Human–Manipulator Interface Based on Multisensory Process via Kalman Filters

Guanglong Du, Ping Zhang, and Di Li

***Abstract*—This paper presents a human–robot interface, which incorporates Kalman filters (KFs) and adaptive multispace transformation (AMT), to track movements of the human hand and control the robot manipulator. This system employs one inertial measurement unit and a 3-D camera (Kinect) to determine the orientation and translation of the human hand, and uses KFs to estimate these. Although KFs can estimate the translation, the translation error increases in a short period of time when the sensor fails to sense hand movement, including handshaking. Therefore, a method to correct the translation error is required. In this paper, the change rate of the human hand is used to determine the posture of the robot. An overdamping strategy is also employed to eliminate the effect of movement sensing failure. Given that a human operator has difficulty operating with high precision due to perceptive and motor limitations, an AMT method is proposed to assist the operator in improving the accuracy and reliability of determining the movement of the robot. The human–manipulator interface is then experimentally tested in a laboratory environment. The results indicate that the system based on the human–manipulator interface can successfully control the robot manipulator.**

***Index Terms*—Adaptive multispace transformation (AMT), human–robot interface, Kalman filter (KF), robot teleoperation.**

I. Introduction

HUMAN INTELLIGENCE is necessary in making a decision and in controlling a robot, particularly in unstructured dynamic environments. Robot teleoperation is essential in highly unstructured environments with unfamiliar objects. Some commonly used human–robot interfaces [1] include joysticks [2], dials [3], and robot–arm replicas [4]. However, using these mechanical contact devices for teleoperation tasks always requires unnatural hand–arm movements.

Another method to transfer complex movements to a remote robot is to track the hand–arm movement of the operator, which is used to complete the required task using inertial sensors, electromagnetic tracking sensors, gloves with angle sensors, and exoskeletal systems [5]. However, these contact devices may hinder the natural human limb movement. Nevertheless, this method is more natural than using mechanical contact devices.

Manuscript received February 6, 2013; revised July 26, 2013 and December 11, 2013; accepted December 18, 2013. Date of publication January 21, 2014; date of current version May 2, 2014. This work was supported by the Fundamental Research Funds for the Central Universities.

G. Du and D. Li are with the School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: medgl@scut.edu.cn; itdili@scut.edu.cn).

P. Zhang is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: pzhang@scut.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2014.2301728

Vision-based techniques are noncontact and seldom hinder the hand–arm movement. Such methods often use physical markers placed on the anatomical body part [6]. Numerous applications [7] rely on the marker-based method of tracking human movement. However, using this method is not always practical because body markers may hinder movement and some highly dexterous tasks and operators can be obstructed. Thus, a markerless approach is better for many applications.

Compared with the image-based tracking method, which uses markers, the markerless approach is less invasive and can eliminate problems of marker occlusion and identification [8]. Many existing markerless tracking techniques can capture images and then compute the movement later [9], [10]. Some techniques can only provide 2-D image information on human movement [11]. The method in [12] provided the position and the orientation of the human hand to control the manipulator. However, initialization requires the human operator to assume a simple posture, that is, a bare arm must be placed in front of a dark background and the hand should be higher than the shoulder. Hence, it is virtually impossible to obtain a precise result under a complex background. These tracking methods cannot be extended for accurate 3-D joint position data. In relation to this, identifying human body parts in different orientations has remained a challenge [13], [14].

The proposed system uses a 3-D camera (Kinect) to locate the hand of the human operator and an inertial measurement unit (IMU) to measure the orientation of the hand. This paper proposes the use of Kalman filters (KFs) to estimate the position and orientation of the hand. In addition, an overdamping strategy and the adaptive multispace transformation (AMT) method are used to eliminate the position noise and improve the accuracy of manipulation, respectively. Experimental results to validate the proposed methods are also presented.

The remainder of this paper is organized in sections. Section II presents the pose measurement system. Section III presents details about the human hand tracking system. Section IV describes the AMT method. The experiments and results are presented in Section V. Section VI provides the discussion, and Section VII presents the conclusions.

II. Orientation Measurement Using IMU

A. Factored Quaternion Algorithm

The factored quaternion algorithm (FQA) [15] is based on the gravity and magnetic field measurements of the Earth; it also estimates the orientation of a rigid body. This algorithm is only applied for the static or slow-moving rigid body. To apply FQA to situations with relatively large linear accelerations,

the KF fusion algorithm is used together with angular rate information to estimate the orientation of dynamic body (slow or fast moving) in the next section.

In the proposed application, a sensor module, which is a strapdown IMU, is attached to the human hand, whose orientation (roll, pitch, and yaw) is to be determined. The IMU sensor consists of one three-axis accelerometer, two two-axis gyroscopes, and one three-axis magnetometer.

Three frames are defined here, namely, body frame $x_by_bz_b$, sensor frame $x_sy_sz_s$, and Earth-fixed frame $x_ey_ez_e$. The sensor frame $x_sy_sz_s$ corresponds to the axes of three orthogonally mounted accelerometers/magnetometers. The body frame $x_by_bz_b$ is assumed to coincide with the sensor frame $x_sy_sz_s$. According to Euler's theorem [16] on finite rotations, the conversion from Euler angles to quaternion is given by

$$q = \begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} \cos(\phi/2)\cos(\theta/2)\cos(\psi/2)+\sin(\phi/2)\sin(\theta/2)\sin(\psi/2) \\ \sin(\phi/2)\cos(\theta/2)\cos(\psi/2)-\cos(\phi/2)\sin(\theta/2)\sin(\psi/2) \\ \cos(\phi/2)\sin(\theta/2)\cos(\psi/2)+\sin(\phi/2)\cos(\theta/2)\sin(\psi/2) \\ \cos(\phi/2)\cos(\theta/2)\sin(\psi/2)-\sin(\phi/2)\sin(\theta/2)\cos(\psi/2) \end{bmatrix} \quad (1)$$

where ϕ, θ, and ψ are the rotation angles of each axis with respect to frame $x_ey_ez_e$.

The four Euler parameters are constrained [17] as

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1 \quad (2)$$

where q_0 is the scalar part, and (q_1, q_2, q_3) is the vector part.

