Reinforcement Learning of Manipulation and Grasping Using Dynamical Movement Primitives for a Humanoidlike Mobile Manipulator

Zhijun Li, *Senior Member, IEEE*, Ting Zhao, Fei Chen, *Member, IEEE*, Yingbai Hu, Chun-Yi Su, *Senior Member, IEEE*, and Toshio Fukuda, *Fellow, IEEE*

***Abstract*—It is important for humanoid-like mobile robots to learn the complex motion sequences in human–robot environment such that the robots can adapt such motions. This paper describes a reinforcement learning (RL) strategy for manipulation and grasping of a mobile manipulator, which reduces the complexity of the visual feedback and handle varying manipulation dynamics and uncertain external perturbations. Two hierarchies plannings have been considered in the proposed strategy: 1) high-level online redundancy resolution based on the neural-dynamic optimization algorithm in operational space; and 2) low-level RL in joint space. At this level, the dynamic movement primitives have been considered to model and learn the joint trajectories, and then the RL is employed to learn the trajectories with uncertainties. Experimental results on the developed humanoidlike mobile robot demonstrate that the presented approach can suppress the uncertain external perturbations.**

***Index Terms*—Dynamic movement primitive (DMP), mobile manipulation, redundancy resolution, reinforcement learning (RL).**

Manuscript received September 15, 2016; revised January 21, 2017 and May 3, 2017; accepted June 11, 2017. Date of publication June 20, 2017; date of current version February 14, 2018. Recommended by Technical Editor Y. Shen. This work was supported in part by the National Natural Science Foundation of China under Grant 61573147, Grant 91520201, and Grant 61625303; in part by the Guangzhou Research Collaborative Innovation Project under Grant 2014Y2-00507; in part by the Guangdong Science and Technology Research Collaborative Innovation Projects under Grant 2014B090901056 and Grant 2015B020214003; in part by the Guangdong Science and Technology Plan Project (Application Technology Research Foundation) under Grant 2015B020233006; and in part by the National High-Tech Research and Development Program of China (863 Program) under Grant 2015AA042303. *(Corresponding author: Zhijun Li.)*

Z. Li, T. Zhao, and Y. Hu are with College of Automation Science and Engineering, South China University of Technology, Guangzhou 510630, China (e-mail: zjli@ieee.org; zt20102011@163.com; 13249144573@163.com).

F. Chen is with the Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genova 16163, Italy (e-mail: fei.chen@iit.it).

C.-Y. Su is with the Department of Mechanical, Industrial, and Aerospace Engineering, Concordia University, Montreal, QC H4B 1R6, Canada (e-mail: chunyi.su@gmail.com).

T. Fukuda is with the School of Mechatronic Engineering, Beijing Institute of Technology, Beijing 100081, China (e-mail: tofukuda@nifty.com).

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/TMECH.2017.2717461

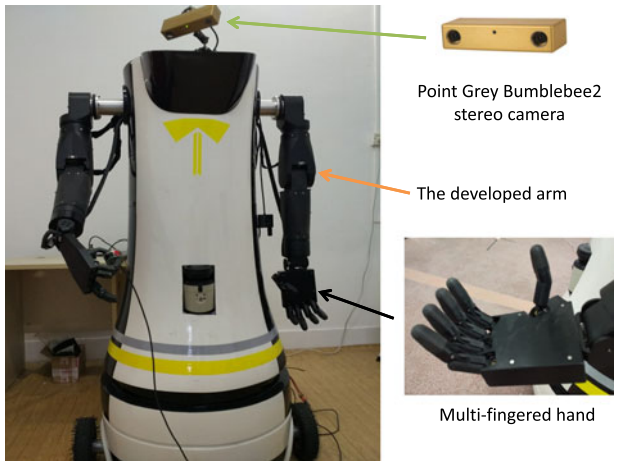


Fig. 1. Humanoidlike mobile manipulator system.

I. Introduction

COMPLEX motion sequences can be performed by human, e.g., grasp in manual assembly task, object manipulation. On the other hand, robots are becoming dexterous and making their way into human everyday environment [1], [2], [4], [7], [24], [25]. In such environments, for given tasks, the robot will have to generate appropriate trajectories based on the sensor feedback and existing knowledge. During the task executing, it is desirable that the robot can adapt trajectories by using the information feedback [8]. However, the external perturbations, lack of information, and errors in the perception of the environment might make the motion sequences deviated from the desired manipulatory actions. Therefore, autonomous exploration or learning is necessary for trajectory generation and makes it adaptable to the task. In this paper, we describe two hierarchies plannings for modulation and flexible learning of a humanoidlike mobile manipulator, so as to be able to reduce the complexity of the visual feedback and handle varying manipulation dynamics and uncertain perturbations. The developed humanoid mobile manipulator consists of a mobile platform with two robotic arms [five degrees of freedom (DOF) each, Fig. 1]. A dexterous hand is mounted at the end of each robot arm for manipulation purpose, being controlled via a controller area network (CAN) bus interface. It is apparent that the humanoidlike mobile manipulator is redundant in manipulation and grasping.

The reinforcement learning (RL) describing a state-action mapping in a discrete form has been extended successfully for robotic applications, such as grasping [30], leg robots [32], and vision-based robot navigation [31]. By maximizing the accumulation of reward functions using simple trial-and-error interactions with the environment, robots can acquire the desired task-specific behaviors. In [28], motor skill learning using the RL was proposed for impedance control of the robot to physically contact the environment. Equilibrium point control method and RL algorithm were integrated to determine the optimized impedance parameters. In [29], the obstacle avoidance technique was investigated by using redundant manipulators during manipulation tasks, where the inverse kinematics and obstacle avoidance in unknown environments were calculated by a dual-neural network (DNN) incorporating RL scheme. However, the main application challenges of the RL in robotic domains include the high-dimensional state and action spaces that bring heavy computational burdens; various physical limits that make applying the RL in actual robotic systems much more difficult than in simulations; uncertainties due to partial observability of the physical environments and inherent perturbation in the sensor measurements; and problems in designing task-dependent external rewards for the robots.

Considering a humanoidlike mobile manipulator with redundant joints, one fundamental issue for the motion sequence generation is to develop suitable redundancy-resolution approaches under various performance criteria to find an optimal solution to the robot kinematics. Many researchers have been working and contributing to the motion generation of humanoid mobile manipulators. In [5], a suboptimal trajectory generation was investigated by considering the mechanical and control constraints, where a parameterized curve was designed for the path planning of the robotic hand. For this purpose, in [11] and [12], the utilization of splines and wavelets was investigated. However, the splines were nonautonomous representations and lacked attractor properties. In [7], the combined planning for the constrained manipulation was investigated and improved by providing user demonstrations of the constrained manipulation motions. In [8], various sources of knowledge had been utilized by the robot to perform novel tasks, and the performance of the robot was improved by frequently occurring and stereotypical tasks. In [9], an elastic roadmap framework capable of satisfying various motion constraints and their respective feedback requirements were presented. From the above-mentioned works, we can see that various constraints exist in the mechanical systems, and an optimization for constraint problems has to be solved in the motion planning. Similarly, there are various physical limits existing in humanoid mobile manipulators, including, for example, the joint angle and velocity limits, and the rotational velocity limits in the manipulator and the wheels, respectively. Thus, it is necessary to develop a scheme that can address constraint optimization problem for physically constrained humanoid mobile manipulators.

