Intelligent Service Robotics

Human-Manipulator Interaction Technology of Dual-Robot Based on Hands Gesture Control

--Manuscript Draft--

Manuscript Number:	JIST-D-17-00062
Full Title:	Human-Manipulator Interaction Technology of Dual-Robot Based on Hands Gesture Control
Article Type:	Original Research
Keywords:	Dual-Robot; hands HCI; hybrid sensor; low-pass filter; Gaussian Mixture Regression Model
Abstract:	The human-manipulator interaction technology based on hands gesture control has the advantages of rich expression, general adaptability, and comfortable interaction, and it is expected to be a new generation of hotspot. Based on the study of one-hand gesture interaction in the early stage of the laboratory, this paper proposes an interaction of two-hand gesture control. The main research contents are as follows: Firstly, the use of the combination of Kinect and Leap Motion hybrid sensor to collect position and motion data of hands, effectively solve the problem of occluded gesture detection. Then the iterated algorithm based of orthogonal matrix is used to calibrate and transform the local coordinate system of hybrid sensor, then process the data with adaptive low-pass filter algorithm based on speed and adaptive weighted fusion to get the final hands gesture. As for the lost gesture prediction, we adopted the Gaussian mixture regression model. Then the adaptive multi-space transformation algorithm was used to transform gesture into motion data. Lastly, the capsule bounding box collision detection algorithm was used to detect robots' collision. At the end of this paper, the feasibility and effectiveness of the research are verified by the experiment of double-robot gesture control.

Human-Manipulator Interaction Technology of Dual-Robot Based on Hands Gesture Control

***Abstract*—The human-manipulator interaction technology based on hands gesture control has the advantages of rich expression, general adaptability, and comfortable interaction, and it is expected to be a new generation of hotspot. Based on the study of one-hand gesture interaction in the early stage of the laboratory, this paper proposes an interaction of two-hand gesture control. The main research contents are as follows: Firstly, the use of the combination of Kinect and Leap Motion hybrid sensor to collect position and motion data of hands, effectively solve the problem of occluded gesture detection. Then the iterated algorithm based on orthogonal matrix is used to calibrate and transform the local coordinate system of hybrid sensor, then process the data with adaptive low-pass filter algorithm based on speed and adaptive weighted fusion to get the final hands gesture. As for the lost gesture prediction, we adopted the Gaussian mixture regression model. Then the adaptive multi-space transformation algorithm was used to transform gesture into motion data. Lastly, the capsule bounding box collision detection algorithm was used to detect robots' collision. At the end of this paper, the feasibility and effectiveness of the research are verified by the experiment of double-robot gesture control.**

Keywords: Dual-Robot; hands HCI; hybrid sensor; low-pass filter; Gaussian Mixture Regression Model

I. Introduction

HUMAN-MANIPULATOR interaction refers to the acquisition of human information through some input devices, so as to transform human intent into instructions that robots can understand, and achieve the intention of dialogue with the robot, and then control the robot movement. Intelligent perception technology is one of the key and core technologies in the human-manipulator interaction technology. It is mainly through the simulation of human biological characteristics based on natural language understanding, image understanding and other related technologies to identify and transform human intent to achieve the intelligent information processing and controlling [1][2]. Mainstream robot human-manipulator interaction uses various types of sensors to simulate human perception, so as to collect human information and build the relationship between people and robots, and the environment and robots. Depending on the ways the human controlling the robot, human-manipulator interaction can be divided into gesture-based control, voice control, expression control, brain wave signal control and so on [3][4][5].

Compared with voice control, expression control, and brain wave signal control, gesture-based control has a greater advantage on flexibility and accuracy. However, there is still a lot of space for development in the real work and life, such as detecting occluded gestures, providing data stability, etc.

At present, human-manipulator interaction based on gesture control can be divided into wear-based human-manipulator interaction technology and vision-based human-manipulator interaction technology according to the different methods of sensor acquisition.

Vision-based human-manipulator interaction mainly uses color images or depth images to identify real-time gestures. Depending on the mode of interaction, vision-based gesture interaction can be divided into two categories [6], one of which pre-defines gestures by computer, and each gesture corresponds to a motion control command. During the operator's interactive process, the sensor identifies the operator's gestures and uses relevant algorithm to match the gestures with gestures from library system, then analyses the gesture information into related motion control commands. Finally the commands are sent to control the robot [7].

Another one is the direct control type. In this type, the robot directly accessing the hands position and their gesture changing data by depth sensor (e.g. Leap Motion, Kinect), which converted data to robot's position and gesture changing instruction, and then controlled the movement of the robot [8][9]. This kind of interactive way is intuitive and convenient. The operator does not need to memorize the complex and cumbersome gesture database, and does not need to be pre-trained, instead, they can control the movement state of the robot qualitatively and quantitatively. Therefore this type of human-manipulator interaction has a great prospect for development.

In the aspect of vision-based human-manipulator interaction, the domestic and foreign researchers are focused on a single hand gesture interaction. In the process of studying the transferring objects between human and robot, Kobayashi Futoshi, *et al* [10] used Leap Motion sensor to identify the position and gesture changes of single hand in real time. The robot's end position and gesture will be based on the location and gesture of the hand to make the appropriate changes to achieve the best object transfer action. In contrast, there are few studies in both hands gesture interaction. G Du and P Zhang [11] identified the gesture data for both hands by the Leap Motion, then the data is handled by Particle filter and Kalman filter to do the remote interaction operation of dual robots. However, they only used one Leap Motion sensor, which can't detect the data when the hands are blocked, so their research needed to reset the hand gestures to avoid the problem. As for research on occlusion of hands gesture, A. Utsumi and J. Ohya [12] used a multi-angle camera to identify the occluded hand gestures. They selected images without any occlusion from the images captured by multiple angles cameras. Chen Bangmin [13] split up the hand images according to the skin color parameters, and used the centroid estimation method to predict the gestures in the case of both hands occluded.