B. Orientation KF

KF is used to estimate the state x of IMU from a set of noisy and incomplete measurements, because both gyroscopes and magnetometer have white noise and random walk [18]. KF is a recursive stochastic technique that estimates the state at time k from the state at time $k-1$. The state transition equation at time k is

$$\begin{aligned} x_k &= A_k \cdot x_{k-1} + B \cdot u_{k-1} + w_{k-1} \\ z_k &= H \cdot x_k + v_k \end{aligned} \quad (3)$$

where x_k is the $n \times 1$ state vector at time k, A_k is a $n \times n$ state transition matrix, B is a $n \times p$ system input matrix, u_{k-1} is a $p \times 1$ vector with deterministic input at time $k-1$, w_{k-1} is a $n \times 1$ process noise vector at time $k-1$, z_k is a $m \times 1$ measurements vector at time k, H is a $m \times n$ observation matrix, and v_k is a $m \times 1$ measurement noise vector. The differential equation of the quaternion q with respect to time is given by

$$\begin{bmatrix} \partial q_0/\partial t \\ \partial q_1/\partial t \\ \partial q_2/\partial t \\ \partial q_3/\partial t \end{bmatrix} = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ v_x/2 \\ v_y/2 \\ v_z/2 \end{bmatrix} \quad (4)$$

where v_x, v_y, v_z are the angular velocity components of IMU in the x_s, y_s and z_s axes, respectively. Given that the state x_k includes the quaternion states and the angular velocities

$$x_k = [\, q_{0,k} \quad q_{1,k} \quad q_{2,k} \quad q_{3,k} \quad v_{x,k} \quad v_{y,k} \quad v_{z,k} \,] \quad (5)$$

where $q_{0,k}$, $q_{1,k}$, $q_{2,k}$, and $q_{3,k}$ are the quaternion states and $v_{x,k}$, $v_{y,k}$, and $v_{z,k}$ are the angular velocities at time k. From (4), the state transition matrix [19] is given as

$$A_k = \begin{bmatrix} 1 & 0 & 0 & 0 & -q_{1,k}\cdot\Delta t/2 & -q_{2,k}\cdot\Delta t/2 & -q_{3,k}\cdot\Delta t/2 \\ 0 & 1 & 0 & 0 & q_{0,k}\cdot\Delta t/2 & q_{3,k}\cdot\Delta t/2 & q_{2,k}\cdot\Delta t/2 \\ 0 & 0 & 1 & 0 & q_{3,k}\cdot\Delta t/2 & q_{0,k}\cdot\Delta t/2 & -q_{1,k}\cdot\Delta t/2 \\ 0 & 0 & 0 & 1 & -q_{2,k}\cdot\Delta t/2 & q_{1,k}\cdot\Delta t/2 & q_{0,k}\cdot\Delta t/2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6)$$

Here, Δt is the sampling time, and $B = 0^{n\times p}$ because there is no control inputs. Angular velocity is used to estimate the quaternion states. Hence, the process noise vector is given as

$$w_k = [0 \quad 0 \quad 0 \quad 0 \quad w_x \quad w_y \quad w_z]^T \quad (7)$$

where w_x, w_y, and w_z are the process noises of the angular velocity. Calibrated gyroscopes are used to measure the angular velocities. Hence, the observation matrix H is given as

$$H = \begin{bmatrix} 0^{3\times 4} & I^{3\times 3} \end{bmatrix}. \quad (8)$$

In order to satisfy (2), the determined quaternion q_k at time k should be normalized as

$$\begin{aligned} q_k &= [\, q_{0,k}/M \quad q_{1,k}/M \quad q_{2,k}/M \quad q_{3,k}/M \,] \\ M &= \sqrt{q_{0,k}^2 + q_{1,k}^2 + q_{2,k}^2 + q_{3,k}^2}. \end{aligned} \quad (9)$$

III. Human Hand Tracking

The human movement tracking approach is driven by two key design goals: computational efficiency and robustness. A single input depth image is segmented into a dense probabilistic body part labeling, with the parts defined to be spatially localized near skeletal joints of interest [20]. Human hand tracking and positioning are carried out by an application program interface (API), which is embedded in the Kinect.

A. Position Model

As aforementioned, the skeleton joint points can be derived from Kinect. Fig. 1(a) shows the 15 skeleton joints in the 3-D space. The 15 skeleton joint points are numbered from top to bottom and left to right. The right arm comprises the parts from the right shoulder to the right wrist. Fig. 1(b) shows that the right arm is $P_3P_4P_5$.

The coordinates of the 15 joint points are referred to as Kinect coordination. The position of the right end-effector can be achieved by the position of P_5.

B. Position KF

Fig. 1(b) shows the coordinate systems. The local frame is attached to a stationary building. Given that the Kinect is placed on a static table, the Kinect frame is relatively static with respect to the local frame. The operator holds the IMU in the manipulation, such that the IMU frame overlaps the hand frame.

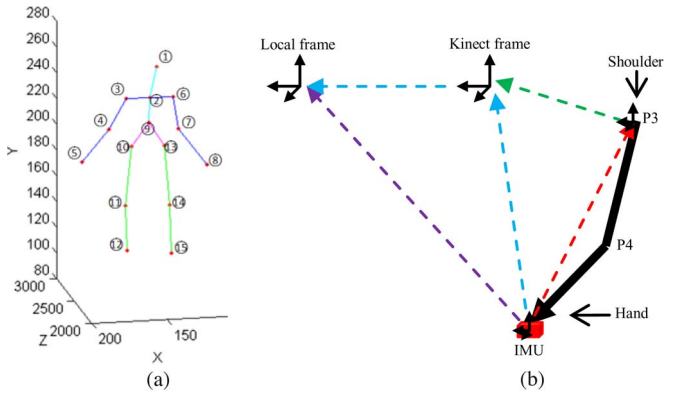

Fig. 1. Skeleton joint points in the 3-D space and the coordinates of the right hand. (a) 1: head; 2: shoulder center; 3: right shoulder; 4: right elbow; 5: right wrist; 6: left shoulder; 7: left elbow; 8: left wrist; 9: hip center; 10: right hip; 11: right knee; 12: right foot; 13: left hip; 14: left knee; 15: left foot. (b) Frames of human hand.