Recently, in [1] and [10], the dynamic movement primitives (DMPs) have emerged as a parameterized movement scheme for trajectories learning due to their parameter linearity, rescaling robustness and continuity. The DMPs are described by a series of second-order differential equations, where the desired motion properties have been encoded. One of the most important features of the DMPs is that perturbations and feedback terms have been taken into account in units of actions describing or generating a particular motion trajectory either in operational or joint space. It is also insensitive to perturbations and the ability of generalization of the trajectory learning. Although by exploiting the special feature of the DMPs, a single manipulator can imitate and learn the trajectory of the given task, and can represent and generate humanlike motion sequences. Until now, to the authors' best knowledge, there is no work investigating dynamic motor primitives in depth for physically constrained redundant mobile manipulators.

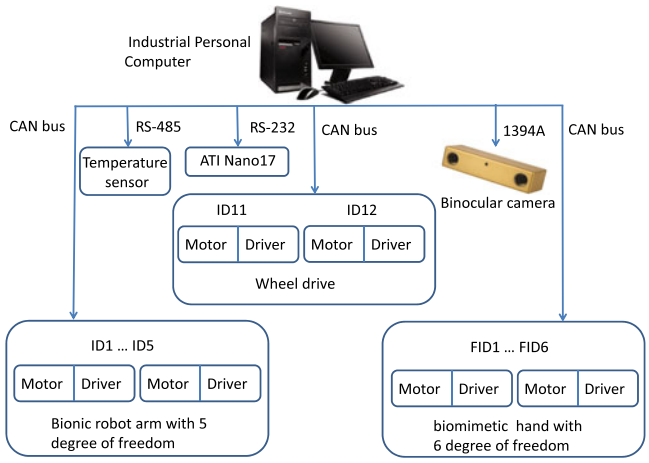


Fig. 2. Overall hardware framework.

This paper proposes a novel scheme for a coupled DMP-based motion sequence planning with redundancy resolution. The vision sensory feedback is used to record the position of the grasped object, and subsequently, the position information is exploited to formulate an operational space target point. Then, a redundancy resolution by optimizing the primal DNN (PDNN) is utilized to improve the motion performance of the robot. The optimal solutions obtained by the PDNN are treated as the target points of the dynamic motion primitives, and then DMPs are employed to model and learn the joint trajectories, whereas the RL improvement with path integrals is used to learn the trajectory models with manipulation uncertainties, after many iterations, bionic robot arm can reach the target position more accurately evenly with unknown external perturbations. Experimental results using the developed humanoidlike mobile robot demonstrate that this approach can suppress the uncertain perturbations compared to the traditional model-based control algorithms.

II. System Description

Fig. 1 illustrates a humanoidlike mobile robot with two differential driven wheels and a humanlike upper body, which has been developed in the South China University of Technology Laboratory. The humanlike upper body includes two robotic arms. There are six revolute joints in each arm, whose length is 1.05 m. The weight of each arm is 4.9 kg. The arm uses a Maxon dc flat brushless motor EC90 for joint 1 and EC45f for the rest joints. The harmonic transmission drives model SHD-17-100-2SH is used in joints 1 and 3, and model SHD-14-100-2SH

TABLE I
EXPERIMENT RESULT OF CAMERA CALIBRATION

Item	Calibration results
Focal length	$f_c = [1032.08148, 1061.13722]^T \pm [58.51339, 58.77151]^T$
Skew	$\alpha_c = [0.00000] \pm [0.00000]$
Principal point	$c_c = [120.32655, 204.91776]^T \pm [81.15977, 65.37172]^T$
Distortion	$k_c = [-0.09214, 0.28902, -0.00056, -0.00040, 0.00000]^T \pm [0.27165, 0.67260, 0.01489, 0.04206, 0.00000]^T$
Pixel error	$\text{err} = [1.88792, 1.53013]^T$

for joints 2, 4, and 5. Each joint uses a high-resolution encoder (1024 pulse/cycle) and Hall effect sensor to measure the displacement angles. The developed dexterous hand with five fingers (including one thumb) in the hand is used for grasping purpose. Each finger contains one joint actively driven by a dc motor with worm gear, and one joint passively driven by a linkage except the thumb. The thumb and index fingers also install ATI Nano17 force sensors and temperature-measuring thermocouple for the tactile sensing and ambient temperature measurement. With both bionic hand and arm, the robot can reach the resolution accuracy less than 1.0 mm, and the repetition accuracy less than 2.0 mm. Fig. 2 shows the system hardware architecture. The industrial personal computer communicates with the binocular camera and Elmo driver through the PCI board and CAN bus, respectively. The joint position information of the multiple fingers, arm joints, and wheels are communicated through CAN bus and controlled by Elmo drivers.

III. VISION SYSTEM

A. Visual Sensor Model

The vision system uses a Bumblebee stereo camera with IEEE-1394 Firewire communication produced by Point Grey. It utilizes two sensor progressive scan charge-coupled device cameras with fixed baseline working at a rate of 20 fps under VGA (640 × 480) mode. Object pose can then be estimated from the depth image and color image. For calibration, we utilize 20 checkerboard images into the calibration algorithm encompassing differential angles in the projection space, and generate enough data to estimate the camera extrinsic parameters [21]. We have calibrated and optimized the parameters through extracting the internal corner from the checkerboard pattern. The result is listed in Table 1. From the data, we can see that the distortion and Pixel error are both small enough for conducting the experiment, and the visual noise has little effect on the estimation of object pose.

B. Three-Dimensional (3-D) Reconstruction

Fig. 3 displays how the images are captured by the stereo camera in active ambient lighting. Similar texture across two coplanar image planes are then matched to search for the correspondence. The relationship between the columns of these corresponding pixels is disparity $\bar{d} = x_l - x_r$, where x_l and x_r denote the column values of left image pixel and right image pixel, respectively.

We make use of block matching algorithm to look for the corresponding pixels between the two images. A size of

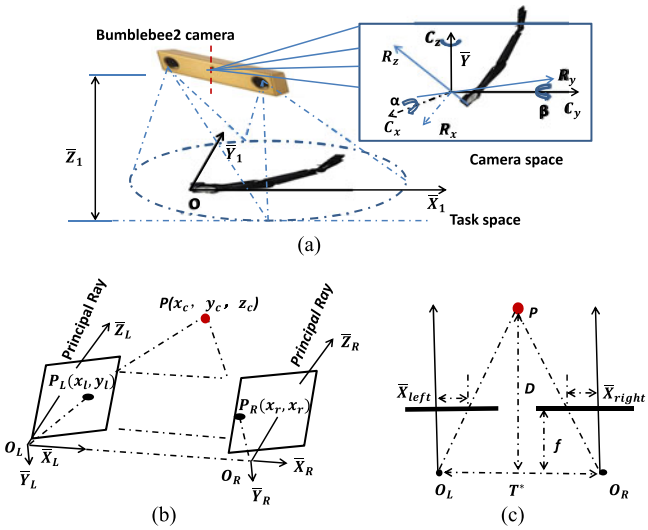


Fig. 3. Structure of the robotic stereo vision. (a) Schematic representation of the robot-camera system. (b) Front view of vision localization. (c) Vertical view of vision localization.

