IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 23, 2019, accepted March 9, 2019, date of publication March 14, 2019, date of current version April 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2905051

Human Intention Understanding From Multiple Demonstrations and Behavior Generalization in Dynamic Movement Primitives Framework

BOYANG TI, YONGSHENG GAO, QIANG LI, AND JIE ZHAO
State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China
Corresponding author: Yongsheng Gao (gaoys@hit.edu.cn)

This work was supported in part by the National Key Research and Development Program under Grant 2017YFB1303002.

ABSTRACT Human's interference in the process of skill learning can improve the performance of the robot greatly. However, learning from demonstration to generate a new action with human behavioral characteristics in the varying situation is challenging. Generally, dynamic movement primitives (DMPs) method can generalize the trajectory imitating the demonstration, but cannot integrate the feature of multiple trajectories of different targets. In this paper, the proposed method contains two aspects of learning and generating. The statistical method Gaussian mixture model and Gaussian mixture regression (GMM-GMR) is used to extract the common characteristic and eliminate the uncertainty of the multiple demonstrations. To exert the ability of DMPs to generate a human-like motion to a new goal, and we model the shape parameter with locally weighted regression (LWR) method. To enhance the ability of DMPs in multiple trajectories learning, we propose the multivariate Gaussian process regression (MV-GPR) method to construct the regression model of shape parameters to reflect the human intentions, with respect to the target position. To verify the feasibility of the proposed method, we design a peg-in-hole experiment with proving generalization and obstacle avoidance performance. The results have shown that the strategy integrated the generalization of DMPs and feature regeneration ability of MV-GPR method, and the generated valid trajectory could achieve the peg-in-hole task with 6-DOF whole-arm avoidance.

INDEX TERMS Dynamic movement primitives, learning from demonstration, MV-GPR, whole-arm obstacle avoidance.

I. INTRODUCTION

In the past decades, the demand for intelligent algorithm to mimic human behavior has been growing drastically, such as in service, extreme and other unconstructed environment. Different from the traditional methods, intelligent algorithm with the machine learning method are trying to make the transition from blind self-learning to empirically learning and the behavior can be generalized in different situations. Based on the prior knowledge, the algorithm can simplify the construction of learning system and reduce the computational performance requirement for the robot.

Learning from Demonstration (LfD) [1] a kind of imitation learning method has been shown to an efficient strategy learning the control policy. Human have the ability to perform complex behavior facing to the changing environment. Through simple programming human can guide robot to learn and perform the complex task. The simplicity and usefulness of algorithm release workers from labour intensive programming work or make it easier to assist human with simple tasks in daily life. LfD can also be classified as a kind of supervised learning. In supervised learning, the agent is presented with labeled training data and learns the mapping between the world state and actions [2]. The challenge of generalizing learned skill is trying to formulate a model mapping between the situation and behavior.

Recently there have been extensive on the dynamical system combination of the movement primitives to learn the demonstration movement [3]–[9], moreover [10]–[12] introducing the sensors information such as force to modulate the learned movement. Reference [13] also extracted variable stiffness in real time to learn the stiffness feature of human not limited to the movement. To improve the ability of the generalization, multiple optimization has been applied to

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

the model construction, such as batch learning [3], [9] and incremental learning [14]–[17]. Another branch is focusing on the probabilistic mixture modeling framework embedded into the dynamical system. The task parameters were inserted into the construction of the model expressed in the coordinate system. This approach is not limited to a specific parameter estimation technique.

Reference to the obstacle avoidance problem, Sampling-based motion planning method has been widely used and developed rapidly including Probabilistic Roadmaps Method (PRM) [18], [19] and Rapidly-Exploring Random Tree (RRT) [20]. Due to the randomization feature, it cannot ensure the path optimality and the computational rapidity. To solve the disadvantage of these aspects, most attentions have tended to focus on narrowing the range of the sampling region such as updating GMM by Greedy expectation-maximization (EM) [21] adopting the principal component analysis (PCA) method to compress sampling data dimension [22] and the retraction policy [23]. However, less the guiding of human, the optimal trajectory is explored without any reference. The LfD method takes the advantage of the human experience information while obstacle is considered during the demonstration, obstacle's information [24] has been inserted into the model learning as the external parameter to modulate the human demonstration to reduce the complexity of robot self-exploring.

The main contributions in this paper are given as follows:

(1) The underlying common features of different demonstrations are integrated by GMM-GMR method, then take the synthetic trajectory into DMPs to model the movement primitives system.

(2) We propose the combination of MV-GPR multi-output regression method and DMPs to optimize the forcing term.

(3) Extract the demonstration shape features including the human intention, and we generalize a new hybrid characteristic trajectory to neighboring points, with the ability of whole-arm obstacle avoidance in unknown obstacle position context.

(4) Finally, we try to verify the proposed method using 6-DOF arm of UR10 with Kinect in the peg-in-hole task, and the task can be performed accurately.

The organization of this paper can be divided into five sections, after the introduction part, we outline the related work of the imitation and generalizing method in the changing context in Sec. II and present the mathematical basis of the main method in Sec. III. Sec. IV describes the experiment settings and the result obtains. Sec. V presents the conclusion of this paper.

II. RELATED WORKS

In general there are two types of approach of robot skill learning from human demonstration. Firstly just to imitate the demonstration and reproduce the same trajectory as human's. Secondly to train a policy generalizing learned skill in a new context from demonstration [7], [25] in addition, it also includes the obstacle avoidance or other affiliated motion [5], [6], [26].

There are two major methods to process the obtained data. Task-parameterized probability approach aims at increasing the generalization ability by exploiting the functional nature of the task parameters. TP-GMM method tries to construct multiple coordinate system corresponding to the objects existing in multi-process task with the orientation in the model [27]. In [28], the approach compacted the TP-GMM in the bimanual tasks, introducing the external task parameters to modulate the end-effecter poses. References [29] proposed the Semitied-GMM model to extract the latent space features, which enables the robot to autonomously deal with different situations in manipulating task. Another method introduced the EMG signal to represent the task environment indirectly and combine the movement encoded by DMPs to complete the complex task [30]. Dynamical movement primitives (DMPs) method utilizes the advantage of its generalization and imitation ability, to generate a trajectory to achieve tracking the goal with property of demonstration. In this paper we focus on describing the second method DMPs.