Let $P(p_x, p_y, p_z)$ be the coordinate of P_5 in the local frame. In the position estimating process, a KF is used to estimate the state P of the position from a set of noise measurement. In this estimation, six measurements are available. These include three acceleration components in the IMU frame, which are from the IMU, and three position components in the local frame. Let the direction of cosine matrix M_{H2S} from the IMU frame (or hand frame) to the local frame be

$$M_{H2L} = \begin{bmatrix} m_{X_x} & m_{Y_x} & m_{Z_x} \\ m_{X_y} & m_{Y_y} & m_{Z_y} \\ m_{X_z} & m_{Y_z} & m_{Z_z} \end{bmatrix}. \tag{10}$$

The acceleration of the hand in the local frame can be expressed as [19]

$$\begin{aligned} \dot{V_x} &= m_{X_x} \cdot A_x + m_{Y_x} \cdot A_y + m_{Z_x} \cdot A_z \\ \dot{V_y} &= m_{X_y} \cdot A_x + m_{Y_y} \cdot A_y + m_{Z_y} \cdot A_z \\ \dot{V_z} &= m_{X_z} \cdot A_x + m_{Y_z} \cdot A_y + m_{Z_z} \cdot A_z - |g_l| \end{aligned} \tag{11}$$

where $|g_l|$ is the magnitude of the local gravity vector, and (A_x, A_y, A_z) are the acceleration measurement components in each axis in the hand frame. The velocity components (V_x, V_y, V_z) in each axis in the local frame can be defined as

$$V_x = \dot{p_x} \quad V_y = \dot{p_y} \quad V_z = \dot{p_z}. \tag{12}$$

From (11) and (12), the state x'_k of the position KF is expressed as

$$x'_k = [p_{x,k}, V_{x,k}, A_{x,k}, p_{y,k}, V_{y,k}, A_{y,k}, p_{z,k}, V_{z,k}, A_{z,k}]. \tag{13}$$

Subscript k represents the state at time k. Based on (11) and (12), the state transition matrix A'_k is defined as

$$A'_k = \begin{bmatrix} 1 & t & m_{X_x} \cdot t^2/2 & 0 & 0 & m_{Y_x} \cdot t^2/2 & 0 & 0 & m_{Z_x} \cdot t^2/2 \\ 0 & 1 & m_{X_x} \cdot t & 0 & 0 & m_{Y_x} \cdot t & 0 & 0 & m_{Z_x} \cdot t \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_{X_y} \cdot t^2/2 & 1 & t & m_{Y_y} \cdot t^2/2 & 0 & 0 & m_{Z_y} \cdot t^2/2 \\ 0 & 0 & m_{X_y} \cdot t & 0 & 1 & m_{Y_y} \cdot t & 0 & 0 & m_{Z_y} \cdot t \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & m_{X_z} \cdot t^2/2 & 0 & t & m_{Y_z} \cdot t^2/2 & 1 & t & m_{Z_z} \cdot t^2/2 \\ 0 & 0 & m_{X_z} \cdot t & 0 & 0 & m_{Y_z} \cdot t & 0 & 1 & m_{Z_z} \cdot t \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{14}$$

Given that the acceleration measurements are affected by the gravitational force and the Z-axis is parallel to the gravity vector, the system input matrix is expressed as

$$B' \cdot u'_{k-1} = \left[0, 0, 0, 0, 0, 0, -|g_l| \cdot t^2/2, -|g_l| \cdot t, 0\right]^T. \tag{15}$$

Accelerations are used to estimate the position states such that the process noise vector becomes

$$w'_k = \left[0, 0, w'_x, 0, 0, w'_y, 0, 0, w_z\right]^T \tag{16}$$

where (w'_x, w'_y, w_z) is the process noise of the hand acceleration.

The Kinect sensor and the IMU are calibrated and initialized; hence, the observation matrix H' for the position estimation is expressed as

$$H' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{17}$$

Therefore, the nine states are observable.

The determined position $P_k(p_{x,k}, p_{y,k}, p_{z,k})$ at time k is the optimal value of the position of the human hand. However, controlling robots in this manner can lead to another problem: the position of the hand might be affected by the moving or trembling human body. To solve this problem, the shoulder frame (X_S, Y_S, Z_S) to be used as the reference frame is obtained with X_S collinear with $\overrightarrow{P_2P_3}$, $P_2P_3P_4$ coplanar with X_SY_S. The position of the human hand is relative to the shoulder instead of the origin of the local frame [see Fig. 1(b)]. Thus, the position of P_5 with respect to the shoulder is

$$P_M(p_x, p_y, p_z) = P_5 - P_3. \tag{18}$$

C. Overdamping Method

Using the API can track the human body and extract 15 skeleton joints, but the skeleton joint position data are unstable. The velocity of the human hand is transmitted to drive the virtual robot EE through the overdamping method (see Fig. 2). In this step, the unstable velocity data (outliers) yielded by the tracking algorithm are filtered. In the second step, the velocity of the virtual robot EE is transmitted to drive the remote robot by the AMT algorithm. In the second step, the influence of the uncorrected movement yielded by human is reduced.

Here, it is assumed that the virtual robot EE is a virtual point. A flexible force $\boldsymbol{F}$ between the human hand P and the virtual point E exists in every sampling period. Assuming that the virtual robot EE completes the simple harmonic motion and the quality of the virtual point is m, the position of the human hand does not directly act on the virtual point E. The movement is achieved by the following steps: P is the driving source, and E is the object to be driven, then, $\boldsymbol{F}$ acts upon E. Finally, $\boldsymbol{F}$ satisfies the relations given by

$$F = \begin{cases} k \cdot D, & D < \tau \\ 0, & D \geq \tau \end{cases} \tag{19}$$

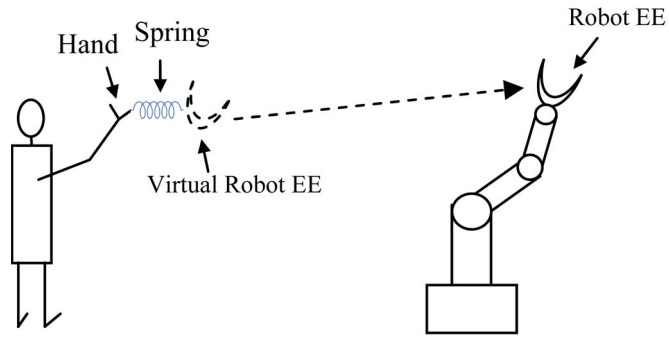

Fig. 2. Model of the overdamping method.

