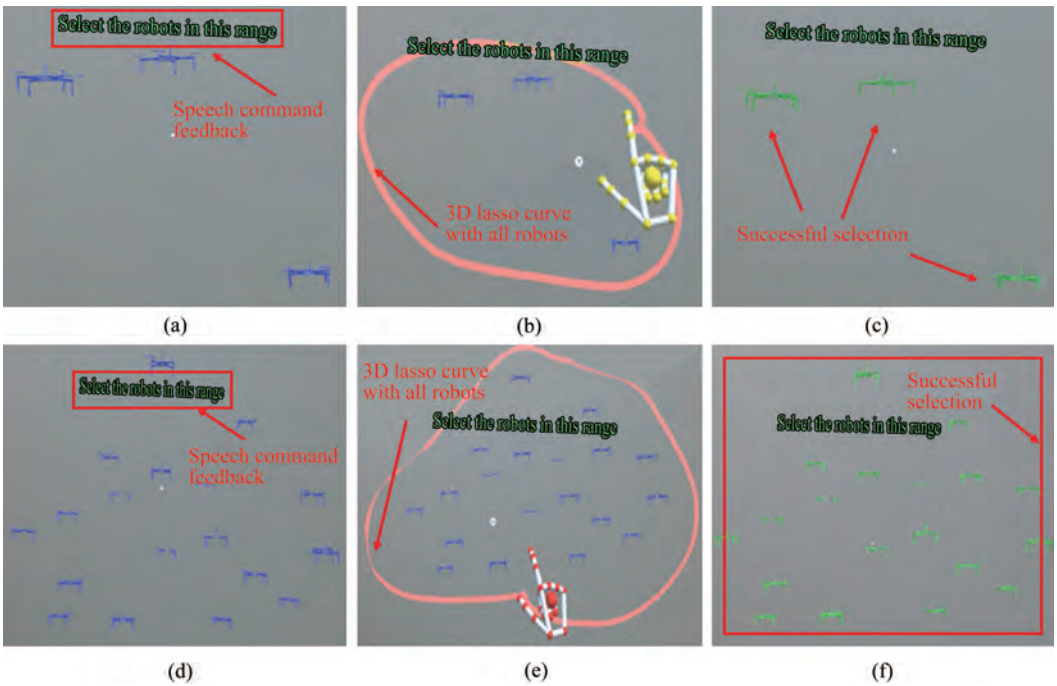

Figure 16 Swarm robot selection process in scenarios 1 and 2.

In Scenario 3, participants were required to select a robot marked as red from among 20 robots. When sending a speech command, a robot set is provided by describing the semantic features of the target robots. Set operations are then performed on the set provided by speech. Finally, target robots are selected (Figure 17).

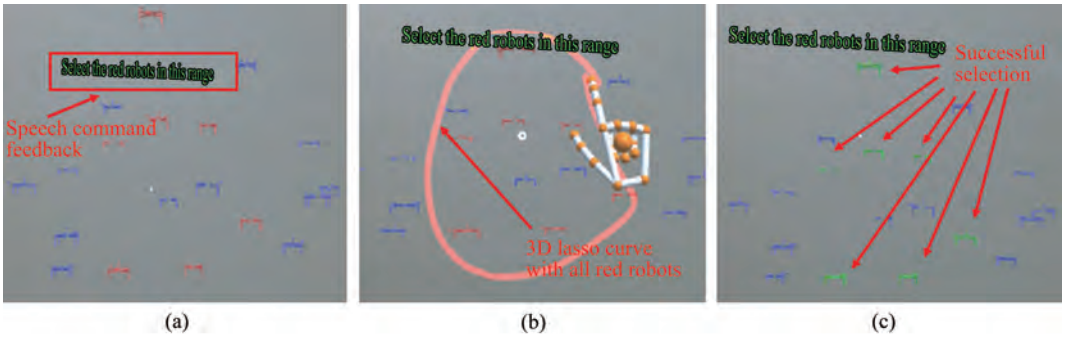

Figure 17 Swarm robot selection process in scenario 3.