In recent years, based on the conventional DMPs, researchers have improved the performance in parameter learning by statistical learning method, or proposed new algorithm citing the idea of DMPs, such as MoMP, ProMP.

ProMP proposed in [31] uses the probabilistic framework to encode the demonstration. Probabilistic method could extract the variance of the human and represent the confidence of the operation process. ProMP has the better inference ability than DMPs, taking all the demonstration shapes into consideration to generalize to a new goal and can also blend or combine different distributions to generate a synthetic trajectory. But ProMP need amount of experimental data obtained from the same environment. MoMP [32] proposed a method to combine and regenerate the movement primitives corresponding to the new goal position and store other task information in the library. Furthermore, it also solved the problem of the terminal velocity limited at zero to complete the striking movement.

DMPs [37] is a kind of one-shot learning method, it can converge to the goal with the property of scaling on spatial and temporal. The original version of DMPs is able to mimic only one demonstration but lacks the property of integrate shape feature of demonstration trajectory of different goals. Most work on DMPs were focusing on the parameter optimization or Task-Parameter introduction. The approach in [5] utilized mixture of GMMs to learn the task-parameterized DMPs, Gaussian mixture model constructed the correlation among forcing term in DMPs, canonical phase and task parameters, another approach [26] compacted the above task-parameter method through integrating multiple function into one function and introducing the Gaussian kernel to expand multi-dimensional task parameter. To adapt to the variable height of obstacle, [38] proposed to reconstruct the forcing term with stylistic factor by SVD. The trajectory was predicted by the mapping from environment feature parameters to trajectory style factors. But the above work planned a path only on the plane, the end-effector's trajectory was 2D. Our method focus on 3D path planning without considering the task parameter.

Above researches on DMPs only obtained one demonstration to train the model, the single demonstration was insufficient to extract the information of the human invention. In [3], it proposed a method using GMM to encoding the forcing term in DMPs with more than one demonstration. Although applying multiple demonstrations in DMPs, each demonstration's goal was constant. In this paper, we propose a method to extract the feature of multiple demonstrations into one trajectory. According to the multi-demonstration of each target, we adopt the GMM method to integrate the underlying common features of multiple demonstrations, furthermore, to regenerate a distribution through GMR.

The forcing term of DMPs commonly is learned by LWR [8], GMM-GMR [3] or Reinforcement Learning such as PoWER [33]. The statistical method learning the parameter of the forcing term like LWR, GMM-GMR and other statistic learning method are focus on training on the same target, lack of integration and extraction of different targets' trajectory. Meanwhile reinforcement learning optimizes parameter by iterative update method, a kind of exploration optimization method, however the human demonstration does not play a major role in the process of skill learning. Reference [4] proposed a method to achieve avoiding both moving and stationary obstacle when manipulate a specify task. But it must detect obstacle position and shape information through vision. In this paper we propose a method integrating the features of the demonstration to different targets without detecting the specific obstacle position, and base on the LWR encoding the forcing term, we used the MV-GPR method to regenerate a new set of weights of basis function to modulate the shape of trajectory to infer the human intention to avoid the obstacle

III. METHOD

We assume that robot can detect the goal position, and the initial end-effector of robot is constant. For a specific task of reaching point such as peg-in-hole, robot senses the position of hole and then plan a no collision path to peg. The proposed method consists of two portions, firstly learning from demonstration, mainly extracting the trajectory shape parameters and the relationship between initial and target position. To reduce the impact of the difference of teachers and the destabilization of teaching movements, introducing the GMM-GMR method to refine and regenerate the reference trajectory for the DMPs encoding and combine the features encoded by DMPs for generalizing a trajectory to peg into the new hole.

Comparing with [7], Gaussian Process Regression (GPR) was applied to estimate the mapping between query points and goals of robot each joints. The input data can be multi-dimension, but output must be one-dimension. So it limits the learning efficiency and is hard to integrate input features into high-dimension output. We adopt MV-GPR method proposed in [34]. In the second part, extract the shape parameters of multiple demonstrations and analyze the human intention in the non-obstacle and obstacle situation with MV-GPR method and reconstruct DMPs model with the new forcing term. The overview of this approach is illustrated in Table 1.

TABLE 1. The algorithm framework of the proposed learning system.

1. Demonstration and Pre-processing
- Obtain demo $\Gamma=\{\xi_{t,i},\xi_{s,i}\}_{i=1}^{M}$ (M demonstrations)
- Regular trajectory by DTW
- GMM encode and GMR regress obtain $P(\xi^s \mid \xi^t) \sim \sum_{k=1}^{K} h_k \mathcal{N}(\hat{\xi}_k, \hat{\Sigma}_k)$
2. Model construction
- model DMPs using the pre-processing data $\Gamma' = \{\xi^t, \xi^s\}$
- Represent the forcing term by LWR and extract the parameter w as the training output $W \in R^{15}$
- Compute the difference position between initial end-effector and target hole $\Delta P \in R^3$ as the training input
- Put the training data $\{\Delta P, W\}$ into the MV-GPR to construct the regression model
3. Reproduction
- Detect the new hole position and compute the difference between the initial end-effector $\Delta P' \in R^3$
- Compute the value of w in new condition $W' \in R^{15}$
- Regenerate the trajectory by DMPs with new w

A. GAUSSIAN MIXTURE MODEL (GMM) AND GAUSSIAN MIXTURE REGRESSION (GMR)

Gaussian mixture model is to use Gaussian probability density function to precisely quantify model. It is a model that decomposes data into several Gaussian probability density functions. We have a dataset collected from kinesthetic guiding $\Gamma=\{\xi_{t,i},\xi_{s,i}\}_{i=1}^{M}$, it consists of M sets of data of dimensionality $T \times 4$. M is the number of the demonstration, T is the duration of each demonstration. t is the timesteps of each data. s represents the position of the end-effector in the Cartesian space $s \in R^3$. The duration of multiple demonstrations may be different. So we adopt DTW [25] method which can ensure the consistency of time.

$$p(k) = \pi_k \tag{1}$$

$$p(\Gamma|k) = \mathcal{N}(\Gamma; \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_k|}} e^{-\frac{1}{2}((\Gamma-\mu_k)^T \Sigma_k{}^{-1}(\Gamma-\mu_k))} \tag{2}$$

$$\mu_k = \begin{bmatrix} \mu_k^I \\ \mu_k^O \end{bmatrix}, \Sigma_k = \begin{bmatrix} \Sigma_k^I & \Sigma_k^{IO} \\ \Sigma_k^{OI} & \Sigma_k^O \end{bmatrix} \tag{3}$$

In which $\{\pi_k, \mu_k, \Sigma_k\}$ are the parameters of the kth Gaussian defining the prior, mean and covariance matrix respectively. Expectation-Maximization (EM) [33] is a local iterative algorithm used in statistics to find probability models that rely on unobservable hidden variables using to solve the Gaussian mixture model. The algorithm need to guarantee the appropriate initial of each component of the mixture model to avoid trapping into the local minima, so k-means clustering method is applied to initialize the mean of each component.