15 pixel window block is used to find the matches by the sum of absolute differences. Depth of the pixel can then be calculated by the following equation:

$$D = T^* \frac{f}{\bar{d}} \tag{1}$$

with the focal length f, the disparity $\bar{d}$, and the baseline T^*. The 3-D reconstruction of the robot workspace is described as

$$\mathcal{Q}\left[\bar{x},\ \bar{y},\ \bar{d},\ 1\right]^T = \left[\bar{X},\ \bar{Y},\ \bar{Z},\ \bar{W}\right]^T \tag{2}$$

where $\bar{x}$ and $\bar{y}$ denote image plane coordinates of the feature point, with the depth D, the projection matrix $\mathcal{Q}$, and 3-D world coordinates $\bar{X}/\bar{W}$, $\bar{Y}/\bar{W}$, $\bar{Z}/\bar{W}$.

C. Object Detection

In this paper, a single-color object in red is easily extracted from the background with the method of color-based segmentation. The image is converted into the $\bar{L} \times \bar{a} \times \bar{b}$ color space. In this space, the Euclidean distance between $(L_1{}^*, a_1{}^*, b_1{}^*)$ and $(L_2{}^*, a_2{}^*, b_2{}^*)$ can be defined as

$$\Delta E_{ab} = \sqrt{(L_2{}^* - L_1{}^*)^2 + (a_2{}^* - a_1{}^*)^2 + (b_2{}^* - b_1{}^*)^2} \tag{3}$$

where the L^* is luminance component, and a^* and b^* are the color components.

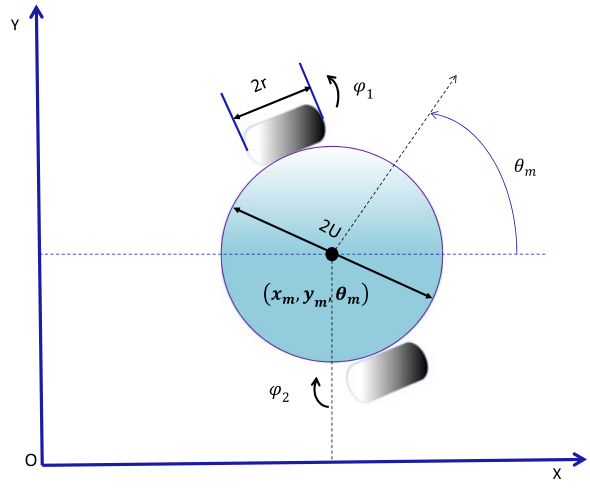


Fig. 4. Kinematic scheme of the wheeled platform.

IV. Kinematics of Humanoidlike Mobile Manipulator and Motion-Force Optimization

Fig. 4 shows that the pose of the mobile manipulator can be described by a three-tuple $X = [x_m, y_m, \theta_m]^T \in R^l$, where x_m and y_m denote the coordinates of the platform centre, θ_m is the orientation angle of the platform in a global world OXY, φ_1, φ_2 denote the angles of driving wheels, $2U$ is the distance between platform wheels, r denotes the wheel radius, $Y \in R^n$ is the joint coordinate of the manipulator mounted on the platform, and n is the degrees of freedom.

The nonholonomic platform is usually subject to $1 \leq k < l$ constraints described in the following form:

$$J(X)\dot{X} = 0 \tag{4}$$

where $l \geq 1$, $J(X) = [\sin\theta_m, -\cos\theta_m, 0]$ denotes the $k \times l$ matrix of the full rank (that is, $\text{rank}(J(X)) = k$). Assume that $\text{Ker}(J(X))$ is spanned over the fields $J_1(X), \ldots, J_{l-k}(X)$, then the analytic driftless kinematic system is described with motion constraint (4) as

$$\dot{X} = N(X)\mathcal{J} \tag{5}$$

where $N(X) = [\eta_1(X), \ldots, \eta_{l-k}(X)]$ and $\text{rank}(N(X)) = l - k$. The auxiliary velocities of the platform can be expressed as $\mathcal{J} = [\dot{\varphi}_2, \dot{\varphi}_1]^T$. For the mobile platform depicted in Fig. 4, and the matrix $N(X)$ is denoted as $N(X) = [r\cos\theta_m/2, r\cos\theta_m/2; r\sin\theta_m/2, r\sin\theta_m/2; r/2U, -r/2U]$.

The pose p of the robot end-effector is expressed as

$$p = f(q) \tag{6}$$

where $p = [p_1, \ldots, p_m]^T$, and $q = [X^T, Y^T]^T$ is the configuration of the mobile manipulator. The m-dimensional nonlinear mapping is described by $f(q) = [f_1(q), \ldots, f_m(q)]^T$, where m indicates the dimension in the task space. We can obtain the following relationship by combining $\dot{q}$ and $\ddot{q}$ with (6) as

$$\dot{q} = Cs, \ \ddot{q} = C\dot{s} + \dot{C}s \tag{7}$$

with $C = \text{diag}[N(X), I_n]$ and the simplified velocity vector of the mobile manipulator is described as $s = [\mathcal{J}, \dot{Y}]^T$, I_n represents the $n \times n$ unit matrix. The derivative $\dot{p}$ can be expressed as

$$\dot{p} = \mathbb{J}s \tag{8}$$

where $\mathbb{J} = JC$, and $J = \partial f/\partial q$ is the Jacobian matrix with $m \times (n + l)$ dimension.

A. Kinematics Redundant Resolution Formulation

It should be noted that since $m < (n + l)$ in redundant mobile manipulators, we have infinitely many solutions. Consider the traditional pseudoinverse-type solution for (8) with one minimum-norm particular solution, $\mathbb{J}^+\dot{p}$ and the same type of solution, $(I - \mathbb{J}^+\mathbb{J})\varpi$, for example

$$\dot{s} = \mathbb{J}^+\dot{p} + (I - \mathbb{J}^+\mathbb{J})\varpi \tag{9}$$

where $\mathbb{J}^+ \in \mathbb{R}^{(n+l)\times m}$ represents the pseudoinverse of $\mathbb{J}$. We can select any vector $\varpi \in \mathbb{R}^{(n+l)}$ as a gradient of optimization performance index if the physical limit of joints, singularity points, as well as environmental barriers need to be considered.