In summary, as for the interaction of hands gestures, on the one hand, it simplifies the process of interaction by avoiding the occlusion of both hands, on the other hand, it estimates gestures with occlusion through the two-dimensional images, but it is difficult to track the position and gestures of the hands in real time. To deal with these problems, the research of this paper mainly uses a simple multi-angle depth camera method to stably identify the three-dimensional position and gestures of the hands in any situation for a long time, and it is simple and easy to expand, and it can effectively solve the tracking and recognizing problem of the occlusion of both hands.

II. Design of Hybrid Sensor Platform

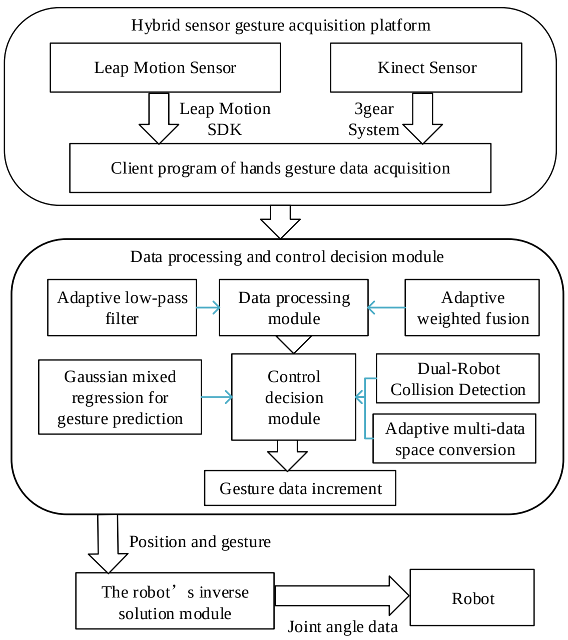


Fig. 1. HMI system framework platform.

Hands gesture interaction is different from single-handed gesture interaction. Mutual occlusion is prone to occur in the process of hands gesture interaction, and a single sensor can't deal with the problem of identifying the occlusion gestures. At present, the mainstream practice of blocking gesture recognition is use machine learning method to identify some specific gestures [14] as well as using estimated and predicted methods to identify the occlusion gestures [13]. But these methods are not able to identify hands gesture stably for a long time. The current mainstream sensors used for real-time tracking recognition gestures include Leap Motion and Kinect, etc. Different types of sensor combinations and different placement position combination can form different solutions. This paper focuses on a comparative analysis of the following three options:

1) Combination of two Kinect sensors, which were placed over the desktop and detected downward. This program was easy to be implemented under the framework of 3Gear System, which can effectively detect a specific obscured gesture in some extent. But the placement effect of the scheme has some limitations, and it can't effectively identify completely occluded objects in the vertical direction, which just like the human binocular visual perception effect.

2) Combination of two Leap Motion sensors, the placement positions can be set in opposite directions: the vertical direction and the orthogonal direction, which means that one Leap Motion was placed on the desktop. Another one was placed on the desktop or in the side direction which was formed above the angle of 90 degrees. Due to the small detection range of the Leap Motion sensor, when the Leap Motion sensors detected closely in the opposite direction, it is prone to different infrared interference, resulting in poor recognition stability.

3) Mixed combination of Leap Motion and Kinect, Leap Motion was placed on the desktop, detecting upward. Kinect was placed over the desktop, detecting downward. This program can effectively solve any case of hands gesture's occlusion, and Kinect sensor is able to detect the distance, which is different from Leap Motion, so the infrared jamming of both sensors carried relatively low impact to the gesture recognition. In addition to that, the range of the detection area of this program is larger than that of option two.

Fig. 2. The hybrid sensor platform based on Leap Motion and Kinect.

To summarize, this paper used option three, the vertical placement scheme, which combines Leap Motion and Kinect 3, to recognize hands gestures, as shown in Fig. 2. Considering the practical application of robot interaction, this paper focuses on the occlusion problem of both hands in the vertical direction. However, the techniques and schemes studied in this paper are suitable for extending ways to solve arbitrary direction occlusion. For gesture occlusion, when the hands gesture does not have an occlusion in the vertical direction, Leap Motion and Kinect can both recognize the position and gestures of the hands, and at this time we need to combine two sensors' data as a final hands gesture data. When the occlusion of hands gesture occurs in the vertical direction, the lower end of the gesture data can be recognized by the Leap Motion sensor below, and the upper gesture data can also be detected by the Kinect sensor. Eventually, all of gesture data are converted to the gesture global coordinate system (In this article we take the local coordinate system of Leap Motion sensor as a gesture global coordinate system), so it can be used as the final hands gesture data. Since Leap Motion sensor and Kinect sensor can stably recognize gesture for a long time, the hybrid sensor platform proposed in this paper can stably and accurately identify any hands gestures for a long time.

III. MULTI-SENSOR COORDINATE SYSTEM TRANSFORMATION

This paper is researching the hybrid sensor gesture recognition system based on Leap Motion and Kinect. Kinect and Leap Motion both have local coordinate system based on the sensor itself. The data identifying the hand gesture is also based on the local coordinate system, and the data obtained by the different sensors need to be unified after filtering to the global coordinate system. We used the Leap Motion coordinate system as the global coordinate system of hands gesture data, and therefore we need to transform the data of Kinect local coordinate system into that of Leap Motion coordinate system. The iterative algorithm based on orthogonal matrix is used to solve the parameters of the coordinate system transformation. The basic idea is based on the least square method. The following model is described in detail for the changes in the Leap Motion and Kinect coordinate systems:

Assume that the hand position data of the N samples are collected in advance and the positions of the hand measured by the Leap Motion sensor and the Kinect sensor are the same point, therefore, the coordinates of the N three-dimensional common point sets in the Leap Motion and Kinect sensors are equal to $\{L_i\}$ and $\{K_i\}, i = 1, 2, \ldots, N$. The points in the Kinect coordinate system are transformed into the points in the Leap Motion coordinate system, and the coordinate transformation model can be constructed as follows [15]:

$$L_i = T + \mu R K_i \tag{1}$$

T represents the translation matrix parameter, μ represents the scale parameter and R denotes the rotation matrix parameter. Since Kinect and Leap Motion belong to the right-hand coordinate system, the rotation matrix satisfies the orthogonal matrix condition constraint:

$$R = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \Rightarrow \begin{cases} a_1 a_2 + b_1 b_2 + c_1 c_2 = 0 \\ a_1 a_3 + b_1 b_3 + c_1 c_3 = 0 \\ a_2 a_3 + b_2 b_3 + c_2 c_3 = 0 \\ a_1^2 + a_2^2 + a_3^2 = 1 \\ b_1^2 + b_2^2 + b_3^2 = 1 \\ c_1^2 + c_2^2 + c_3^2 = 1 \end{cases} \tag{2}$$

For the common point set of multiple samples, the basic idea of the least squares method is trying to make the equation (1) possible by estimating the optimal set of parameters $\{\hat{T}, \hat{\mu}, \hat{R}\}$, that is to say, the overall mean square error is the smallest, and the objective equation of the optimization model is:

$$min \sum_{i=1}^{N} \left\| L_i - (T + \mu R K_i) \right\|^2 \tag{3}$$

In summary, the method of solving the coordinate system transformation based on the orthogonal matrix is a nonlinear programming model with the formula (3) as the objective equation and formula (2) as the constraint condition. Since the objective equation and the constraint condition are non-linear, it is difficult to solve it directly by Lagrange Multiplier method (adjustment theory). The literature [15] proposed a simple iterative approach to solve the parameters, and the specific process is as follows:

For the formula (1), we set Taylor series expansion in the vicinity of the parameter $\{\hat{T}, \hat{\mu}, \hat{R}\}$ and get the formula (4):

$$L_i = T^0 + \mu^0 R^0 K_i + dT^0 + R^0 K_i d\mu + \mu^0 dRK_i \tag{4}$$

The formula (4) is written in the form of error equations and we get the formula (5):

$$\begin{aligned} V_i &= T^0 + \mu^0 R^0 K_i + dT^0 + R^0 K_i d\mu + \mu^0 dRK_i - L_i \\ &\triangleq P_i X + Q_i \qquad i = 1, 2, \ldots, N \end{aligned} \tag{5}$$

Among them:

V_i represents residual vector and

$$V_i = \begin{bmatrix} V_{x,i} & V_{y,i} & V_{z,i} \end{bmatrix}^T \tag{6}$$

$$X = \begin{bmatrix} \mathrm{dT}_x^0 & \mathrm{dT}_y^0 & \mathrm{dT}_z^0 & d\mu & da_1 & da_2 & da_3 & db_1 & db_2 & db_3 & dc_1 & dc_2 & dc_3 \end{bmatrix}^T \tag{7}$$

$$Q_i = \begin{bmatrix} T_x^0 \\ T_y^0 \\ T_z^0 \end{bmatrix} + \mu^0 \begin{bmatrix} a_1^0 & a_2^0 & a_3^0 \\ b_1^0 & b_2^0 & b_3^0 \\ c_1^0 & c_2^0 & c_3^0 \end{bmatrix} \begin{bmatrix} K_x \\ K_y \\ K_z \end{bmatrix}_i - \begin{bmatrix} L_x \\ L_y \\ L_z \end{bmatrix}_i \tag{8}$$

According to the orthogonal matrix constraint of formula (2), the Taylor series expansion is also carried out in the vicinity of the parameter set. The conditional equation is:

$$BX + W = 0 \tag{9}$$

Among them:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 2a_1^0 & 2a_2^0 & 2a_3^0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2b_1^0 & 2b_2^0 & 2b_2^0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2c_1^0 & 2c_2^0 & 2c_3^0 \\ 0 & 0 & 0 & 0 & a_2^0 & a_1^0 & 0 & b_2^0 & b_1^0 & 0 & c_2^0 & c_1^0 & 0 \\ 0 & 0 & 0 & 0 & a_3^0 & 0 & a_1^0 & b_3^0 & 0 & b_1^0 & c_3^0 & 0 & c_1^0 \\ 0 & 0 & 0 & 0 & 0 & a_3^0 & a_2^0 & 0 & b_3^0 & b_2^0 & 0 & c_3^0 & c_2^0 \end{bmatrix} \tag{10}$$

$$X = \begin{bmatrix} \mathrm{dT}_x^0 & \mathrm{dT}_y^0 & \mathrm{dT}_z^0 & d\mu & da_1 & da_2 & da_3 & db_1 & db_2 & db_3 & dc_1 & dc_2 & dc_3 \end{bmatrix}^T \tag{11}$$

$$W = \begin{bmatrix} a_1^{02} + a_2^{02} + a_3^{02} - 1 \\ b_1^{02} + b_2^{02} + b_3^{02} - 1 \\ c_1^{02} + c_2^{02} + c_3^{02} - 1 \\ a_1^0 a_2^0 + b_1^0 b_2^0 + c_1^0 c_2^0 \\ a_1^0 a_3^0 + b_1^0 b_3^0 + c_1^0 c_3^0 \\ a_2^0 a_3^0 + b_2^0 b_3^0 + c_2^0 c_3^0 \end{bmatrix} \tag{12}$$

Combining with the error equation of formula (5) and the conditional equation of formula (9), we can obtain the corresponding X by using the conditional indirect adjustment method. However, in the actual calculation, the conditional equation is usually transformed into a pseudo-observation equation, and given the power, and then as for the pseudo-observation equation, we use conventional indirect adjustment method to solve the equation efficiently. In this paper, the error weight of the observation equation is set to 1, and the error weight of the pseudo-observation equation is set to 100 in order to solve the problem.