Gaussian mixture regression [36] is a regression technique reproduce the distribution of the new input, a kind of weighted sum method of the GMM.

$$p(\xi^s|\xi^t, k) \sim \mathcal{N}(\hat{\xi}_k, \hat{\Sigma}_k) \tag{4}$$

$$\hat{\xi}_k = \mu_k^s + \Sigma_k^{st}(\Sigma_k^t)^{-1}(\xi^t - \mu_k^t) \tag{5}$$

$$\hat{\Sigma}_k = \Sigma_k^s - \Sigma_k^{st}(\Sigma_k^t)^{-1}\Sigma_k^{ts} \tag{6}$$

$$P(\xi^s|\xi^t) \sim \sum_{k=1}^{K} h_k \mathcal{N}(\hat{\xi}_k, \hat{\Sigma}_k) \tag{7}$$

B. DYNAMIC MOVEMENT PRIMITIVES (DMPS)

Dynamic movement primitives is a method learning the motor action, which is able to encode both rhythmic and discrete movement. Our work only focuses on the discrete movement. The original formulation in paper [38], can solve the problem arose by changing goal position, such as tracking the moving object. A separated DMPs can be learned in different degrees of freedom (DOF), corresponding to the robotic arm, representing the joint of robot arm or the xyz axis position of end-effector in Cartesian space respectively. The canonical system has the function of normalizing time. x is set to 1 at the beginning generally and it monotonically decays to zero. τ is the temporal scaling factor, we set it equal to the execute time. α_x determines the convergence of x.

$$\tau\dot{x} = -\alpha_x x \tag{8}$$

The prototype of DMPs is a second-order damped spring system model [8]

$$\tau\ddot{y} = \alpha_z(\beta_z(g - y) - \dot{y}) + f \tag{9}$$

$$\tau\dot{z} = \alpha_z(\beta_z(g - y) - z) + f \tag{10}$$

$$\tau\dot{y} = z \tag{11}$$

α_z and β_z are the positive constants, and f is a nonlinear forcing term normalized linear combined of basis function. The width and center of basis function is set heuristically. The forcing term has the influence on the shape of trajectory, nevertheless the weight of each kernel is that determining factor.

$$f(x) = \frac{\sum_{i=1}^{N}\Psi_i(x)w_i}{\sum_{j=1}^{N}\Psi_j(x)}x(g - y_0) \tag{12}$$

$$\Psi_i(x) = \exp(-\frac{1}{2\sigma_i^2}(x - c_i)^2) \tag{13}$$

For adapting different goals autonomously, we online modify the forcing term, and employ $x(g - y_0)$ in $f(x)$. Conditioning on given position of the end-effector, we use difference methods to compute the velocity and acceleration. Inserting the above computed information in f_{target} obtain

$$f_{t\arg et} = \tau^2\ddot{y}_{demo} - \alpha_z(\beta_z(g - y_{demo}) - \tau\dot{y}_{demo}) \tag{14}$$

Now the problem converts into $J = min(f_{target} - f)$, the weight w is learned by approximate functions. In this paper we learn the parameter w of the demonstration by LWR, and use the MV-GPR mentioned in the Sec. C to regenerate.

C. MULTIVARIATE GAUSSIAN PROCESS REGRESSION (MV-GPR)

Gaussian process regression has been widely applied in the research of learning from demonstration, it is a powerful and effective method to process nonlinear regression problems. Despite the excellent properties of GPR, there are still exiting some obvious deficiencies. The majority of GPR models are implemented for single response variables, thus it is hard to solve the multi-response variables. So we apply the method MV-GPR to break this limitation. MV-GPR is a more straightforward method, and can be implemented in the same way as the conventional GPR, where the model settings, derivations and computations are all directly performed in matrix form, rather than vectorizing the matrices as done in the existing methods. This multi-output method can solve regression between the change of goal position and the multi-weight parameters in forcing term of dynamical movement primitives. Apply the covariance of the each shape parameters in the regression model, it could improve optimization.

The definition of multivariate Gaussian process is the first step of MV-GPR. Like the definition of Gaussian process, the multivariate Gaussian process should be a set of random vector-valued variables, and any number of random vector-valued variables have a matrix-variable Gaussian distribution. Therefore, the multivariate Gaussian process is defined as follows.

F is a multivariate Gaussian process of variable x, with mean function u: X $\mapsto R^d$ and kernel function k represents the covariance: X $\times$ X $\mapsto R$ and the positive semi-definite parameter matrix $\Omega \in R^{d\times d}$, if any finite collection of vector-valued variables have a joint matrix-variate Gaussian distribution

$$[f(x_1)^T, \ldots, f(x_n)^T]^T \sim MN(M, \Sigma, \Omega), n \in \mathrm{N} \tag{15}$$

we denote $f \sim MGP(u, k, \Omega)$. $f, u \in R^d$ are the row vectors consisting of the function $\{f_i\}_{i=1}^d$ and $\{\mu_i\}_{i=1}^d$ respectively. Moreover, $M \in R^{n\times d}$ with $M_{ij} = \mu_j(x_i)$ represents mean matrix, $\Sigma \in R^{n\times n}$ with $\Sigma_{ij} = k(x_i, x_j)$ and $\Omega \in R^{d\times d}$ represent the column covariance matrix and row covariance matrix respectively

Given n pairs of difference between the goal and initial position and the parameter w of LWR learning the forcing term in DMPs from the demonstration. $\{(x_i, y_i)\}_{i=1}^n, x_i \in R^p, y_i \in R^d$. These data satisfy the following assumption,

$$f \sim MGP(u, k', \Omega) \tag{16}$$

$$y_i = f(x_i), \quad for\ i = 1, \ldots, n \tag{17}$$

where $k' = k(x_i, x_j) + \delta_{ij}\sigma_n^2$, $\delta_{ij} = 1$ *if* $i = j$, otherwise $\delta_{ij} = 0$. Following the common set of GPR, the mean vector $u = 0$.