In order to measure the performance of the swarm robot's selection interaction, we analyze the proposed method's interactive efficiency and accuracy. To measure interactive efficiency, we measure the time to complete a selection interaction. In order to evaluate the statistical significance of the differences between participants, we analyzed the completion time using single factor repeated-measures ANOVA, with a chosen significance level of $p < 0.05$. In order to measure the accuracy of the selection interaction, we introduced inclusion rate, error rate, and interaction success rate. The inclusion rate is defined as the ratio of the number of target robots selected to the number of target robots. The error rate is the ratio of the number of non-target robots selected to the number of non-target robots. Interaction success rate is defined as the ratio of successful selection interactions to the total number of interactions.

When the proposed multichannel selection instruction is used for the selection task of virtual scenario 3, participants used a 3D lasso containing all target robots and some non-target robots to preliminarily limit the set of selected robots. Moreover, at the same time, the constraint that the target robots should be red is provided via speech. The robot selection set generated by the above two constraints realizes the selection of the target robots through the intersection operation. As shown in Table 2, 20 groups of selection interaction experiments were performed (20 times for each participant). According to the ANOVA results, there is a statistically significant difference in the completion time for different participants, $p = 5.94\times 10^{-8} < 0.05$. The ANOVA showed that different people use the system with different interactive efficiency. However, the average interaction times are approximate (participant 1: 5.068s, participant 2: 5.852s, participant 3: 5.697s, participant 4: 5.725s, and participant 5: 5.133s). The average interaction time of the proposed multichannel selection method was 5.495s. The average inclusion rate was 99.6%, the average error rate was 0, and the interaction success rate was 95%. This is because the controller judges through his own experience that the target robots have a special semantic attribute, while the non-target object does not. Therefore, a precise robot set provided via speech improves the efficiency and accuracy of interaction.

Table 2 The result of swarm robot selection interaction experiment

Interactive method	Average time(s)	Average inclusion rate(%)	Average error rate(%)	Interaction success rate(%)
Multichannel selection instructions	5.495	99.6	0	95
Lasso selection instructions	10.835	91.9	6.25	20

6 Conclusion

The proposed multichannel HSI system with augmented reality realizes interactions such as robot selection, trajectory motion, and state control. Among them, the 3D gesture interaction channel is used to realize the batch picking of swarm robots. The natural language instruction channel is used to realize the natural language instruction understanding. Finally, augmented reality is used for immersive visual

feedback. Instruction tag filling achieves information complementation and collaboration between the 3D gesture interaction channel and the natural language instruction channel, which makes the interaction between humans and swarm robots natural and efficient.

The multichannel interaction method based on 3D gesture interaction and natural language instructions proposed in this paper provides new ideas for the research on HSI. However, there are some shortcomings of our study. Since the gesture sensor used by the system is installed on a head-mounted augmented reality device, its working range changes with the movement of the controller's head. Moreover, it is easy for the operator to exceed the detection range. Therefore, we have to increase the number of sensors or install the sensors at positions that remain stationary during the interaction, such as on shoulders, for reliable and stable gesture data acquisition. In addition, the robot selection instruction requires the target robots to have a distinguishing feature. Moreover, the feature must be predefined in the natural language instruction understanding control instruction library; thus, the application scenario has limitations, which need to be addressed in future studies.

Declaration of competing interest

We declare that we have no conflict of interest.