In summary, the main algorithm steps for solving the model parameters using the iterative algorithm based on orthogonal form are:

1) Initialize the parameter set

$$\left\{T_x^0, T_y^0, T_z^0, \mu^0, a_1^0, a_2^0, a_3^0, b_1^0, b_2^0, b_3^0, c_1^0, c_2^0, c_3^0\right\} \tag{13}$$

Under normal circumstances, the following parameters are desirable:

$$\begin{bmatrix} T_x^0 \\ T_y^0 \\ T_z^0 \end{bmatrix} = \begin{bmatrix} \frac{1}{N}(\sum_{i=1}^{N} K_{x,i} - \sum_{i=1}^{N} L_{x,i}) \\ \frac{1}{N}(\sum_{i=1}^{N} K_{y,i} - \sum_{i=1}^{N} L_{y,i}) \\ \frac{1}{N}(\sum_{i=1}^{N} K_{z,i} - \sum_{i=1}^{N} L_{z,i}) \end{bmatrix}, \mu^0 = 1, \begin{bmatrix} a_1^0 & a_2^0 & a_3^0 \\ b_1^0 & b_2^0 & b_3^0 \\ c_1^0 & c_2^0 & c_3^0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

2) According to the formula (5) we get *3N* error equations, and according to formula (9) we get *N* conditional equations;

3) The conditional equation is transformed into the pseudo - observation equation, and the power is given, and the indirect adjustment method is used to solve the correction quantity of the parameter set.

4) Calculate and update the latest values of the parameter set based on the corrections;

5) According to the size of the correction, we need to determine whether to meet the convergence requirements. If it is not satisfied, repeating steps 2) ~ 4) until the convergence requirements are met.

IV. Fusion of Hands Gesture Data

This paper uses a hybrid sensor platform based on Leap Motion and Kinect to identify both hands gestures. When there is an occlusion of hands gesture, the position and gesture of each hand can only be detected by one of the sensors in the hybrid sensor platform. The lower hand can only be recognized by Leap Motion, and the upper hand can only be detected by Kinect, so we use the data to do the filtering and coordinate system transformation to get the final hands position. In order to further improve the accuracy of gesture recognition, we need to consider the two sensors' data, and make fusion of their data. This paper uses a self-adaptive weight fusion algorithm [17], whose idea is based on the principle of minimizing the mean squared error, and the optimal weighting factor of each sensor is found in an adaptive way, so that the target value can be optimal after the fusion is complete.

Assume that the gesture detection of both hands is at the same time, the measured value of the Leap Motion sensor is z_L after filtering, and Kinect sensor is filtered and transformed to the measurement of z_K in the Leap Motion coordinate system, and the real gesture data of hand is x. Assume that the measurements are independent of each other and the random measurement errors are v_L, v_K, and the errors obey the Gaussian distribution with the mean 0 and the variances are σ_L^2, σ_K^2, so the following formula can be obtained:

$$\begin{cases} Z_L = x + v_L \\ Z_K = x + v_K \end{cases} \tag{15}$$

Assume that we use linear weighted fusion, the estimated position of the hand gesture is:

$$\hat{x} = w_L z_L + w_K z_K \tag{16}$$

$\{w_L, w_K\}$ were the weight measurements of Leap Motion and Kinect respectively.

The main principle of self-adaptive weighted fusion algorithm is as follow: Under the premise of $\hat{x}$ is the unbiased estimation of x, the estimation error is optimized so that the estimated mean square error is minimized. Assume that the formula of estimation error is $\tilde{x} = x - \hat{x}$, so the model of the self-adaptive weight fusion algorithm can be expressed as follows:

$$\begin{cases} min \quad E(\tilde{x}^2) \\ s.t. \quad E(\tilde{x}) = 0 \end{cases} \tag{17}$$

After the constraint condition is expanded:

$$E(\tilde{x}) = E(x - w_L(x+v_L) - w_K(x+v_K)) = 0 \Rightarrow w_L + w_K = 1 \tag{18}$$

The derived result of the formula (18) is substituted into the objective function of formula (17):

$$\begin{aligned} E(\tilde{x}^2) &= E((x - w_L(x+v_L) - w_K(x+v_K))^2) \\ &= E({w_L}^2 {v_L}^2 + (1-w_L)^2 {v_K}^2 + 2w_L(1-w_L)v_L v_K) \\ &= {w_L}^2 \sigma_L^2 + (1-w_L)^2 \sigma_K^2 \end{aligned} \tag{19}$$

If we get the partial derivatives of *X* from formula (19), we can get the minimum value of the weight from the target equation, the optimal solution is:

$$\frac{\partial E(\tilde{x}^2)}{\partial w_L} = 0 \Rightarrow \begin{cases} w_L = \dfrac{1}{\sigma_L^2(\frac{1}{\sigma_L^2} + \frac{1}{\sigma_K^2})} \\ w_K = \dfrac{1}{\sigma_K^2(\frac{1}{\sigma_L^2} + \frac{1}{\sigma_K^2})} \end{cases} \tag{20}$$

From the above equations, we can see that the optimal weight factor depends on the variance of the measurement error from each sensor. The initial variance can be obtained according to the measured value of the sensor in advance by the following algorithm:

Since the random measurement error variables of Leap Motion and Kinect are independent from each other, according to the time domain estimation method [18], the variance of measurement errors can be calculated by the following formula:

$$\begin{cases} \sigma_L^2 = E(v_L^2) = R_{LL} - R_{LK} \\ \sigma_K^2 = E(v_K^2) = R_{KK} - R_{KL} \end{cases} \tag{21}$$

R_{LL} is the autocovariance function of z_L, R_{KK} is the autocorrelation function of z_K, R_{LK}, R_{KL} are the cross covariance function of z_L, z_K respectively, and the values are equal.