According to the definition of multivariate Gaussian process, the collection of function $[f(x_1)^T, \ldots, f(x_n)^T]^T$ satisfy the joint matrix-variate Gaussian distribution.

$$[f(x_1)^T, \ldots, f(x_n)^T]^T \sim MN(0, K', \Omega) \tag{18}$$

where K' is the $n \times n$ column covariance matrix of X, where the (i, j)-th component is $[K']_{ij} = k'(x_i, x_j)$. We define the kernel as Automatic Relevance Determination(ARD)

After training the model of MV-GPR, we predict the new parameter w with the test position of a new target hole collected through the vision system. The difference position between target and initial is $X_* = [x_{n+1}, \ldots, x_{n+m}]^{\mathrm{T}}$ and the predictive target w is $f_* = [f_{*1}, \ldots, f_{*m}]^{\mathrm{T}}$. The joint distribution of the training observations and the predictive vector of w is given as follow.

$$\begin{bmatrix} Y \\ f_* \end{bmatrix} \sim MN\left(0, \begin{bmatrix} K'(X,X) & K'(X_*,X)^{\mathrm{T}} \\ K'(X_*,X) & K'(X_*,X_*) \end{bmatrix}, \Omega\right) \tag{19}$$

Based on the principle of the condition distribution of Gaussian process, we can derive multivariate Gaussian process predictive distribution as follow.

$$p(f_*|X, Y, X_*) = MN(\hat{M}, \hat{\Sigma}, \hat{\Omega}) \tag{20}$$

where

$$\begin{aligned} \hat{M} &= K'(X_*, X)^{\mathrm{T}} K'(X, X)^{-1} Y, \\ \hat{\Sigma} &= K'(X_*, X_*) - K'(X_*, X)^{\mathrm{T}} K'(X, X)^{-1} K'(X_*, X), \\ \hat{\Omega} &= \Omega. \end{aligned} \tag{21}$$

Furthermore we also derive the expectation and covariance.

$$\begin{aligned} E[f_*] &= \hat{M} \\ &= K'(X_*, X)^{\mathrm{T}} K'(X, X)^{-1} Y \end{aligned} \tag{22}$$

$$\begin{aligned} cov(vec(f_*^{\mathrm{T}})) &= \hat{\Sigma} \otimes \hat{\Omega} \\ &= [K'(X_*, X_*) \\ &\quad - K'(X_*, X)^{\mathrm{T}} K'(X, X)^{-1} K'(X_*, X)] \otimes \Omega \end{aligned} \tag{23}$$

The forcing term of DMPs is the function mentioned as Sec. III-B. We reconstruct the function as follow

$$\begin{aligned} \mathrm{F} &= \Phi^{\mathrm{T}}\mathrm{W} \\ \Phi &= \left\{ \begin{matrix} \frac{\phi_1(x_1)}{\sum_{i=N}^{N} \phi_i(x_1)} x_1 \Delta G & \ldots & \frac{\phi_1(x_T)}{\sum_{i=N}^{N} \phi_i(x_T)} x_T \Delta G \\ \vdots & \ddots & \vdots \\ \frac{\phi_N(x_1)}{\sum_{i=N}^{N} \phi_i(x_1)} x_1 \Delta G & \cdots & \frac{\phi_N(x_T)}{\sum_{i=N}^{N} \phi_i(x_T)} x_T \Delta G \end{matrix} \right\} \\ \mathrm{W} &= \left\{ \begin{matrix} \omega_1(x) \\ \ldots \\ \omega_N(x) \end{matrix} \right\} \\ \Delta G &= (g - y_0) \end{aligned} \tag{24}$$

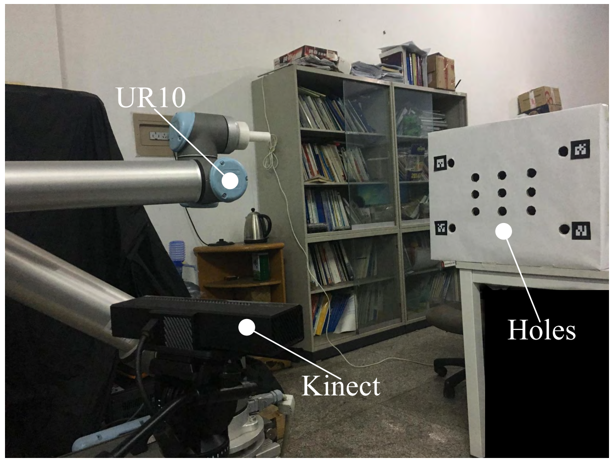

FIGURE 1. The setting of the experimental environment.

where N is the number of basis function and T is the number of timesteps. We can extract the weights of DMPs from the observation data, $\Xi = \{W^1 \ldots W^J\}$. J is the target number of demonstration.

Through the vision system the position of the target is available and the initial position of the robot is set stationary, $P_{target}\{x_i, y_i, z_i\}$ $i = 1 \cdots N$, $P_0\{x_0, y_0, z_0\}$

The difference between target and robot end-effector $\Delta P = P_{target} - P_0$ is used as input data of the MV-GPR model, and output the weights of the radial basis function.