Considering the Jacobian (9) of the mobile manipulator, we choose the following objective function as

$$\begin{aligned} &\text{minimize} && \dot{s}^T\dot{s}/2 && (10)\\ &\text{subject to} && \mathbb{J}s = \dot{p} && (11)\\ & && s_{\min} \leqslant s(t) \leqslant s_{\max} && (12)\\ & && \dot{s}_{\min} \leqslant \dot{s}(t) \leqslant \dot{s}_{\max} && (13) \end{aligned}$$

we denote the boundedness of the joint position and velocity as $s_{\min}$ and $s_{\max}$, $\dot{s}_{\min}$ and $\dot{s}_{\max}$, respectively, which could be rewritten as

$$\begin{aligned} &\text{minimize} && \dot{s}^TQ\dot{s}/2 + b^T\dot{s} && (14)\\ &\text{subject to} && \mathbb{J}s = d && (15)\\ & && \xi_{\min} \leqslant As \leqslant \xi_{\max} && (16) \end{aligned}$$

where $Q = I, b = 0, d = \dot{p}$, $\xi_{\min}$ and $\xi_{\max}$ are the boundedness of the velocity, and $A = \text{diag}\,[a_1, \ldots, a_n]$ is the positive definite diagonal matrix.

Consider the joint physical constraints, the could be formulated in the quadratic program (QP) problem

$$\begin{aligned} &\text{minimize} && x^TQx/2 + b^Tx && (17)\\ &\text{subject to} && J_R(q)x = d && (18)\\ & && \xi_{\min} \leqslant Ax \leqslant \xi_{\max} && (19) \end{aligned}$$

where the decision vector x is defined as $\dot{q}$ in velocity-level schemes, and the coefficients (i.e., W, b, d, $\xi_{\min}$, and $\xi_{\max}$) are defined corresponding to a specific redundancy-resolution scheme.

B. Grasping-Force Optimization Formulation

Considering object grasping problem with a dexterous hand, there are three kinds of contact model: frictionless point contact, point contact with friction, and soft-finger contact. In this paper, we consider the point contact model with friction and the finger joint torque is used to satisfy the constraints of the frictions.

The contact wrench is denoted by $f_g = [f_1^T, \ldots, f_l^T]^T \in R^{\iota h}$, where ι denotes the number of grasping contact points, $f_i \in R^h$ is the ith contact vector with dimension h, and $f_e \in R^6$ is the generalized external force acting on the object. The contact wrench includes three vectors: one normal vector $f_{i,z}$ to the vertical plane, and two vectors $f_{i,x}$, $f_{i,y}$ on the tangent plane. We can describe the friction constraint as

$$\frac{1}{\mu_i}\sqrt{f_{i,x}^2 + f_{i,y}^2} \leq f_{i,z} \tag{20}$$

$$\tau_L \leq \tau \leq \tau_U \tag{21}$$

with the tangential friction coefficient μ_i at the ith contact point and the upper and lower joint torque bound τ_L and τ_U, respectively.

It is easy to have the following equations as [36]:

$$f_e = G f_g \tag{22}$$

$$A^T(\vartheta) f_g + \tau_e = \tau \tag{23}$$

with full-rank grasp matrix $G \in R^{6\times \iota h}$, the external torque τ_e, the hand Jacobian matrix $A(\vartheta) \in R^{\iota h \times \varrho}$, and the finger joint angle ϑ. Therefore, considering the above-mentioned equations, we can present optimization problem of the grasping force as seeking the optimal grasp wrench, which can minimize the internal forces acting on the object, and satisfy the friction cone constraints within the null space of the grasp matrix G. We can present the frictional inequalities (20) and the torque limit constraint (21) as

$$F(c) = \text{diag}[F_1(c_1), \ldots, F_i(c_i), \ldots, F_n(c_n)] > 0 \tag{24}$$

where $F_i(f_i)$ is the symmetric matrix

$$F_i(c_i) = \begin{bmatrix} f_{i,z} + \dfrac{f_{i,x}}{\mu_i} & \dfrac{f_{i,y}}{\mu_i} \\ \dfrac{c_{i,y}}{\mu_i} & f_{i,z} - \dfrac{f_{i,x}}{\mu_i} \end{bmatrix} \tag{25}$$

$$T(f_g, \vartheta, \tau_e) = \text{diag}(\tau_B) > 0 \tag{26}$$

with

$$\tau_B = \begin{bmatrix} \tau_{B,L} \\ \tau_{B,H} \end{bmatrix} = \begin{bmatrix} A^T(\vartheta) f - \tau_L + \tau_e \\ -A^T(\vartheta) f + \tau_h - \tau_e \end{bmatrix} \tag{27}$$

and $(\tau_{B,L})$ and $(\tau_{B,H})$ are the lower and upper limits, respectively. Therefore, the frictional and joint torque constraints can be described as

$$\mathcal{T} = \text{diag}(F, T) > 0. \tag{28}$$

The matrix $\mathcal{T}$ with a linear constraint can be described as [36]

$$A\zeta(\mathcal{T}) = b \tag{29}$$

with $\zeta(\mathcal{T}) = [f_g(F)^T, \tau_B(T)^T]^T$, $S = \begin{bmatrix} G & 0 & 0 \\ A^T(\vartheta) & -I & 0 \\ A^T(\vartheta) & 0 & I \end{bmatrix}$, and $b = \begin{bmatrix} f_e \\ \tau_L - \tau_e \\ \tau_H - \tau_e \end{bmatrix}$, where the null matrix equal to 0 and the identity matrix of proper dimension is I. With the purpose of minimizing the cost function $\Phi(\mathcal{T})$, considering the symmetric positive definite matrix Q, and the boundedness $\beta_{\min}$ and $\beta_{\max}$ of $\mathcal{T}$, the optimization can be defined as [36]

$$\Phi(\mathcal{T}) = \mathcal{T}^T Q \mathcal{T} + b^T \mathcal{T}(t) \tag{30}$$

$$\text{subject to} \quad S\zeta(P) = b \tag{31}$$

$$\beta_{\min} \leqslant \mathcal{T} \leqslant \beta_{\max}. \tag{32}$$

C. Neurodynamics Optimization

In this paper, the QP problems (17)–(19) and (30)–(32) can be simplified into a gradient linear variational inequalities (LVI) based PDNN [15] (17)–(19).

First, we can make the QP problem (17)–(19) and (30)–(32) to be a set of LVI. The primal-dual equilibrium vector $u^* \in \Omega = \{v | u^- \leq v \leq u^+\}$ with appropriate dimension, meets the following constraint:

$$(u - u^*)^T(\mathbb{M}u^* + \rho) \geqslant 0 \quad \forall u \in \Omega \tag{33}$$

where the variable vector u and its upper/lower bounds are denoted by $u^+ = [\xi^+, +y^+, +y^+]^T, u^- = [\xi^-, -y^+, -y^+]^T$. Given arbitrary i, the elements $y_i^+ \gg 0$ in y^+ are sufficiently positive to represent $+\infty$. The coefficients $\mathbb{M}$ and ρ are described as

$$\mathbb{M} = \begin{bmatrix} Q & -G^T & S^T \\ G & 0 & 0 \\ -S & 0 & 0 \end{bmatrix}, \ \rho = \begin{bmatrix} b \\ -\mathcal{W} \\ C \end{bmatrix}. \tag{34}$$

Theorem 1: [15]: The QP problem (17)–(19) and (30)–(32) could be transformed into the LVI problem (33).