where k represents the Hooke parameters, τ represents the threshold of elastic limit of the spring, and D is the distance between P and E. When D is bigger than τ, E cannot move up through the force F since the spring breaks at the limit. As a result, an error occurs during the extraction of the joint points, that is, the front and next positions of joint points would suddenly change, leading to the sudden change for D. When the distance between P and E is bigger than τ, then $\boldsymbol{F}$ is 0. This condition implies that E is not pulled by $\boldsymbol{F}$, which eliminates the effect of sudden change caused by outlier data of joint points. When E is very close to P or when D is bigger than τ, $\boldsymbol{F}$ is 0. Due to a certain velocity of E and if no force acts on the E, E continues moving, and the system becomes uncontrollable. To solve this problem, a virtual resistance μ is introduced when E is moving, which is expressed as

$$\mu = \begin{cases} \gamma \dot{D}, & D < \tau \\ +\infty, & D \geq \tau \end{cases} \tag{20}$$

where γ is the damping coefficient.

Assuming that the sampling time of this algorithm is Δt, when $D > \tau$, which is considered due to the outlier data of extracting joint points, the traction force acted on E is 0. E ought to stop immediately because the resistance is infinite. When $D \leq \tau$, the system is in criticality damping vibration.

When hands are moving slowly or remain still, a dithering phenomenon exists when the extracted joint points randomly tremble around real joint points to some extent.

The overdamping model is introduced to solve this problem. When the human hand P vibrates in a small range around E, the gravitational force between P and E is smaller than the resistance acting on E. The force cannot move E to move.

Let $\omega = (k/m)$; $\beta = (\gamma/2m)$, when $\omega^2 < \beta^2$, the algorithm model is overdamping, and the velocity of the virtual point is

$$\dot{P_M} = \frac{dD}{dt}. \tag{21}$$

IV. AMT

Humans have inherent perceptive limitations (e.g., perception of distance) and motor limitations (e.g., physiological tremor), which prevent them from doing certain tasks precisely and smoothly [21].

To improve the visual and motor performances of teleoperation interfaces, the modified version of AMT is applied. In the proposed method, an interface that introduces two scaling processes is used to link the human operator working space (WS) to the robot WS. The first change scales the movement produced by the human operator. Another change of scale is applied between the virtual unit vector K of the central axis of the robot EE and the robot movements. Such changes in scale modify the robot speed, thus improving performance.

The scaling vector S is used to coordinate the actions of the human hand in master space (MS) and the movement of the virtual point K in the visual space (VS). Another scaling variable u is used to relate the VS space to the robot WS. S and u are functions of the distance r between robot EE and the target. When $S < 1$, the movement of K decelerates, whereas $S > 1$ accelerates the movement of K.

As the movement of the virtual position in VS is affected by vector S, let $S = [S_{\text{pos}}, S_{\text{ori}}]$, where S_{pos} is the scaling vector of the position, and S_{ori} is the scaling variable of the orientation. If the Euler angular velocity in MS is $\dot{O}_M$ and $\dot{O}_V$ is the angular velocity in VS, then the following equation is obtained:

$$\dot{O_V} = S_{\text{ori}} \cdot \dot{O_M}. \tag{22}$$

Assume that $\dot{P}_M$ is the velocity vector of hand movement in MS and $\dot{P}_V$ is the speed of the vector K in VS, then the mapping vector of $\dot{P}_M$ in the VS is $\dot{P}'_M$ [see Fig. 3(a)]. If $S_{\text{pos}} = [s_k s_{k_\perp}]$, $K_\perp$ is vector perpendicular to K, and $K_\perp$, K, and $\dot{P}'_M$ are coplanar, then the speed components $(\dot{p}_k, \dot{p}_{k_\perp})$ of $\dot{p}_V$ are defined in (23), shown at the bottom of the page. When $s_k = s_{k_\perp}$, the speed $\dot{p}_V$ results in $\dot{p}_v = s_k \dot{p}'_M$. Meanwhile, $s_k < s_{k_\perp}$ means that it requires greater precision in the direction of the central axis of EE. The direction, which is perpendicular to the central axis, requires greater precision with $s_k > s_{k_\perp}$. In this paper, $s_k > s_{k_\perp}$ is used such that the operator can easily center the axis of the hole and quickly insert the peg into the hole.

The arrangements of WS and VS are shown in Fig. 3(b). The speed $\dot{p}_W$ is expressed as

$$\dot{P_W} = \left(\dot{P_V} + \dot{p} - (P_W - P_C) \frac{du}{dr} \dot{r} \right) \frac{1}{u} \tag{24}$$

where P_C is the position of the zoom center in VS, P_W is the position of the EE in WS, and $\dot{p}$ is the panning speed.

$$\dot{p_k} = \begin{cases} s_k \cdot \left(\left| \dot{P'_M} \right| \cdot \cos\theta \right) \cdot K = s_k \cdot K \cdot \dot{P'_M} \cdot K, & S_k \leq \delta_k \\ 0, & S_k > \delta_k \end{cases}$$

$$\dot{p_{k_\perp}} = \begin{cases} s_{k_\perp} \cdot \left(\left| \dot{P'_M} \right| \cdot \sin\theta \right) \cdot K_\perp = s_{k_\perp} \cdot K_\perp \cdot \dot{P'_M} \cdot K_\perp, & S_{k_\perp} \leq \delta_{k_\perp} \\ 0 & S_{k_\perp} > \delta_{k_\perp} \end{cases} \tag{23}$$

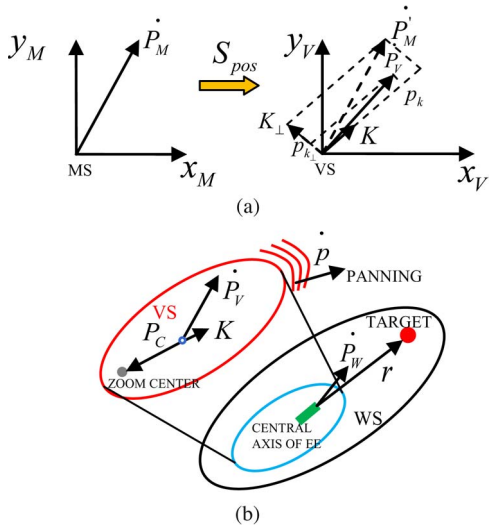

Fig. 3. Arrangement between VS and WS with the velocity vectors.