Assume that the number of sensor measurements at current time is *n*, the estimated time domain of R_{LL} is $R_{LL}(n)$, the estimated time domain of R_{KK} is $R_{KK}(n)$,the estimated time domain of R_{LK} is $R_{LK}(n)$, so we can conclude that:

$$\begin{cases} R_{LL}(n)=\frac{1}{n}\sum_{i=1}^{n}(z_L(i)-\mu)(z_L(i)-\mu)=\frac{n-1}{n}R_{LL}(n-1)+\frac{1}{n}(z_L(n)-\mu)(z_L(n)-\mu) \\ R_{KK}(n)=\frac{1}{n}\sum_{i=1}^{n}(z_K(i)-\mu)(z_K(i)-\mu)=\frac{n-1}{n}R_{KK}(n-1)+\frac{1}{n}(z_K(n)-\mu)(z_K(n)-\mu) \\ R_{LK}(n)=\frac{1}{n}\sum_{i=1}^{n}(z_L(i)-\mu)(z_K(i)-\mu)=\frac{n-1}{n}R_{LK}(n-1)+\frac{1}{n}(z_L(n)-\mu)(z_K(n)-\mu) \end{cases} \quad (22)$$

μ is the mean value of the sampled data, $\mu=\frac{z_L+z_K}{2}$. Therefore, the variance of the measurement error in the Leap Motion and Kinect sensors can be obtained from the time domain estimation.

In practice, due to environmental interference, equipment reliability and other reasons, the variance of the measurement error of the sensor is not fixed, and the fusion algorithm can also adjust the variance of the sensor online. Therefore, in this experiment, before the experiment of human-manipulator interaction began, we let Leap Motion and Kinect sensor to collect 100 frames of data. When the data is becoming stable, we started the human-manipulator interaction experiment. The estimated time domain of the current time R_{LL}, R_{KK}, R_{LK} is calculated according to the above formula, and it is regarded as the initial estimated time domain of human-manipulator interaction experiment.

As the interactive experiments carried on, the continuous collection of data, and constantly calculating the variance of the Leap Motion and Kinect sensors at each frame, and the self-adaptive adjustment of the weight of each sensor ensured that the entire human-manipulator interaction experiment has a great accuracy of the data fusion.

V. EXPERIMENT

A. Design of Experiment

In the experiment of this paper, we collected the 200-frame three-dimensional position based on the Leap Motion and Kinect hybrid sensor platforms which proposed in this paper, and then we used the common point set to train the parameters. The convergence condition is set as the correction of the translation parameters in the range of 10^{-3}mm, the correction of the scale parameter is in the range of 10^{-7}, the correction amount of the rotation matrix parameter is in the range of 10^{-10} [16]. Finally, the calculation results of the iteration and convergence are:

$$\begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = \begin{bmatrix} -41.9472 \\ -32.0917 \\ 95.6835 \end{bmatrix}, \mu = 1.000909, R = \begin{bmatrix} 0.96749 & -0.00291 & 0.07691 \\ 0.03366 & 0.99403 & -0.04566 \\ 0.03729 & -0.08361 & 1.02158 \end{bmatrix} \quad (23)$$

For Kinect's hand position data $(x,y,z,\alpha,\beta,\gamma)^T$, it is transformed into a 4x4 pose matrix, and then multiplied by the coordinate transformation matrix which is shown above, and we can get the position matrix $\tilde{A}$ in the Leap Motion coordinate system. The formula is shown as follow:

$$\tilde{A}_{4\times4} = \begin{bmatrix} \mu R_{(3\times3)} & T_{(3\times1)} \\ 0 & 1 \end{bmatrix} A_{(4\times4)} \quad (24)$$

The dual-robot cooperative screw-tightening experimental platform is composed of human, robot and human-manipulator interaction system based on hands gesture. The human-manipulator interaction system is a bridge between people and robots, which can do real-time detection of hands gestures through a hybrid sensor platform based on Leap Motion and Kinect. After running data processing module and control decision module, the system drives the dual-robots to complete the screw-tightening test.

The real scene of the dual-robot cooperative screw-tightening experiment is shown in Fig. 3. The combination of the Leap Motion sensor placed in the surface of the table and the Kinect sensor placed over the table is called the gesture recognition platform based on hybrid sensor, in which we need to operate robots in the intermediate range. The left hand is used to control the movement of the left robot and the right hand is used to control the movement of the right robot.

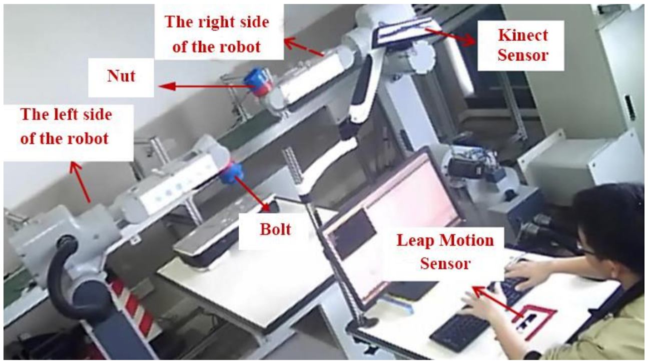


Fig. 3. The real scene of the dual-robot cooperative screwing experiment.

In this experiment, we used other method described in paper[11] as a comparison, which only uses one Leap Motion to collect hands gesture, so as to verify the effectiveness of the method proposed in this paper.