Given J pairs of observation $\{(\Delta P_{axis}^i, W_{axis}^i)\}_{i=1}^J$ axis represent the x, y, z axis in Cartesian space respectively and $\Delta P_{axis} \in R$, $W_{axis} \in R^N$ We train the model through the formula (15)-(23) with above data. While, the hole transfers to another place, the new position detected by Kinect will be passed into the model to get the corresponding shape parameter ω of the new trajectory. By deriving from the formula (20)-(21), we obtain

$$E(W_*) = \hat{M} = K'(\Delta P_*, \Delta P)^{\mathrm{T}} K'(\Delta P, \Delta P)^{-1} W \tag{25}$$

$$\begin{aligned} \hat{\Sigma} &= K'(\Delta P_*, \Delta P_*) \\ &\quad - K'(\Delta P_*, \Delta P)^{\mathrm{T}} \\ &\quad \times K'(\Delta P, \Delta P)^{-1} K'(\Delta P_*, \Delta P). \end{aligned} \tag{26}$$

$$\begin{aligned} \mathrm{cov}(vec(W_*^{\mathrm{T}})) &= \hat{\Sigma} \otimes \hat{\Omega} = [K'(\Delta P_*, \Delta P_*) \\ &\quad - K'(\Delta P_*, \Delta P)^{\mathrm{T}} K'(\Delta P, \Delta P)^{-1} \\ &\quad \times K'(\Delta P_*, \Delta P)] \otimes \Omega \end{aligned} \tag{27}$$

With the multi-dimension output parameter w, We generalize a new peg-in-hole behavior by DMPs.

IV. EXPERIMENTS

In this section we test the performance of the method proposed in this paper through two experiments. In the experiment setup part, we select UR10 robot and the vision embedded in the robot system by hand-eye calibration. The fixed marker on the hole-box is detected by the Kinect RGB-D

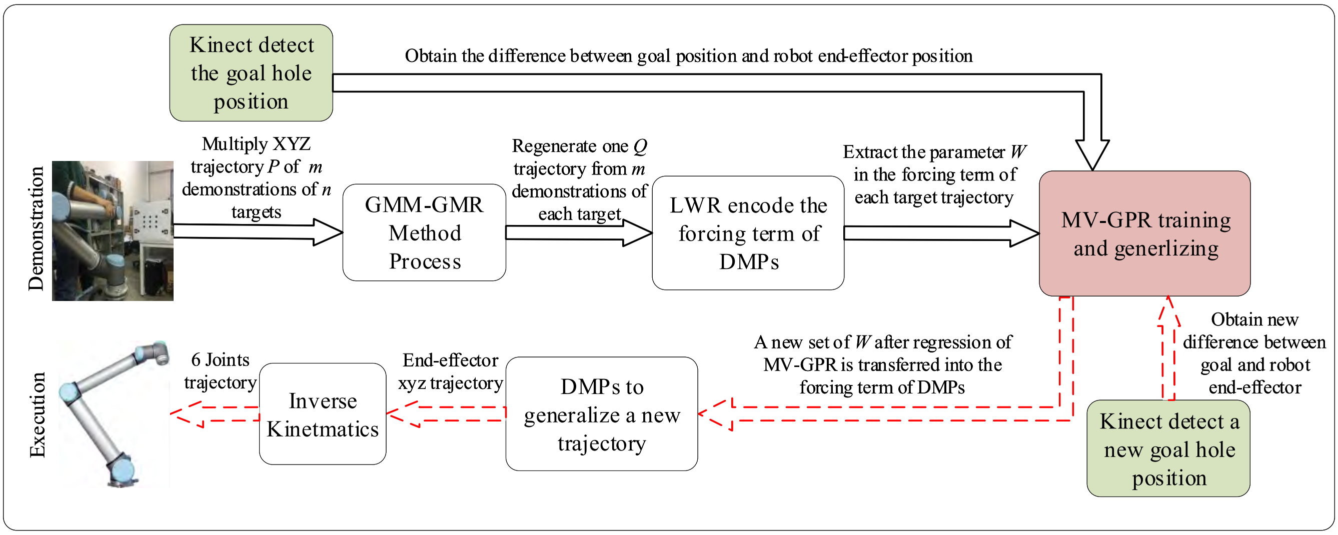

FIGURE 2. The block diagram of GMM-GMR data processing and MV-GPR-DMPs behavior generalizing.

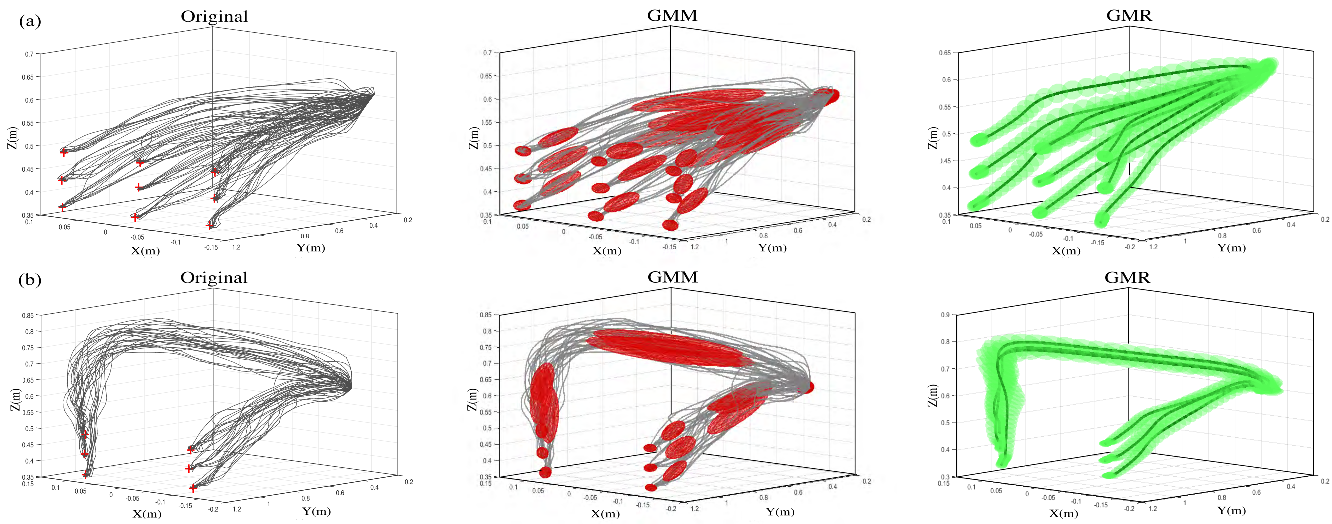

FIGURE 3. (a) In generalizing task, human dragged UR10 robot to the nine targets. (b) In the obstacle avoidance task, human demonstrated the UR10 robot to the six targets. These trajectories were encoded by the GMM method, and we selected 4 Gaussian models to characterize the curve (selected by Bayesian information criterion [39]). Obtain the normal trajectory by GMR.

camera using the wrapper of ROS to locate the position of the target hole, to simplify the experiment, we set the orientation of the target and end-effecter fixed as to verify the path planning ability.