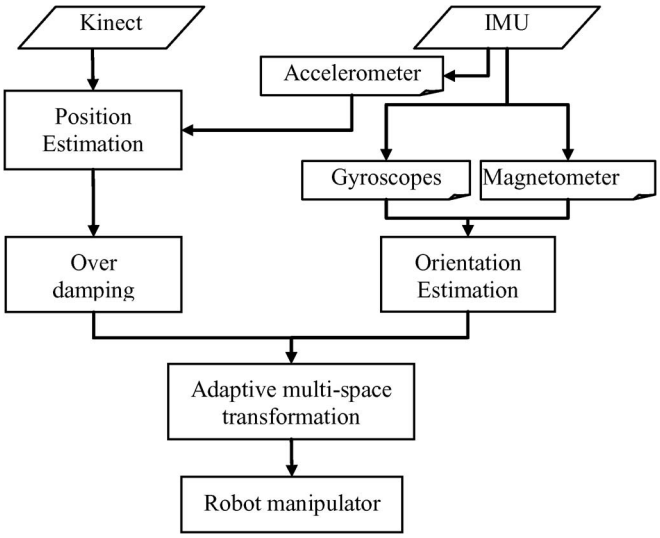

Fig. 4. Human–robot interface system process.

Fig. 4 shows the structure and data stream of our system.

V. Experiments

A. Environment of Experiment

To verify the proposed method, two experiments were designed. In Experiment 1, a series of five tests were carried out to copy the human hand movement in the human–robot–manipulator interfaces for robot teleoperation. The tests were designed to compare the proposed method without AMT and the method presented in [6] in terms of the accuracy of controlling the robot movements in real time to perform object manipulation tasks, as well as the ability of the manipulator to copy human hand–arm movement. Joint angles can be achieved by the solution of robot inverse kinematics. Given a pose of end-effector, the angle of each joint can be achieved by the robot inverse kinematics [23].

Our method without AMT is direct position and orientation control, which is the same as that shown in [6]. During each test, an operator moved the arm in the working volume to perform a pick-and-place object-manipulation task. The task was to pick up an object and to place the object on a target, with the object and target edges aligned [see Fig. 5(a)]. To pick and place the object, the controlling robot must close and open its claws. Here, the movement of the left arm was used to give the commands. When the left wrist was higher than the left elbow, the command corresponded to closing the claws. When the left wrist was higher than the left shoulder, the command corresponded to opening the claws. The object was a rigid plastic block, 180 mm high, 90 mm wide, and 50 mm thick. The target was a 180 mm $\times$ 90 mm square paper. The 2-D error $E_{\mathrm{pos},2D}$ in translation was the distance between the center of the target and that of the object. The 2-D error $E_{\mathrm{ori},2D}$ in rotation was the angle θ between the edge of the target and that of the object [see Fig. 6(c)].

Fig. 5. Environments of the experiment. (a) Environment of Experiment 1. The robot picks up the object and places it on the target paper. (b) Environment of Experiment 2. The robot inserts the peg into a hole.

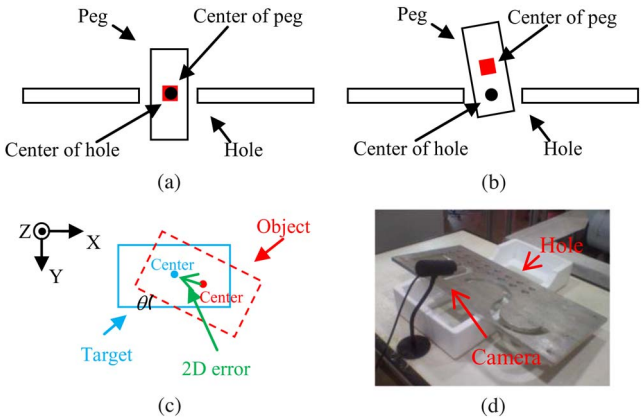

Fig. 6. Definition of errors and experiment platform of Experiment 2. (a) Screw inserted in the hole. (b) Definition of 3-D error. (c) Definition of 2-D error. (d) Experiment platform of Experiment 2.

In Experiment 2, a series of five tests was carried out to evaluate whether the accuracy of the proposed approach was sufficient to enable a human operator to control the movements of the robot in real time to perform peg-into-hole manipulation tasks [see Fig. 5(b)]. Given that the maximum crack between the peg and the hole was 3 mm and the errors of the method shown in [6] were more than 3 mm, only our method with AMT was carried out in Experiment 2. In this experiment, the steel plate had 16 holes [see Fig. 6(d)]. The peg is a cylinder with a radius of 7.5 mm and a length of 50 mm. The radius of the hole was 9 mm, and the size of the steel plate was 300 mm $\times$ 500 mm. The inclinations of the steel plate in X, Y, and Z were

TABLE I
NOMINAL LINK PARAMETERS IN THE DH MODEL FOR THE GOOGOL GRB3016 ROBOT

DH / Joint	a (mm)	α (radian)	d (mm)	θ (radian)
1	150	$-\pi/2$	250	0
2	570	$-\pi$	0	$-\pi/2$
3	150	$\pi/2$	0	0
4	0	$-\pi/2$	650	0
5	0	$-\pi/2$	0	$-\pi/2$
6	0	0	-200	0

8°, 0°, and 0°, respectively. A camera with 1280 (H) × 960 (V) picture elements and 22 frame/s frequency was mounted around the robot EE to capture the images of the robot and the target. In this experiment, the peg was fixed in the robot EE, and the steel plate was placed in the special position. The feedback display consisted of a 23-in 1600 × 1200 pixel resolution monitor, which was placed on the ground. The display window was configured in such a way that the total WS 400 mm × 300 mm was visualized on a unitary scale, implying a spatial resolution of 4 pixels/mm. To assist the operator in locating the object, a virtual system was built, in which a virtual camera was fixed under the peg. The task was to control the movement of the robot EE to insert the peg into a hole of the steel plate. In this experiment, $S_{\text{ori}} = r/M_r$ and $S_{\text{pos}} = [\sqrt{r}/M_r r/M_r]$ were used, where r is the distance between the peg and the hole, and M_r is the distance from the initial position of the peg to the hole. Hence, the robot EE rapidly moved in the direction of the axis of the peg and slowly moved in the direction perpendicular to the axis of the peg. The velocity of the robot EE was reduced along with r.