Second, the LVI (33) is equivalent to the following piecewise linear equations [15]:

$$P_\Omega\left(u - (\mathbb{M}u + \rho)\right) - u = 0 \tag{35}$$

where $P_\Omega(\cdot)$ stands for the projection operator onto Ω and defined as $P_\Omega(u) = [P_\Omega(u_1), \ldots, P_\Omega(u_\kappa)]^T$ with

$$P_\Omega(u_i) = \begin{cases} u_i^- & \text{if } u_i < u_i^- \\ u_i & \text{if } u_i^- \leqslant u_i \leqslant u_i^+ \\ u_i^+ & \text{if } u_i > u_i^+. \end{cases}$$

It is feasible to set up a dynamical system to solve the linear projection equation (35) by following the dual dynamical system design approach. We can simply define the error function:

$$E(u) = P_\Omega\left(u - (\mathbb{M}u + \rho)\right) - u. \tag{36}$$

Obviously, the optimization purpose is to devise an iteration approach so that the deviation in (36) converges to zero.

In order to obtain a primal-dual equilibrium vector u^* in (33), we can examine $u(t)$, for $t = 0, 1, 2, \ldots, \iota$. If $u(t) \neq u^*$, we use the following iteration approach to solve (36):

$$u(t+1) = u(t) + \frac{\|E(u(t))\|^2}{\|(I + \mathbb{M}^T)E(u(t))\|^2}\Gamma E(u(\iota)) \tag{37}$$

where Γ stands for a positive parameter in design, which is chosen as $\Gamma = 10^{10}$ in the experiments.

Theorem 2: [15]: Considering the state vector $u(t)$ of (37) converging to an equilibrium point u^* from any initial state, the first n elements are the optimal solution F_g^* to the QP problem in (17)–(19). Moreover, the exponential convergence can be achieved for a constant $\varrho > 0$ such that $\|E(u)\|_2^2 \geqslant \varrho\|u - u^*\|_2^2$.

Remark 1: The implementation of the PDNN only needs simple vector or matrix augmentation and operation. Compared with the SQP methods, which utilize a traditional gradient descent and need repeatedly calculating the Hessian matrix, it has higher computational complexity than the PDNN [34].

V. RL With Dynamical Movement Primitives

In order to fulfill the eye-hand coordination, the complete motion can be decomposed into goal reaching and object grasping. Considering the redundancy of the mobile manipulator, the proposed two hierarchies plannings consist of high-level online redundancy resolution based on the neural-dynamic optimization algorithm in operational space, and low-level RL in joint space. In the previous section, the problem of redundant planning in operational space and the corresponding solution have been proposed. In this section, the redundant resolutions are regarded as the target positions in the joint space, and the DMP-based RL is employed to learn the trajectories with uncertainties.

The previous works commonly assume that the humanoid mobile manipulator has full geometry and pose knowledge of the objects for the manipulation planning tasks [26], [27]. However, that planning might not be successful in practice due to uncertainty perturbations, even if theoretical grasps and motion planning are successful. Nevertheless, the DMPs can be employed to generate goal directed movements with generalization and learning either in joint or task space due to its capability of being robust to those perturbations.

A. Dynamical Movement Primitives

The DMP is a flexible representation for motion primitives [1] and [10], being expressed as

$$\frac{1}{\varsigma}\dot{z}_t = \alpha_z(\beta_z(g - y_t) - z_t) + f_t \tag{38}$$

$$\frac{1}{\varsigma}\dot{y}_t = z_t \tag{39}$$

$$\frac{1}{\varsigma}\dot{x}_t = \alpha_x x_t \tag{40}$$

$$f_t = h_t^T(\theta + \varepsilon_t) \tag{41}$$

where y_t and $\dot{y}_t$ indicate the trajectory position and trajectory velocity, respectively, z_t and x_t are internal states, α_z, β_z, α_x, ς are the scale factors, θ influences the shape of the trajectory, which determines the movement shape, e.g., the joint trajectory $[\ddot{q}_t, \dot{q}_t, q_t]^T$ over time. The goal g influences the final configuration of the movement, e.g., the joint positions at the end of the movement, and f_t is defined as a nonlinear function allowing the generation of arbitrary complex movements, which consists of the basis functions $h_t \in R^{\zeta\times 1}$ by a piecewise linear function approximator with Gaussian weighting kernels as

$$h_t = \frac{\sum_{i=1}^{N}\omega_i(x)x_t}{\sum_{i=1}^{N}\omega_i(x)}(g - y_0) \tag{42}$$

$$\omega_i(x) = \exp(-\frac{1}{2\sigma_i}(x_t - c_i)^2) \tag{43}$$

with the width σ_i and the center c_i. We can see that the function f_t is related to a phase variable x_t, which varies from 1 to 0 during a movement by (40).

B. RL With Optimal Trajectory

In the RL, many works have been proposed by empirical learning and statistical technique for the parameter θ updating. Integrated probability-weighted RL method is used in this paper. From the definition (38)–(41), we can define a cost function as

$$S(\varsigma_i) = \phi_{t_N} + \sum_{j=i}^{N-1} r_{t_j} + \frac{1}{2}\sum_{j=i}^{N-1}(\theta_{t_j} + \varepsilon_{t_j})^T M_{t_j}^T R M_{t_j}(\theta_{t_j} + \varepsilon_{t_j}) \tag{44}$$

$$M_{t_j} = \frac{R^{-1}g_{t_j}g_{t_j}^T}{g_{t_j}^T R^{-1} g_{t_j}} \tag{45}$$

where $S(\varsigma_i)$ is the finite horizon cost, ϕ_{t_N} is the terminal cost at time t_N, r_{t_j} is the immediate cost, and ε_{t_j} represents the Gaussian white noise added to the parameter space. According to the stochastic optimal control theory [17], the definition of path integral for DMPs can be written as

$$\mu_{t_i} = \int P(\varsigma_i)\mu(\varsigma_i)d\varsigma_i \tag{46}$$

where μ_{t_i} is the local control and $P(\varsigma_i)$ is the probability, which are described as $P(\varsigma_i) = \frac{e^{-\frac{1}{\lambda}S(\varsigma_i)}}{\int e^{-\frac{1}{\lambda}S(\varsigma_i)}d\varsigma_i}$ and $\mu(\varsigma_i) = \frac{R^{-1}g_{t_i}g_{t_i}^T}{g_{t_i}^T R^{-1}g_{t_i}}\varepsilon_{t_i}$, respectively.