The following tests are carried out for the discrete motion: (1) peg into a new hole to evaluate the ability to extract the characteristics of multi-objective demonstration trajectory. (2) take a stationary obstacle appearing during both teaching and execution step into consideration, the regenerated trajectory extracts the human intentions of the multi-objective demonstration trajectory to avoid the obstacle without knowing the specific obstacle's position.

The common part of The two experiments is training data acquisition. Firstly we demonstrate the peg-in-hole task by kinesthetic teaching, set the robot to gravity compensation mode, and record the end-effector's trajectory to the same target ten times. Using DTW method to regular time series trajectory before adopting the GMM-GMR method to encode these trajectories into the distribution offline and then reconstruct a trajectory. Fig. 1 shows the environment setting. The algorithm framework of the whole system is shown in Fig. 2.

A. PEG-IN-HOLE WITH GENERALIZATION ABILITY

Firstly, we drag UR10 to teach the skill of inserting into the nine holes. As shown in Fig. 3(a), there are ten trajectories to each holes. And these ninety trajectories are encoded by GMM, and regressed by GMR. The trajectories obtained by regression extract the characteristic of these

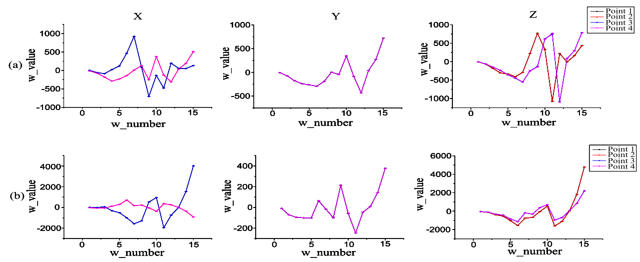

FIGURE 4. Mean and variance of the w value obtained by MV-GPR method in generalizing task. (a) presents the value of w in the generalizing task. (b) presents the value of w in the obstacle avoidance task.

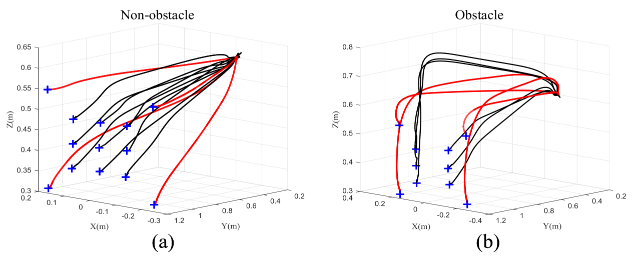

FIGURE 5. The generalizing trajectories of during the non-obstacle (a) and obstacle (b) peg-in-hole task. The black curves represent the normal trajectories processed by GMM-GMR, and the red curves represent the generalized trajectories processed by MV-GPR and DMPs. The symbol '+' represents the target hole.

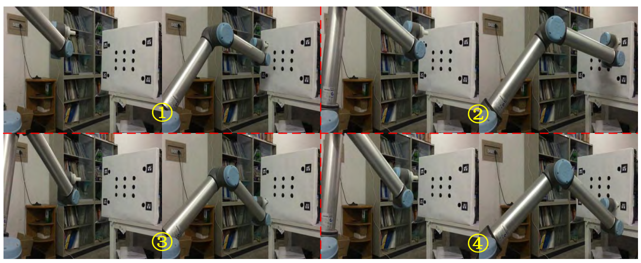

FIGURE 6. The snapshots during the generalizing task to the four targets.

ninety trajectories by statistical method. Comparing with the average method, the statistical method can eliminate the jerky signal during the kinesthetic teaching and obtain a smoother curve. We use these as normal trajectories to model the DMPs.

After using the LWR method to encoded the forcing term in DMPs of these normal trajectories, we obtain each demonstration feature parameters w. we choose the number of w heuristically as 15, which can encode the forcing term smoothly and accurately. After that we take the w value of the nine normal trajectories as the training data into the MV-GPR regression model, and the four new targets position detected by Kinect after processing as the test input data. As Fig. 4(a) shown, the feature parameters of trajectories to the four new targets are obtained. Corresponding to the position of the four targets, the w value of x and z axis are divided into two parts (x : $\{1, 3; 2, 4\}$, z : $\{1, 2; 3, 4\}$ where number is the mark of the hole, holes classified in the same group are approximately on the same axis), because the four holes are on the same plane x-z, so the values of w on y axis are similar. In Fig. 5(a), the trajectories of the four new targets could represent the proposed method ability of feature integration and generalization.

B. PEG-IN-HOLE WITH AVOIDANCE OBSTACLE ABILITY

For highlighting the application characteristics of this method obviously. We design a 6-DOF whole-arm obstacle avoidance task. With the intervention of human, we can get the rid of the limitation of the redundant degree of freedom. In this task, we just demonstrate only six holes, because of the existence of the obstacle in the 3D configuration space of the 6-DOF arm, some position cannot reach anyway. Firstly we set the stationary obstacle right in front of the hole-box without any external detection to this obstacle. Then we get the sixty trajectories to achieve the hole with avoidance ability. As Fig. 3(b) shown, the process is the same as the above mentioned.

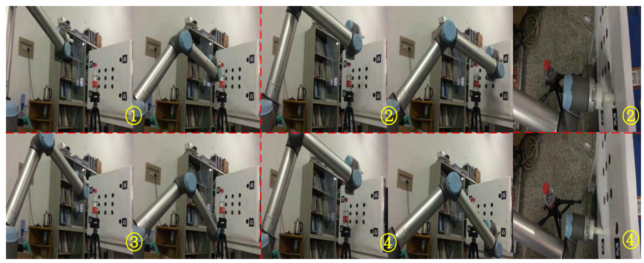

FIGURE 7. The snapshots during the generalizing task with obstacle avoidance to the four targets.