A GOOGOL GRB3016 robot model was used in the current experiments. Table I lists the nominal link parameters of the robot.

To evaluate the manipulation accuracy, 3-D position errors between the peg and the hole were proposed. Another calibrated camera with 1280 (H) × 960 (V) picture elements was used to measure the 3-D position errors [see Fig. 6(d)]. In the initiation of the system, the camera measured the coordinates of the center of the hole with respect to the camera coordinate system. After the peg was inserted into a hole, the camera then measured the depth and the direction of insertion by detecting the edge of the peg. Then, the system calculated the center of the peg with respect to the camera coordinate system. If (x_o, y_o, z_o) and (x_t, y_t, z_t) are the centers of the peg and the hole, respectively [see Fig. 6(a) and (b)], then the 3-D position error E can be written as

$$E_{3D} = \sqrt{(x_o - x_t)^2 + (y_o - y_t)^2 + (z_o - z_t)^2}. \quad (25)$$

B. Results of Experiments

In Experiment 1, the proposed method without AMT and that presented in [6] are compared in terms of accuracy and operating time. Table II shows the errors for the five tests with the two methods. In our method without AMT, the 2-D errors for the five tests ranged from −2.71 to 5.53 mm in X translation, −4.66 to 3.31 mm in Y translation and 0.8° to 1.4° in Z rotation, with the mean errors (MEs) of 2.93 mm, 2.93 mm, and 1.08°, respectively. In the method presented in [6], the 2-D errors for the five tests ranged from −4.52 to 4.96 mm in X translation, −5.02 to 5.35 mm in Y translation, and 1.9° to 4.2° in Z rotation, with the MEs of 4.12 mm, 4.25 mm, and 2.86°, respectively. Compared with the method presented in [6], the total translation errors of the proposed method without AMT decreased to 1.69 mm. The Z rotation errors decreased to 1.78°. The translation errors of the proposed method are slightly smaller than that of the method in [6], but the rotation errors of our method are less than half of that of the method in

TABLE II
ERRORS FOR ALL FIVE TESTS

Times	Our method without AMT				The method [6]			
	X mm	Y mm	Total mm	Z deg	X mm	Y mm	Total mm	Z deg
1	2.41	-4.66	5.24	0.9	4.96	-5.02	7.06	2.3
2	-2.71	3.31	-4.27	1.1	-4.52	-3.87	-5.95	3.5
3	2.15	-1.74	2.77	0.8	-3.98	3.29	-5.16	2.4
4	-1.88	-2.13	-2.84	1.2	-3.97	5.35	6.66	1.9
5	5.53	2.82	-6.21	1.4	3.20	-3.78	-4.95	4.2
MEs	2.93	2.93	4.26	1.08	4.12	4.25	5.95	2.86

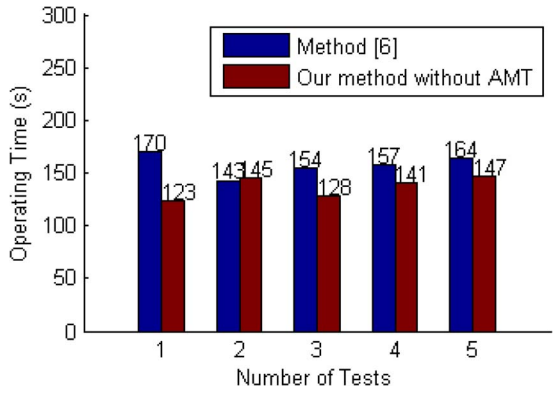

Fig. 7. Comparison of operation time.

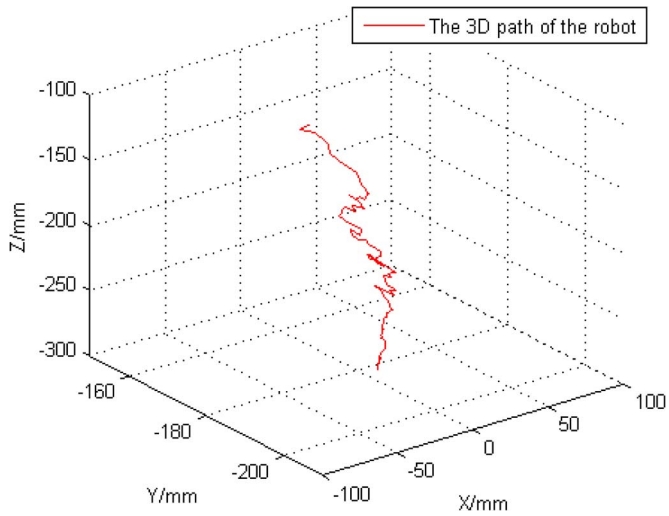

Fig. 8. Three-dimensional path of the robot.