The development environment of Leap Motion sensor is SDK2.3.1 version, and the first generation of Kinect sensor, and the development environment of 3Gear System is v0.9.36 version. The two robots with six degrees of freedom are the GRB3016 robots from Googol Technology Limited, and D-H parameters are shown as TABLE I, the nominal diameter of the nut and bolt is 75.5 mm and 75 mm, and screw thread lengths are 30mm and 35mm, respectively, as shown in Fig. 4 below:

TABLE II
THE DH PARAMETERS OF 6-DOF ROBOT

Joint	a(mm)	α(rad)	d(mm)	θ(rad)
1	150	-π/2	250	0
2	570	-π	0	-π/2
3	150	π/2	0	0
4	0	-π/2	650	0
5	0	-π/2	0	-π/2

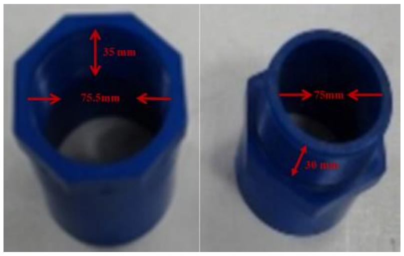
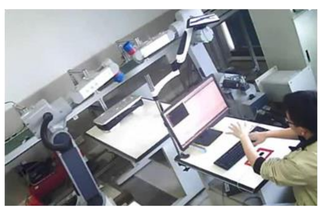


Fig.4. nut (left side) and bolt (right side).

B. The Experiment Results and Analysis

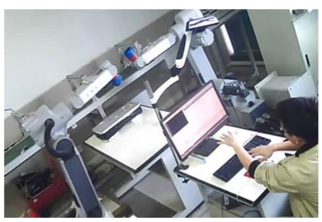
A key technique for human-manipulator interaction based on hands gesture control is the real-time detection of hands gestures with the presence of occlusion. Fig. 5 shows the process of human-manipulator interaction when there is an occlusion of hands gesture. Fig. 5.(a) shows that the hands are in a separate state, each hand gesture can be identified by Leap Motion sensor and Kinect sensor; Fig. 5.(b) represents after moving a distance, there will be a mutual occlusion of the hands gestures in the vertical direction; Fig. 5.(c) and Fig. 5.(d) indicates that human-manipulator interaction platform is still able to detect up and down movement and front and rear movement of both hands when there is an occlusion of hands gesture. At this time, the upper hand gestures are identified by the Kinect sensor, and the lower hand gestures are identified by the Leap Motion sensor.

(a) Hands separated state



) Hands are covered

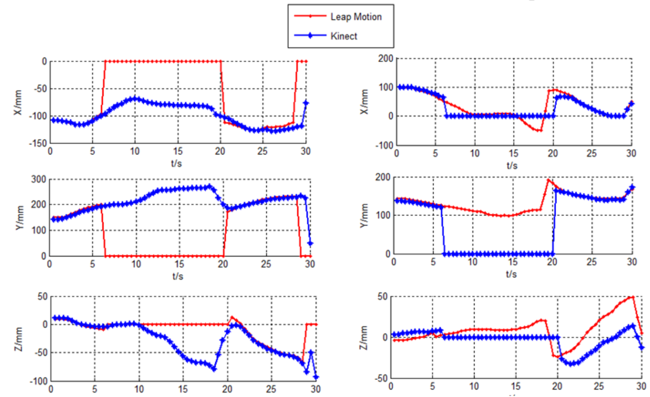
(c) Hands up and down when they are blocked



(d) Hands move back and forth when they are blocked

Fig. 5. Hands gestures movement when there is an occlusion of hands gesture.

Fig. 6 shows the hands tracking effect of Leap Motion sensor and Kinect sensor when there is an occlusion of hands gesture. Fig. 6.(a) indicates the trajectory of each axis of the left hand, and Fig. 6.(b) indicates the trajectory of each axis of the right hand. It can be seen that in the time range of 6 seconds to 20 seconds, the operator operates in the case where the hands are in a block and the left hand is at the top of the right hand, and Leap Motion sensor can't effectively detect the data of the left hand, while Kinect sensor can effectively detect the data of the right hand. As for the right hand, Leap Motion sensor can effectively detect the data of the right hand, but the Kinect sensor can't do that. For intuitive observation, Fig. 6 uses a zero value to show the data that can't be effectively detected. It can be seen that the hybrid sensor platform based on Leap Motion and Kinect can effectively track the data in real time when the hands are blocked.



(a) Left hand trajectory of each axis

(b) Right hand trajectory of each axis

Fig. 6. the track data of both hands identified each sensor.

The initial separation distance of the end of the double robot is 1.76 m, and after the movement of the dual robot guided by the operator's reasonable hand gestures, the two robots are constantly adjusting and approaching, and eventually screwing the bolt onto the nut, the final effect is shown in Fig. 7.

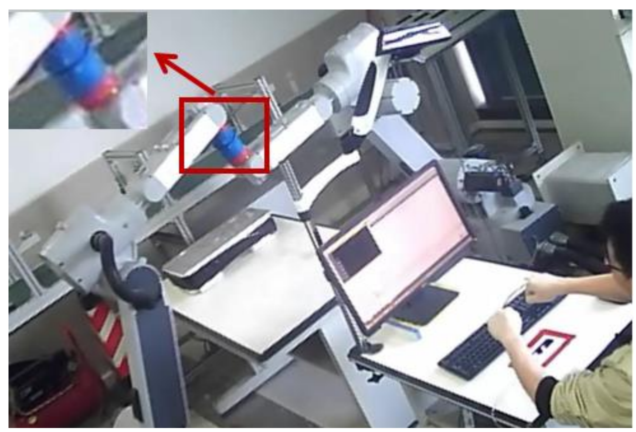


Fig. 7. The dual-robot screwing effect.

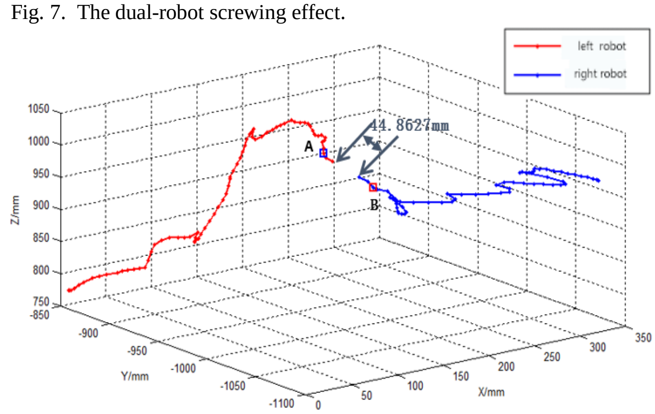


Fig. 8. Motion trajectories at the end of the dual-robot.