Given an initial value of θ, generating stochastic parameters $\theta + \varepsilon_t$ at every time step, and considering the definition (46), we can obtain the equations of the update of the parameter vector θ, which can be written as [37]

$$\begin{aligned}\theta_{t_i}^{\text{update}} &= \int P(\varsigma_i)\frac{R^{-1}g_{t_i}g_{t_i}^T(\theta + \varepsilon_{t_i})}{g_{t_i}^T R^{-1}g_{t_i}}d\varsigma_i \\ &= \int P(\varsigma_i)\frac{R^{-1}g_{t_i}g_{t_i}^T\varepsilon_{t_i}}{g_{t_i}^T R^{-1}g_{t_i}}d\varsigma_i + \frac{R^{-1}g_{t_i}g_{t_i}^T}{g_{t_i}^T R^{-1}g_{t_i}}\theta \\ &= \delta\theta_{t_i} + \frac{R^{-1}g_{t_i}g_{t_i}^T}{g_{t_i}^T R^{-1}g_{t_i}}\theta \\ &= \delta\theta_{t_i} + M_{t_i}\theta \end{aligned} \tag{47}$$

$$\delta\theta_{t_i} = \int P(\varsigma_i)\frac{R^{-1}g_{t_i}g_{t_i}^T\varepsilon_{t_i}}{g_{t_i}^T R^{-1}g_{t_i}}d\varsigma_i. \tag{48}$$

From (47), we can see that $\theta_{t_i}^{\text{update}}$ is time dependent. In order to obtain one single time-independent parameter vector θ^{update}, the

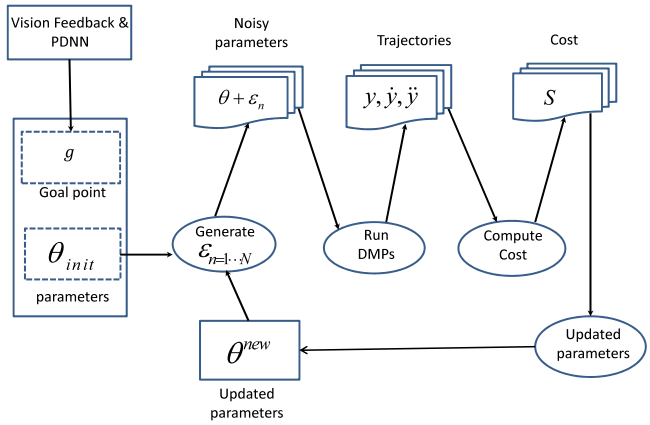


Fig. 5. Generic loop of the RL algorithm.

vectors $\theta_{t_i}^{\text{update}}$ should be averaged over time t_i as

$$\theta^{\text{update}} = \frac{1}{N}\sum_{t=0}^{N-1}\theta_{t_i}^{\text{update}} = \frac{1}{N}\sum_{t=0}^{N-1}\delta\theta_{t_i} + \frac{1}{N}\sum_{t=0}^{N-1}M_{t_i}\theta_{t_i}. \tag{49}$$

Considering the parameter vector to be constant and time independent, by eliminating the project matrix M_{t_i} onto the range space of g_{t_i} under the metric R^{-1}, we can rewrite (49) as

$$\theta^{\text{update}} = \frac{1}{N}\sum_{t=0}^{N-1}\delta\theta_{t_i} + \frac{1}{N}\sum_{t=0}^{N-1}\theta_{t_i} = \frac{1}{N}\sum_{t=0}^{N-1}\delta\theta_{t_i} + \theta \tag{50}$$

where θ^{update} can be obtained by the average value of $\delta\theta_{t_i}$ and adding the current value of θ in each iteration, we can modify the cost function (44) as

$$S(\varsigma_i) = \phi_{t_N} + \sum_{j=i}^{N-1} r_{t_j} + \frac{1}{2}\sum_{j=i}^{N-1}(\theta + M_{t_j}\varepsilon_{t_j})^T R(\theta + M_{t_j}\varepsilon_{t_j}). \tag{51}$$

Finally, Fig. 5 demonstrates the generic loop of the generalized path integral algorithm. We can summarize the calculation procedure as follows:

$$P(\varsigma_i) = \frac{e^{-\frac{1}{\lambda}S(\varsigma_i)}}{\int e^{-\frac{1}{\lambda}S(\varsigma_i)}d\varsigma_i} \tag{52}$$

$$S(\varsigma_i) = \phi_{t_N} + \sum_{j=i}^{N-1} r_{t_j} + \frac{1}{2}\sum_{j=i}^{N-1}(\theta + M_{t_j}\varepsilon_{t_j})^T R(\theta + M_{t_j}\varepsilon_{t_j}) \tag{53}$$

$$\delta\theta_{t_i} = \int P(\varsigma_i)\frac{R^{-1}g_{t_i}g_{t_i}^T\varepsilon_{t_i}}{g_{t_i}^T R^{-1}g_{t_i}}d\varsigma_i \tag{54}$$

$$[\delta\theta]_j = \frac{\sum_{k=1}^{K}(N-i)\omega_{j,t_i}[\delta\theta_{t_i}]_j}{\sum_{i=0}^{N-1}(N-i)\omega_{j,t_i}} \tag{55}$$

$$\theta^{\text{update}} = \theta + \delta\theta. \tag{56}$$

Algorithm 1: The Update Rule.

Given:

1. Initialize a cost function r_t, terminal cost ϕ_{t_N}, a stochastic parameterized policy f_t, g_t, Σ_ε, θ, initial state x_0.

Repeat until the convergence of R using stochastic parameters $\theta + \varepsilon_t$:

2. **For** $k = 1, \ldots, K$, calculate (52), and (53);
3. **For** $i = 1, \ldots, (N-1)$, calculate (54) and (55);
4. Update θ using $\theta + \delta\theta$;
5. Calculate $\phi_{t_N} + \sum_{i=0}^{N-1}(r_{t_i} + \theta_t^T R\theta_t)$.

C. RL With Optimal Target

By updating the parameter vector θ, the DMP is estimated with the new θ^{update}, without noise, i.e., $\varepsilon = 0$, to generate different movements with lower costs. The process then continues with the new θ as the basis for exploration. However, for different position of end-effectors, we need to learn the goal exploration by the DMP parameter optimization, the goal and trajectory exploration are both extended as

$$\frac{1}{\varsigma}\dot{g}_t = \alpha_g(g + \{\epsilon^g\}_k - g_t) \tag{57}$$

$$\frac{1}{\varsigma}\dot{z}_t = \alpha(\beta(g_t - y_t) - \dot{y}_t) + h_t^{\mathrm{T}}(\theta + \{\varepsilon_t\}_k). \tag{58}$$

As similar as trajectory exploration, the goal exploration can be added with the Gaussian noise ϵ^g with the variance δ_g and the center c_g. If α_g is chosen to be large enough, then the goal state g_t can converge immediately to $g + \epsilon_k^g$ as follows:

$$P(\varsigma_{0,k}) = \frac{\exp(-\frac{1}{\lambda}S(\varsigma_{0,k}))}{\sum_{n=1}^{K}\exp(-\frac{1}{\lambda}S(\varsigma_{0,n}))} \tag{59}$$

$$\delta g = \sum_{k=1}^{K}[P(\varsigma_{0,k})\,\epsilon_k^g] \tag{60}$$

$$g^{\text{update}} = g + \delta g \tag{61}$$

where $P(\varsigma_0, k)$ is the probability, which is calculated by the total cost of the trajectory. It should be noted that $P(\varsigma_0, k)$ in (59) is equivalent to (52) with $t = 0$. Probability-weighted averaging (60) and goal updating (61) follow similar procedure when updating θ.