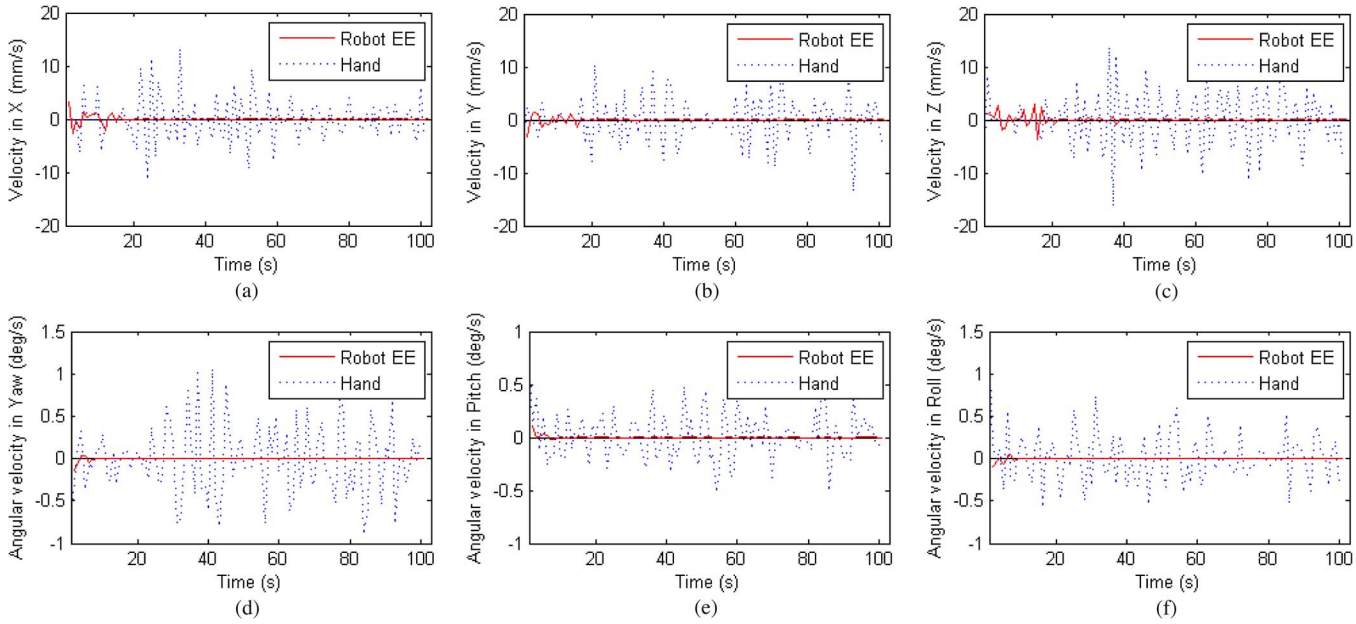

Fig. 9. Peg tracking results. (a) Velocity in X of robot EE and human hand. (b) Velocity in Y of robot EE and human hand. (c) Velocity in Z of robot EE and human hand. (d) Angular velocity in yaw of robot EE and human hand. (e) Angular velocity in pitch of robot EE and human hand. (f) Angular velocity in roll of robot EE and human hand.

[6]. In our method, high-precision equipment (i.e., IMU) was used to measure the orientation of the human hand. Thus, the rotation errors are smaller in the proposed method. However, due to perceptive and motor limitations, the human operator has difficulty carrying out the high-precision operation. In Experiment 2, the AMT was used to assist the operator to perform a high-precision manipulation (peg-into-hole).

Given that the method presented in [6] needs to execute big movements to control the robot manipulator, its operating time is longer than that of the current method. The mean time of the method presented in [6] is 157.6 s, which is more than that of the proposed method (136.8 s), as shown in Fig. 7.

In Experiment 2, AMT was used to improve the accuracy of the manipulation. The peg started from the initial position and moved for 103 s before it was inserted into the hole. The test results of tracking the peg are shown in Fig. 9.

Fig 8 shows the 3-D path of the peg during the insertion. Fig. 9(a)–(c) shows the velocity (position) of the robot EE and the human hand in X, Y, and Z. Fig. 9(c)–(e) shows the angular velocity (orientation) of the robot EE and the human hand in the yaw, pitch, and roll. The 3-D error between the peg and the hole is 3.1 mm. From Fig. 9, it can be observed that the amplitude of the human hand remains unchanged, but the robot runs down steadily. In the experiment, the operator did not need to move cautiously because a big movement of the hand had little or even no effect on the robot. With the reduction in the distance between the robot EE and the target, the amplitude of the velocity (position and orientation) decreased.

Due to the AMT method, the peg can be inserted into the small hole through our human–robot interface. The operator repeated the corrective actions until the target is reached because the movement space is limited. In this experiment, the operator needed to make a fast reflex to withdraw the hand when his hand reached the border of the movement space. In the command transmission module, the withdrawing movement was recognized by the overdamping method and then discarded. The robot EE did not move at all; hence, the overdamping protected the robot.

VI. Discussion

This paper presents a robot control method based on hybrid sensors. The translation and the orientation of the robot manipulator are controlled by copying the movement of the human hand. Compared with electromagnetic contact devices, such as hand joystick and data gloves, the proposed method rarely hinders natural human limb movements and even enables the operator to focus on his own task instead of thinking of how to deconstruct the commands into some simple orders. Compared with the markerless noncontact method [6], [12], the proposed method has high accuracy and strong stability. This system can be used immediately without any initialization. This control system can be also used outdoors. The lighting effect can be ignored, and the system does not need accurate image processing to extract the 3-D coordinates, thus allowing the system to work in more severe environments (e.g., highlights and darkness) because this algorithm uses infrared distance measurement to obtain arm data. Moreover, the algorithm presented in [12] requires a bare hand to recognize the color of skin; otherwise, the algorithm cannot be used to extract the hand data. In comparison, the proposed algorithm does not require a bare hand, and the operator can wear gloves when using the system in the cold outdoor working environment. This feature therefore expands the application range of the system.

At present, the contact interfaces used in teleoperation for the multiple robot manipulators often require multiple operators [24]. In the current paper, the right hand is used to control a robot manipulator. The proposed method can be also expanded to control dual manipulators with a Kinect and two IMUs using both hands.

VII. Conclusion

This paper presents a human–robot interface system that utilizes a hybrid sensor configuration consisting of an IMU and a 3-D camera. KFs were developed to estimate the position and the orientation of the human hand more stably. To eliminate the effect of the occlusion and the failure of movement sensing, an overdamping method was proposed in this paper. Since humans have inherent perceptive and motor limitations, the high-precision tasks cannot be performed directly by a human. Therefore, an AMT system was employed o assist the operator in performing the high-precision tasks.

In the experiments, picking and placing as well as peg-into-hole tasks were carried out. The accuracy and the efficiency of the proposed method and that presented in [6] were also compared. The feasibility of manipulation for high-precision tasks was validated in Experiment 2. Using the proposed system, the experimental results showed that the manipulation can be correctly carried out, and the manipulation error was reduced.