Fig. 8 shows the changes of the track of dual robots' end position, in which AB represents the position of bolts nuts which is coming into contact with the double end of the robot, at this time the distance between AB is 66.1607mm. Theoretically when bolts and nuts just approaching, the distance between the two robot end positions is the length of the bolt's length plus nut thread length, which is 65mm. Finally, the distance between the two robots is 44.8627mm, which means that the bolt nut is screwed in at a distance of 21.298mm (more than 20mm), so the screwing experiment can be considered successful.

In addition to the effectiveness of human-manipulator interaction, the performance analysis of human-manipulator interaction system based on hands gesture control includes two aspects: the stability of robot movement and the high efficiency of human-manipulator interaction.

For the stability of robot, Fig. 9 shows two robots of 6 joint angles rotational speed state during the progress of interaction. It can be seen from the results that the speed of movement of the two robots is basically the same, and the maximum velocity of joint 1, 2, 3, 4, 5 and 6 are 1.075 degrees / second, 1.467 degrees / second, 3.739 degrees / second, 0.225 degrees / second, 1.1746 degrees / second and 0.756 degrees / second, respectively. Each joint angle has different level of influence on the position and posture of the robot, so there are differences between their velocities. Though the movement of joint 3 is the fastest, its movement is relatively stable for the actual experimental environment.

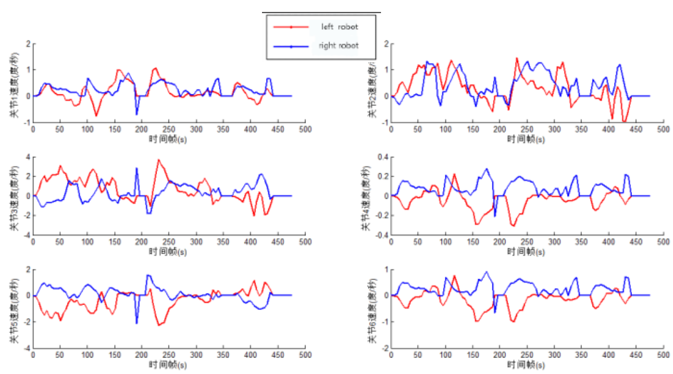


Fig. 9. Velocity of 6 joints angles. The x-axis represents the time frame (second), the y-axis represents the velocity of each joint (angle/second)

This paper proposes that, based on the Leap Motion Hands and Kinect mixed sensor platform, it can have a gesture interaction, which is able to recognize the existence of obscured hands gesture in a stable and long time, and reduce the reset times during the interaction gesture, thus improving the efficiency of this hands gesture interaction. For that, this paper compares hybrid sensor platform based on a Leap Motion and Kinect with a single Leap Motion sensor platform for hands-on gesture interaction[11].

In a single Leap Motion sensor platform, a Leap Motion was placed in front of the robots, and the operator stood in front of the Leap Motion. The disadvantage of this method is that the gesture must be reset to achieve further interaction when the hand gestures are occluded. The experiment was repeated three times to tighten the screw.

TABLE III is the statistics of each experiment's time consuming. It can be seen that the average time of proposed hybrid sensor platform for the screwing experiment is 48.5s, which is better than the average time of a single sensor Leap (60.4s). With improved operator proficiency based on hands gesture control double robot collaboration, tasks are less time-consuming and more interactive.

TABLE IV
A COMPARISON OF THE TIME CONSUMING OF EACH SCREWING EXPERIMENT

Experimental platform	Time consuming			
	Experiment 1	Experiment 2	Experiment 3	average
Mixed sensor	51.2s	47.7s	46.5s	48.5s
Single sensor [11]	62.3s	59.8s	59.1s	60.4s

In summary, in this paper, the double robot HMI system based on double hands gesture control shows the double robot has good movement stability and high efficiency of interaction in the process to complete the task of carrying out cooperation. In addition, the operator doesn't need to deliberately reset the gesture during the interaction, making the interaction process more natural and convenient.

VI. CONCLUSION

In this paper, we proposed a human-manipulator interaction technology of dual-robot based on hands gesture control for the lack of research in the related field, provided a new method for the research in multi-robot cooperation and man-machine collaboration. In this paper, the way of interaction is based on the Leap Motion and Kinect mixed sensor platform, so as to monitor hands gesture stably for a long time, and effectively solved the problem of the human-manipulator interaction stability with the existence of obscured hands gesture, and reduced the reset times during the interaction gesture, thus improving the efficiency of this hands gesture interaction.

The dual-robot human-manipulator interaction system that based on hands gesture control including two parts, which are the hybrid sensor gesture acquisition platform and the data processing and control decision module. This paper introduced the key technologies used in the two parts in detail, and got the staged achievements. It can be seen that the average time of proposed hybrid sensor platform for the screwing experiment is better than that of a single sensor Leap. With improved operator proficiency based on hands gesture control double Robot collaboration, tasks are less time-consuming and more interactive.

In this paper, the double robot HMI system based on double hands gesture control shows the double robot has good movement stability and high efficiency of interaction in the process to complete the task of carrying out cooperation. In addition, the operator doesn't need to deliberately reset the gesture during the interaction, making the interaction process more natural and convenient.

The dual-robot human-manipulator interaction system based on hands gesture control proposed in this paper giving the operator the interaction ability of two robots, implemented the function of controlling the movement of two robots with hands gesture. We hope this paper to have some value in the field of multi-robot cooperation and man-machine collaboration, and provide the reference value for the next generation of the new type human-manipulator interaction research.

REFERENCES

[1] Y. H. Zhu, "Intellisense Leading Development Hotspot of Future Robot," Journal on Science & Technology Review, vol. 33, no. 23, 2015.