In Fig. 4(b), we get the generalized parameter w through the MV-GPR. We can find that w value of x axis has the complete difference from the generalizing task without obstacle, because of the existence of obstacle, the value of w in x and z have dramatically increased. The manipulation of task can be seen from Fig. 7. The robot can avoid the obstacle without knowing the specific position just analyzing the shape of the demonstration trajectory. The specific path is presented in Fig. 5(b). In Fig. 8, from two viewpoints, the shape of the path generated with the avoidance ability imitating the demonstration. In the plane of x-y and y-z when violating into the region of the obstacle, it generated a high curvature fragment trajectory within the range of demonstration.

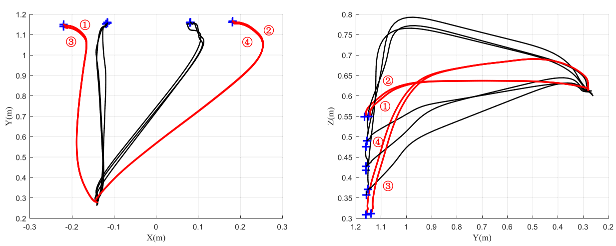

FIGURE 8. x-y and y-z orientation of the generalized trajectory in the obstacle avoidance task.

The target holes on the same axis have the similar shape. As above summarized, the trajectories are integrated with the human intention, so we can conclude that the method of MV-GPR can generalize a new set of feature parameters according to the new target under the reference of the demonstration.

V. CONCLUSION

In contrast to previous research on the DMPs in the robot learning from demonstration, we used the machine learning method GMM-GMR to encode the demonstration to refine the learning skills. The DTW method was used to regular the time series of demonstration preparing for GMM encoding. For getting rid of the shortcomings of DMPs just modelling a single trajectory, we adopted the statistical regression method MV-GPR to construct the shape parameter model of DMPs. This approach generalized the trajectory synthesizing the shape feature corresponding to the target position detected by Kinect. The generalization performance and flexibility were verified on the real robot arm UR10 for the point to point extrapolation reaching and 6-DOF whole-arm obstacle avoidance task. The results showed that with the combination characteristics of DMPs and MV-GPR this approach could extract and understand the human intention, and then the 6-DOF arm could achieve the obstacle avoidance without knowing the specific obstacle position.

This paper only described the aspect of trajectory skill learning. In the future work, we will take the obstacle position into consideration in dynamic environment, and combine it with the human intention to modulate the robot trajectory to achieve dynamic obstacle avoidance. We also plan to extend our proposed work on the force feedback control embedded in DMPs and incorporate the end-effecter orientation into the generated behavior and draw on the idea like ProMP, to introduce Bayesian regression method into the existing method.

REFERENCES

[1] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends Cognit. Sci.*, vol. 3, no. 6, pp. 233–242, 1999. doi: 10.1016/S1364-6613(99)01327-3.

[2] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robot. Auton. Syst.*, vol. 57, no. 5, pp. 469–483, 2009. doi: 10.1016/j.robot.2008.10.024.

[3] C. Yang, C. Chen, W. He, R. Cui, and Z. Li, "Robot learning system based on adaptive neural control and dynamic movement primitives," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 777–787, Mar. 2019. doi: 10.1109/TNNLS.2018.2852711.

[4] A. M. Ghalamzan, and M. Ragaglia, "Robot learning from demonstrations: Emulation learning in environments with moving obstacles," *Robot. Auton. Syst.*, vol. 101, pp. 45–56, Mar. 2018. doi: 10.1016/j.robot.2017.12.001.

[5] A. Pervez and D. Lee, "Learning task-parameterized dynamic movement primitives using mixture of GMMs," *Intell. Service Robot.*, vol. 11, no. 1, pp. 61–78, Jan. 2018. doi: 10.1007/s11370-017-0235-8.

[6] T. Matsubara, S.-H. Hyon, and J. Morimoto, "Learning parametric dynamic movement primitives from multiple demonstrations," *Neural Netw.*, vol. 24, no. 5, pp. 493–500, Jun. 2011. doi: 10.1016/j.neunet.2011.02.004.

[7] D. Forte, A. Gams, J. Morimoto, and A. Ude, "On-line motion synthesis and adaptation using a trajectory database," *Robot. Auton. Syst.*, vol. 60, no. 10, pp. 1327–1339, Oct. 2012. doi: 10.1016/j.robot.2012.05.004.

[8] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: Learning attractor models for motor behaviors," *Neural Comput.*, vol. 25, no. 2, pp. 328–373, 2013. doi: 10.1162/NECO_a_00393.

[9] T. Gašpar, B. Nemec, J. Morimoto, and A. Ude, "Skill learning and action recognition by arc-length dynamic movement primitives," *Robot. Auto. Syst.*, vol. 100, pp. 225–235, Feb. 2018. doi: 10.1016/j.robot.2017.11.012.

[10] T. Petrič, A. Gams, L. Colasanto, A. J. Ijspeert, and A. Ude, "Accelerated sensorimotor learning of compliant movement primitives," *IEEE Trans. Robot.*, vol. 34, no. 6, pp. 1636–1642, Aug. 2018. doi: 10.1109/TRO.2018.2861921.

[11] R. Vuga, B. Nemec and A. Ude, "Speed adaptation for self-improvement of skills learned from user demonstrations," *Robotica*, vol. 34, no. 12, pp. 2806–2822, Dec. 2016. doi: 10.1017/S0263574715000405.

[12] B. Nemec *et al.*, "Transfer of assembly operations to new workpiece poses by adaptation to the desired force profile," in *Proc. ICAR*, Nov. 2013, pp. 1–7.

[13] C. Yang, C. Zeng, C. Fang, W. He, and Z. Li, "A DMPs-based framework for robot learning and generalization of humanlike variable impedance skills," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 3, pp. 1193–1203, Jun. 2018. doi: 10.1109/TMECH.2018.2817589.

[14] R. Samant, L. Behera, and G. Pandey, "Adaptive learning of dynamic movement primitives through demonstration," in *Proc. IJCNN*, Jul. 2016, pp. 1068–1075.

[15] K. Muelling, J. Kober, and J. Peters, "Learning table tennis with a mixture of motor primitive," in *Proc. 10th IEEE-RAS Int. Conf. Humanoid Robots*, Dec. 2010, pp. 411–416.