VI. Experiment Verification

A. Experimental Description

To verify the effectiveness of the proposed approach, the experiment is partitioned into reaching target position and object

TABLE II
EXPERIMENT PARAMETERS

Name	Parameters and Results
The length vector of three links	$[26.50, 33.50, 15.00]^T$ cm
Parameters of DMPs equation	$\alpha = \alpha_z = 25, \beta = \beta_z = 25/4, \alpha_g = \alpha_x = 25/3, \varsigma = 0.5$
Parameters of cost function	$Q_q = 10^{-6}, R = 10^{-4}, R_N = Q_N = 1000$
The position of the target object	$g_w = [0,0.6,0.85]^T m$
The target position in joint space	$g_j = [0.9557, -1.2789, 0.8829, -1.5659, 0.7599, 1.7135, 1.7135]^T$ rad
The optimization scheme in (17)–(20)	$\mu = 4$, and $\lambda = 10^8$
The mass of the goal object	$m = 0.25$ kg
The grab friction coefficient	$\mu = 0.82$

grasping with the uncertain external perturbations using developed experimental platform, as shown in Fig. 1. First, given the landmark of the spatial target object, whose positions can be calculated by Bumblebee stereo camera. Second, the target position (g_w) in operational space is transformed into the joint angle position (g_j) in the joint space using the PDNN optimization, where the front five elements represent, respectively, the goal of the five joint positions of robotic arm, and the last two elements represent, respectively, the joint positions of the two wheel angles of the mobile platform. Third, the RL using the DMP technique is utilized to learn joint target trajectory, and the robotic hand is driven to the target position and grasp the given object.

In the vision feedback, we can obtain the coordinates of the feature point from color information and the known geometrical relationship among the five colored circular blobs of the particular mark, then the pose of the end-effector through image coordinates and Cartesian coordinates. In addition, we can compute the orientation of the arm on the image plane in each control cycle from the image coordinates of the colored circular blobs geometrical centers.

The DMP used in this task has ten dimensions: the position of the end-effector (three dimensions), the orientation in quaternion (three dimensions), and the joint angles of the hand (four dimensions). The goal of the DMP is initialized with one desired trajectory by the PDNN, whose corresponding policy parameters are obtained by training the DMP through the supervised learning. The parameters θ are also optimized in the process of RL, which decides mainly the learning curve shape.

The cost function is chosen as $S(\varsigma_i) = \phi_{t_N} + \int_{t_i}^{t_N} (r_t + \frac{1}{2}\theta_t^T R\theta_t)dt$, where, ϕ_{tN} represents the terminal cost at t_N, $\phi_{t_N} = Q_N(q_t - g)^T(q_t - g) + R_N \dot{q_t}^T \dot{q_t}$, q_t and $\dot{q_t}$, respectively, represent the actual position and velocity value of the joint. $r_t + \frac{1}{2}\theta_t^T R\theta_t$ expresses time cost, $r_t = \frac{1}{2}(Q_q \ddot{q_t}^T \ddot{q_t})$, $\theta_t = \theta + \varepsilon_t$. The experiment parameters are listed in Table 2.

B. Experimental Result and Analysis

Grasping task execution takes 6 s. Through several experiments, we have achieved satisfactory results. Table 2 shows the experimental parameters and results. The grasped cylindrical object is shown in Fig. 6. Based on the description in Section III, the images of the left and right cameras have parallax on the horizontal axis. From Fig. 5, we can find that the target points

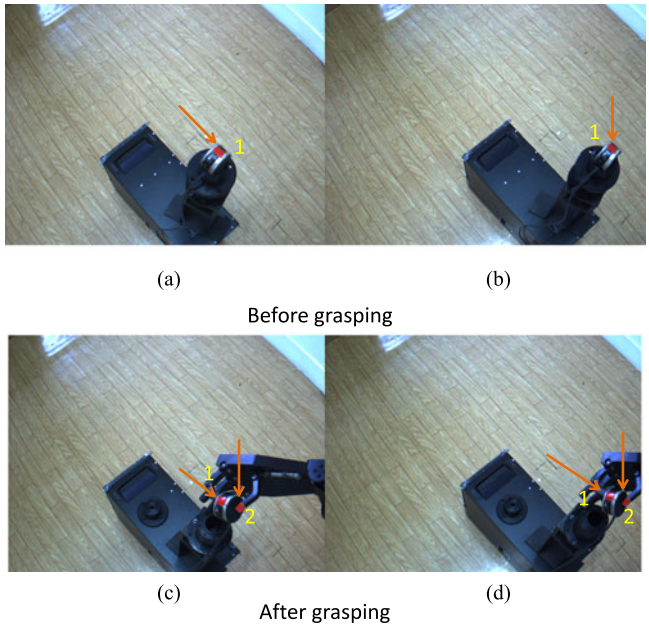


Fig. 6. Mobile manipulator grasps the object. (a) Object landmark in the left camera. (b) Object landmark in the right camera. (c) Object and grasping landmark in the left camera. (d) Object and grasping landmark in the right camera.

1 and 2 have very small parallax in the left and right cameras. Therefore, the experimental results show good performance of the vision system.

The experimental results are shown in Figs. 7–9. Fig. 7(a)–(e) and Fig. 9(a)–(b) demonstrate that the actual trajectories converge to the desired trajectories in each joint after 80 iterative learning. They almost converge to the target positions in Fig. 7(a)–(e). For the comparison, we have conducted the comparison experiments without the RL technique in the same condition. The comparison results are shown in Fig. 10, where the dashed line represents the desired trajectory by the planning of the DMP, and the solid line shows the trajectory planning without learning. We can see that the trajectory planned by the DMP cannot converge to the desired trajectory well. The main reason is that the external disturbances which have much influence on the planned trajectory are not considered.

By comparing the two experiments in Fig. 10 and Fig. 7(a)–(e), we can see that the RL algorithm can suppress the disturbances and uncertainties. Fig. 7(f) and Fig. 8(a)–(d)

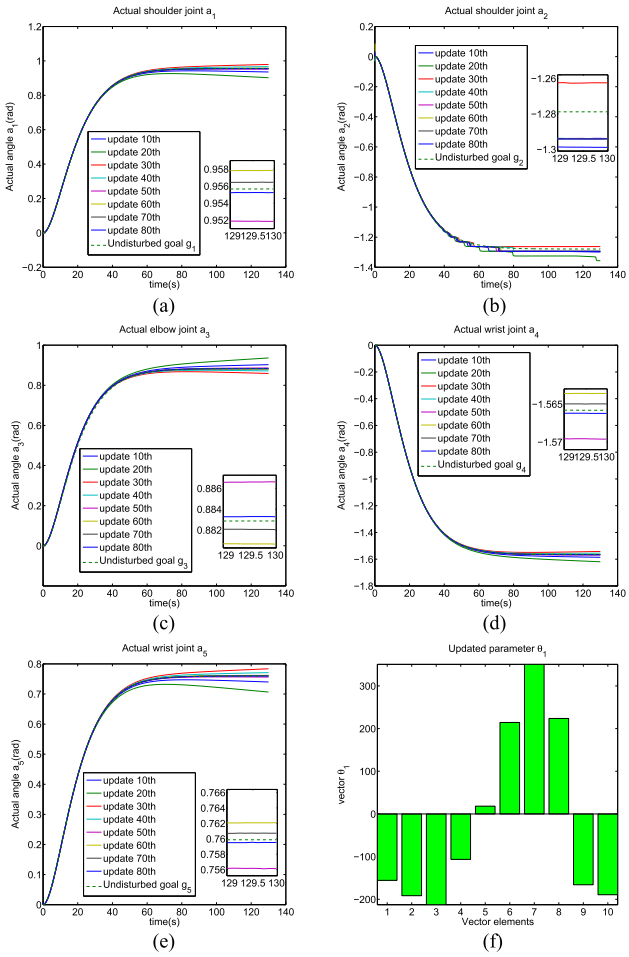


Fig. 7. The joint trajectories of the learning parameters. (a) Trajectory of the joint 1. (b) Trajectory of the joint 2. (c) Trajectory of the joint 3. (d) Trajectory of the joint 4. (e) Trajectory of the joint 5. (f) Joint 1 learning parameter.