References

1 Alfeo A L, Cimino M G C A, de Francesco N, Lazzeri A, Lega M, Vaglini G. Swarm coordination of mini-UAVs for target search using imperfect sensors. Intelligent Decision Technologies, 2018, 12(2): 149–162
DOI:10.3233/idt-170317

2 Li K, Ni W, Wang X, Liu R P, Kanhere S S, Jha S. Energy-efficient cooperative relaying for unmanned aerial vehicles. IEEE Transactions on Mobile Computing, 2016, 15(6): 1377–1386
DOI:10.1109/tmc.2015.2467381

3 Zhang Q, Gong Z K, Yang Z Q, Chen Z Q. Distributed convex optimization for flocking of nonlinear multi-agent systems. International Journal of Control, Automation and Systems, 2019, 17(5): 1177–1183
DOI:10.1007/s12555-018-0191-x

4 Krause J, Ruxton G D, Krause S. Swarm intelligence in animals and humans. Trends in Ecology & Evolution, 2010, 25 (1): 28–34
DOI:10.1016/j.tree.2009.06.016

5 Vassev E, Hinchey M, Nixon P. A formal approach to self-configurable swarm-based space-exploration systems. 2010 NASA/ESA Conference on Adaptive Hardware and Systems. Anaheim, CA, USA, IEEE, 2010, 83–90
DOI:10.1109/ahs.2010.5546276

6 Kolling A, Walker P, Chakraborty N, Sycara K, Lewis M. Human interaction with robot swarms: a survey. IEEE Transactions on Human-Machine Systems, 2016, 46(1): 9–26
DOI:10.1109/thms.2015.2480801

7 Krishnamurthy P, Khorrami F. A distributed monitoring approach for human interaction with multi-robot systems. In: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. Vienna Austria, New York, NY, USA, ACM, 2017
DOI:10.1145/3029798.3038327

8 Savla K, Frazzoli E. A dynamical queue approach to intelligent task management for human operators. Proceedings of the IEEE, 2012, 100(3): 672–686
DOI:10.1109/jproc.2011.2173264

9 Setter T, Fouraker A, Kawashima H, Egerstedt M. Haptic interactions with multi-robot swarms using manipulability. Journal of Human-Robot Interaction, 2015, 4(1): 60–74
DOI:10.5898/jhri.4.1.setter

10 Franchi A, Secchi C, Ryll M, Bulthoff H, Giordano P. Shared control: balancing autonomy and human assistance with a

group of quadrotor UAVs. IEEE Robotics & Automation Magazine, 2012, 19(3): 57−68
DOI:10.1109/mra.2012.2205625

11 Wang Z J, Schwager M. Kinematic multi-robot manipulation with no communication using force feedback. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). Stockholm, Sweden, IEEE, 2016, 427−432
DOI:10.1109/icra.2016.7487163

12 Gromov B, Gambardella L M, di Caro G A. Wearable multi-modal interface for human multi-robot interaction. In: 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). Lausanne, Switzerland, IEEE, 2016, 240−245
DOI:10.1109/ssrr.2016.7784305

13 Erat O, Isop W A, Kalkofen D, Schmalstieg D. Drone-augmented human vision: exocentric control for drones exploring hidden areas. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(4): 1437−1446
DOI:10.1109/tvcg.2018.2794058

14 Tsykunov E, Agishev R, Ibrahimov R, Labazanova L, Tleugazy A, Tsetserukou D. SwarmTouch: guiding a swarm of micro-quadrotors with impedance control using a wearable tactile interface. IEEE Transactions on Haptics, 2019, 12(3): 363−374
DOI:10.1109/toh.2019.2927338

15 Zhang Y H, Du Y, Pan F, Wei Y. Intelligent vehicle path tracking algorithm based on cubic B-spline curve fitting. Journal of Computer Applications, 2018, 38(6): 1562−1567(in Chinese)

16 Zhai Y, Xu W Y, Zhang Q. Judgment of topological relation between point and polygon or polyhedron. Computer Engineering and Design, 2015, 36(4): 972−976(in Chinese)
DOI:10.16208/j.issn1000-7024.2015.04.026

17 Du G L, Chen M X, Liu C B, Zhang B, Zhang P. Online robot teaching with natural human−robot interaction. IEEE Transactions on Industrial Electronics, 2018, 65(12): 9571−9581
DOI:10.1109/tie.2018.2823667