[16] M. Hazara and V. Kyrki, "Model selection for incremental learning of generalizable movement primitives," in *Proc. ICAR*, Jul. 2017, pp. 359–366.

[17] F. Stulp, E. A. Theodorou, and S. Schaal, "Reinforcement learning with sequences of motion primitives for robust manipulation," *IEEE Trans. Robot.*, vol. 28, no. 6, pp. 1360–1370, Dec. 2012. doi: 10.1109/TRO.2012.2210294.

[18] L. E. Kavraki, P. Svestka, J.-C. Latombe, and M. H. Overmars, "Probabilistic roadmaps for path planning in high-dimensional configuration spaces," *IEEE Trans. Robot. Autom.*, vol. 12, no. 4, pp. 566–580, Aug. 1996. doi: 10.1109/70.508439.

[19] L. E. Kavraki, M. N. Kolountzakis, and J.-C. Latombe, "Analysis of probabilistic roadmaps for path planning," *IEEE Trans. Robot. Autom.*, vol. 14, no. 1, pp. 166–171, Feb. 1998. doi: 10.1109/70.660866.

[20] J. J. Kuffner and S. M. LaValle, "RRT-connect: An efficient approach to single-query path planning," in *Proc. ICAR*, Apr. 2000, pp. 995–1001.

[21] F. Zha, Y. Liu, X. Wang, F. Chen, J. Li, and W. Guo, "Robot motion planning method based on incremental high-dimensional mixture probabilistic mode," *Complexity*, vol. 2018, Nov. 2018, Art. no. 4358747. doi: 10.1155/2018/4358747.

[22] S. Dalibard and J.-P. Laumond, "Linear dimensionality reduction in random motion planning," *Int. J. Robot. Res.*, vol. 30, no. 12, pp. 1461–1476, Oct. 2011. doi: 10.1177/0278364911403335.

[23] M. Saha, J.-C. Latombe, Y.-C. Chang, and F. Prinz, "Finding narrow passages with probabilistic roadmaps: The small-step retraction method," *Auto. Robots*, vol. 19, no. 3, pp. 301–319, Dec. 2005. doi: 10.1007/s10514-005-4748-1.

[24] T. Alizadeh, M. Malekzadeh, and S. Barzegari, "Learning from demonstration with partially observable task parameters using dynamic movement primitives and Gaussian process regression," in *Proc. AIM*, Jul. 2016, pp. 889–894.

[25] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 286–298, Apr. 2007. doi: 10.1109/TSMCB.2006.886952.

[26] F. Stulp, G. Raiola, A. Hoarau, S. Ivaldi, and O. Sigaud, "Learning compact parameterized skills with a single regression," in *Proc. 13th IEEE-RAS Int. Conf. Humanoid Robots (Humanoids)*, Oct. 2013, pp. 417–422.

[27] S. Calinon, D. Bruno, and D. G. Caldwell, "A task-parameterized probabilistic model with minimal intervention control," in *Proc. ICAR*, May/Jun. 2014, pp. 3339–3344.

[28] J. Silvério, L. Rozo, S. Calinon, and D. G. Caldwell, "Learning bimanual end-effector poses from demonstrations using task-parameterized dynamical systems," in *Proc. IROS*, Sep. 2015, pp. 464–470.

[29] A. K. Tanwani and S. Calinon, "Learning robot manipulation tasks with task-parameterized semitied hidden semi-Markov model," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 235–242, Jan. 2016. doi: 10.1109/LRA.2016.2517825.

[30] C. Yang, C. Zeng, Y. Cong, N. Wang, and M. Wang, "A learning framework of adaptive manipulative skills from human to robot," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1153–1161, Feb. 2019. doi: 10.1109/TII.2018.2826064.

[31] A. Paraschos, C. Daniel, J. R. Peters, and G. Neumann, "Probabilistic movement primitives," in *Proc. NIPS*, 2013, pp. 2616–2624.

[32] K. Mülling, J. Kober, O. Kroemer, and J. Peters, "Learning to select and generalize striking movements in robot table tennis," *Int. J. Robot. Res.*, vol. 32, no. 3, pp. 263–279, Mar. 2013. doi: 10.1177/0278364912472380.

[33] J. Kober and J. R. Peters, "Policy search for motor primitives in robotics," in *Proc. NIPS*, 2009, pp. 849–856.

[34] Z. Chen, B. Wang, and A. N. Gorban. (2019). "Multivariate Gaussian and Student−t process regression for multi-output prediction." [Online]. Available: https://arxiv.org/abs/1703.04455

[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.

[36] D. H. Grollman and O. C. Jenkins, "Sparse incremental learning for interactive robot control policy estimation," in *Proc. ICRA*, May 2008, pp. 3315–3320.

[37] S. Schaal, "Dynamic movement primitives—A framework for motor control in humans and humanoid robotics," in *Proc. AMAM*, Mar. 2003, pp. 261–280.

[38] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Movement imitation with nonlinear dynamical systems in humanoid robots," in *Proc. ICRA*, May 2002, pp. 1398–1403.

[39] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

BOYANG TI received the B.S. degree in mechanical engineering from the Dalian University of Technology, Dalian, China, in 2017. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Robotics and System, Harbin Institute of Technology. His research interests include robotic path planning, robotic skill learning, and cooperative manipulation.

YONGSHENG GAO received the B.Sc., M.Sc., and Ph.D. degrees from the State Key Laboratory of Robotics Institute, Harbin Institute of Technology, Harbin, China, in 1994, 2001, and 2007, respectively, where he is currently a Professor. His research interests include pathological tremor suppress, tele-operation robot, and biomedical signal processing.

QIANG LI received the B.Sc. degree in mechanical engineering from the Dalian University of Technology, Dalian, China, in 2017. He is currently pursuing the master's degree with the School of Mechatronics Engineering, Harbin Institute of Technology, Harbin, China. His research interest includes robotic systems and control.

JIE ZHAO received the B.S. and Ph.D. degrees in mechatronics engineering from the Harbin Institute of Technology, Harbin, China, in 1990 and 1996, respectively, where he is currently a Professor with the School of Mechatronics Engineering and also the Director of the State Key Laboratory of Robotics and System. His research interests include industrial robots and bionic robots.