References

[1] K. B. Cho and B. H. Lee, "Intelligent lead: A novel HRI sensor for guide robots," *Sensors*, vol. 12, no. 6, pp. 8301–8318, Jun. 2012.

[2] C. Mitsantisuk, S. Katsura, and K. Ohishi, "Force control of human–robot interaction using twin direct-drive motor system based on modal space design," *IEEE Trans. Ind. Electron.*, vol. 57, no. 4, pp. 1338–1392, Apr. 2010.

[3] S. Hirche and M. Buss, "Human-oriented control for haptic teleoperation," *Proc. IEEE*, vol. 100, no. 3, pp. 623–647, Mar. 2012.

[4] T. Ando, R. Tsukahara, and M. Seki, "A haptic interface 'Force Blinker 2' for navigation of the visually impaired," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4112–4119, Nov. 2012.

[5] K. Kiguchi, S. Kariya, and K. Watanabe, "An exoskeletal robot for human elbow motion support-sensor fusion, adaptation, and control," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 31, no. 3, pp. 353–361, Jun. 2001.

[6] J. Kofman, X. Wu, T. J. Luu, and S Verma, "Teleoperation of a robot manipulator using a vision-based human–robot interface," *IEEE Trans. Ind. Electron.*, vol. 52, no. 5, pp. 1206–1219, Oct. 2005.

[7] J. H. Park, Y. D. Shin, J. H. Bae, and M. H. Baeg, "Spatial uncertainty model for visual features using a Kinect sensor," *Sensors*, vol. 12, no. 7, pp. 8640–8662, Jun. 2012.

[8] S. Verma, "Vision-based markerless 3D human-arm tracking," M.S. thesis, Dept. Mech. Eng., Univ. Ottawa, Ottawa, ON, Canada, 2004, pp. 323–356.

[9] G. L. Du, P. Zhang, J. H. Mai, and Z. L. Li, "Markerless Kinect-based hand tracking for robot teleoperation," *Int. J. Adv. Robot. Syst.*, vol. 9, no. 36, pp. 1–12, 2012.

[10] K. C. C. Peng, W. Singhose, and D. H. Frakes, "Hand-motion crane control using radio-frequency real-time location systems," *IEEE/ASME Trans. Mechatronics*, vol. 17, no. 3, pp. 464–471, Jun. 2012.

[11] M. Khezri and M. Jahed, "A neuro-fuzzy inference system for sEMG-based identification of hand motion commands," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1952–1960, May 2011.

[12] J. Kofman, S. Verma, and X. H. Wu, "Robot–manipulator teleoperation by markerless vision-based hand–arm tracking," *Int. J. Optomechatron.*, vol. 1, no. 3, pp. 331–357, Jan. 2007.

[13] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 11, pp. 3592–3607, Nov. 2011.

[14] E. Ueda, Y. Matsumoto, M. Imai, and T. Ogasawara, "A hand-pose estimation for vision-based human interfaces," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 676–684, Aug. 2003.

[15] X. Yun, E. R. Bachmann, and R. B. McGhee, "A simplified quaternion-based algorithm for orientation estimation from Earth gravity and magnetic field measurements," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 3, pp. 638–650, Mar. 2008.

[16] J. Diebel, *Representing Attitude: Euler Angles, Unit Quaternions, and Rotation Vectors.* Stanford, CA, USA: Stanford Univ. Press, 2006, pp. 136–148.

[17] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics*, 3rd ed. Reading, MA, USA: Addison-Wesley, 2002, pp. 131–162.

[18] J. A. Corrales, F. A. Candelas, and F. Torres, "Hybrid tracking of human operators using IMU/UWB data fusion by a Kalman filter," in *Proc. 3rd ACM/IEEE Int. Conf. Human Robot Interaction*, Amsterdam, The Netherlands, 2008, pp. 193–200.

[19] S. H. P. Won, F. Golnaraghi, and W. W. Melek, "A fastening tool tracking system using an IMU and a position sensor with Kalman filters and a fuzzy expert System," *IEEE Trans. Ind. Electron.*, vol. 56, no. 5, pp. 1782–1792, May 2009.

[20] J. Shotton, T. Sharp, A. Kipman, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Commun. ACM*, vol. 56, no. 1, pp. 116–124, Jan. 2013.

[21] L. M. Munoz and A. Casals, "Improving the human–robot interface through adaptive multispace transformation," *IEEE Trans. Robot.*, vol. 25, no. 5, pp. 1208–1213, Oct. 2009.

[22] R. Paul, *Robot Manipulators: Mathematics, Programming, and Control*. Cambridge, MA, USA: MIT Press, 1982, pp. 50–55.

[23] G. Antonelli, S. Chiaverini, and G. Fusco, "A new on-line algorithm for inverse kinematics of robot manipulators ensuring path tracking capability under joint limits," *IEEE Trans. Robot. Autom.*, vol. 19, no. 1, pp. 162–167, Feb. 2003.

[24] R. Marin, P. J. Sanz, P. Nebot, and R. Wirz, "A multimodal interface to control a robot arm via the web: A case study on remote programming," *IEEE Trans. Ind. Electron.*, vol. 52, no. 6, pp. 1506–1521, Dec. 2005.

Guanglong Du received the M.S. and Ph.D. degrees in human-computer interaction from the South China University of Technology, China, in 2009 and 2013, respectively.

He is currently a Postdoctor with the School of Mechanical and Automotive Engineering, South China University of Technology. His research interests include machine vision, human–computer interaction, and virtual fixtures.

Ping Zhang received the B.S. degree in mechanical engineering, and the M.S. and Ph.D. degrees in robotics from Tianjin University, Tianjin, China, in 1985, 1988, and 1994, respectively.

He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include intelligent networked robotics, intelligent networked manufacturing, human–computer interaction, and real-time embedded systems.

Di Li received the B.Sc. and M.Sc. degrees from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree in control from the South China University of Technology, Guangzhou, China, in 1993.

She is currently a Professor with the School of Mechanical and Automotive Engineering, and the Head of the Institute of Optical, Mechanical, and Electronic Integration, South China University of Technology. Her research interests include control systems, embedded systems, computerized numerical control systems, and machine vision.