[2] T. M. Wang, D. S. Chen, Y. Tao, *et al*, "Intelligent Machines That Will Change the World," Journal on Thinking on the Development of Intelligent Robots, vol. 21, pp. 16-22, 2015.

[3] S Waldherr, R Romero, S Thrun, "A Gesture Based Interface for Human-Robot Interaction," Journal on Autonomous Robots, vol. 9, no. 2, pp. 151-174, 2000.

[4] R Stiefelhagen, C Fugen, R Gieselmann, *et al*, "Natural human-robot interaction using speech, head pose and gestures," IEEE/RSJ International Conference on Intelligent Robots & Systems, vol. 3, no. 1, pp. 2422-2477, 2004.

[5] C Escolano, J Antelis, J Minguez, "Human brain-teleoperated robot between remote places," IEEE International Conference on Robotics & Automation, 2009, pp. 4430-4437.

[6] TI Cerlinca, SG Pentiuc, "Hand posture recognition for human-robot interaction," Journal on International Journal for Light and Electron Optics, vol. 125, no. 9, pp. 2148-2153, 2007.

[7] Y. M. Chen, Y. H. Zhang, "Robot Interaction Technology Research Based on Hand Gesture Recognition," Journal on Robot, vol. 31, no. 4, pp. 351-356, 2009.

[8] Q. C. Pang, "3D Gesture Data Processing Technology for Human-Manipulator Interaction Research," Doctoral dissertation, South China University of Technology, 2015.

[9] G. L. Du, P Zhang, *et al*, "Robot teleoperation using a vision-based manipulation method," Journal on International Conference on Audio Language & Image Processing, 2010, pp. 945-949.

[10] F Kobayashi, K Okamoto, "Robot-human handover based on position and posture of human hand," Journal on Joint International Conference on Soft Computing & Intelligent Systems, 2014, pp. 918-921.

[11] G. Du, P Zhang, "A Markerless Human-Robot Interface Using Particle Filter and Kalman Filter for Dual Robots," IEEE Transactions on Industrial Electronics, vol. 62, no. 4, pp. 2257-2264, 2015.

[12] A Utsumi, J Ohya, "Multiple-hand-gesture tracking using multiple cameras," IEEE Computer Society Conference on Computer Vision & Pattern Recognition, vol. 1, pp. 1473-1473,1999.

[13] B. M. Chen, "Gesture Tracking Research and Implementation of Gesture Recognition Application Platform," Doctoral dissertation, Sun Yat-sen University, 2011.

[14] RB Mapari, G Kharat, "Real time human pose recognition using leap motion sensor," IEEE International Conference on Research in Computational Intelligence & Communication Networks, 2015, pp. 323-328.

[15] Y. Chen, Y. Z. Shen, *et al*, "A Simplified Model of Three Dimensional Datum Transformation Adapted to Big Rotation Angle," Journal on Geomatics and Information Science of Wuhan University, vol. 29, no. 12, pp.1101-1105, 2004.

[16] Z. P. Lv, J. C. Wu, "Comparison of Three-dimensional Coordinate Transformation Algorithm," Journal on Journal of Geodesy and Geodynamics, vol. 36, no. 1, pp. 35-39, 2016.

[17] S. Gao, Q. Pan, *et al*, "Multi-sensor Adaptive Filter Data Fusion Algorithm," Journal on Journal of Electronics & Information Technology, vol. 30, no. 8, pp. 1901-1904, 2008.

[18] Y. L. Zhai, Y. S. Dai, "Study of Adaptive Weighted Fusion Estimated Algorithm of Multisensor Data," Journal on Acta Metrologica Sinica, vol. 19, no. 1, pp. 69-75, 1998.

[19] Kanda T, Miyashita T, Osada T, *et al*, "Analysis of Humanoid Appearances in Human-Robot Interaction," IEEE Transactions on Robotics, vol. 24, no. 3, pp. 725-735, 2008.

[20] J Sherrah, S Gong, "Resolving Visual Uncertainty and Occlusion through Probabilistic Reasoning," Journal on British Machine Vision Conference, 2000, pp. 252-261.

[21] S. Chen, H. Ma, *et al*, "Hand Gesture Based Robot Control System Using Leap Motion," Journal on International Conference on Intelligent Robotics and Applications, 2015, pp. 581-591.

[22] G. Wang, L. Qian, et al, "Continuous Tool Wear Prediction Based on Gaussian Mixture Regression Model," Journal on The International Journal of Advanced Manufacturing Technology, vol. 66, no. 9, pp. 1921-1929, 2013.

[23] LM Munoz, A Casals, "Improving the Human-Robot Interface Through Adaptive Multispace Transformation," IEEE Transactions on Robotics, vol. 25, no. 5, pp. 1208-1213, 2009.

[24] G. L. Du, P. Zhang, "Human-Manipulator Interface Using Hybrid Sensors with Kalman Filters and Adaptive Multi-Space Transformation," Journal on Measurement, vol. 55, no. 3, pp. 413-422, 2014.

[25] K Khosheiham, SO Elberink, "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications," Journal on Sensors, vol. 12, no. 2, pp. 1437-1454, 2012.

[26] Y. K. Liu, Y. M. Zhang, "Toward Welding Robot with Human Knowledge: A Remotely-Controlled Approach," IEEE Transactions on Automation Science and Engineering, vol. 12, no. 2, pp. 769-774, 2015.

[27] CS Fahn, KY Chu, "Hidden-Markov-Model-Based Hand Gesture Recognition Techniques Used for a Human-Robot Interaction System," Journal on Springer Berlin Heidelberg, vol. 6762, no. 1, pp. 248-258, 2011.

[28] F Weichert, D Bachmann, B Rudak, et al, "Analysis of thhe Accuracy and Robustness of the Leap Motion Controller," Journal on Sensor, vol. 13, no. 5, pp. 6380-6393, 2013.