represent the parameter learning result of every joint of the robotic right arm. Fig. 11(a) shows the total cost of five joints learning in joint space, after 20 iterations, and the total cost converges to a stable value, which indicates that the trajectory is convergent and the end-effector of the mobile manipulator is very close to the target point. From these figures, we can see that the humanoid mobile robot first moves forward and then the end-effector of the right arm reaches the target point after 80 learning, and then grasps the object. The grasping positions are defined by three grasp points: $[1.50, -11.00, 3.00]^T$ mm, $[-11.00, 11.00, 3.00]$ mm, and $[14.00, 11.00, 3.00]$ mm. The transformation matrix G can be described as $G^T = [0, 0, 1, -0.11, -0.01, 0; 1.0, 0, 0, 0, 0.03, 0.11; 0, 1, 0, -0.03, 0, 0.01; 1.0, 0, 0, 0, 0.03, -0.11; 0, 0, 1.0, 0.11, 0.12, 0; 0, -1.0, 0, 0.03, 0, 0.12; 1.0, 0, 0, 0, 0.03, -0.11; 0, 0, 1.0, 0.11, -0.14, 0; 0, -1.0, 0, 0.03, 0, 0.14]$. We can obtain the desired torque of each finger according to the friction coefficient and the weight of the target through the recurrent neural network and regulate the contact forces with the feedback value of the force sensor. Fig. 11(b) shows the optimized contact force and the grasping force. Fig. 11(c)–(d) shows three grasping fingers trajectories.

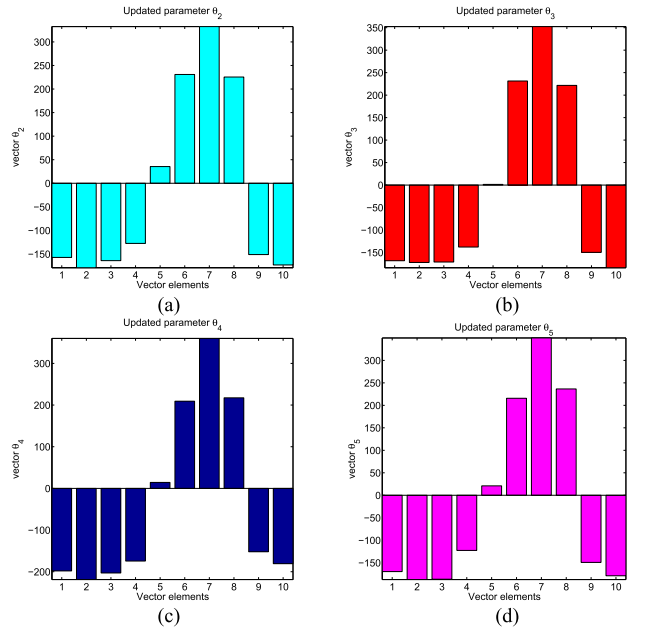


Fig. 8. Learning parameters of every joint of the robotic arm. (a) Joint 2 learning parameter θ_2. (b) Joint 3 learning parameter θ_3. (c) Joint 4 learning parameter θ_4. (d) Joint 5 learning parameter θ_5.

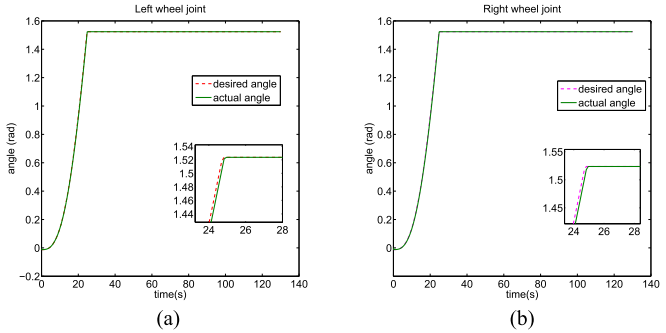


Fig. 9. Joint trajectories of the mobile base. (a) Left wheel learning parameter. (b) Right wheel learning parameter.

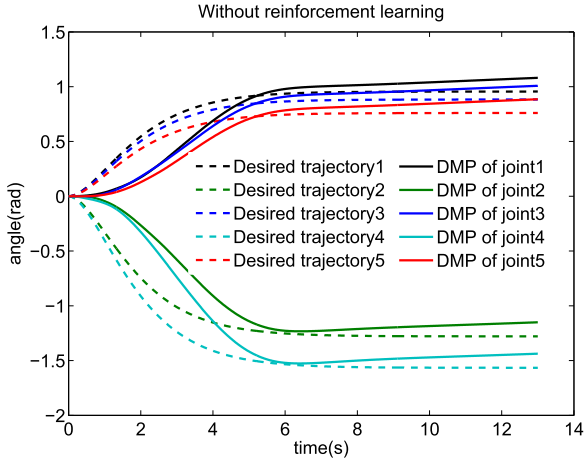


Fig. 10. DMP planning trajectory without RL.

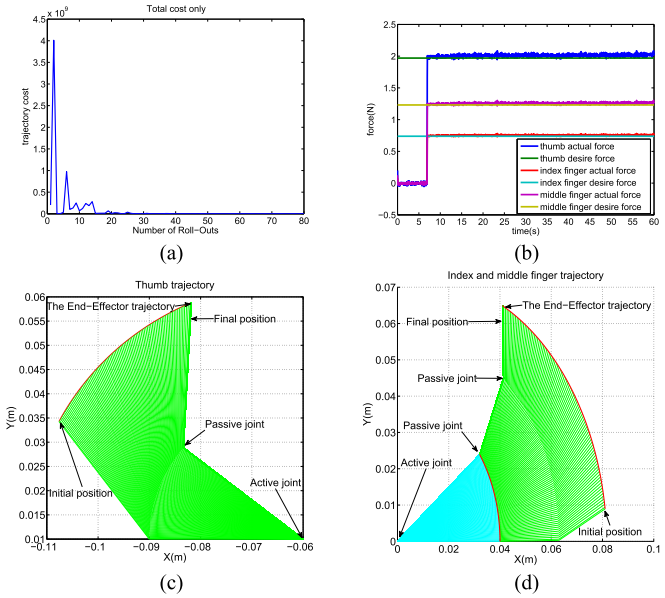


Fig. 11. Trajectories of fingers. (a) Total cost. (b) Contact force. (c) Thumb trajectory. (d) Index and middle fingers trajectory.

VII. Conclusion

In this paper, we have presented an RL strategy for manipulating and grasping a mobile manipulator, which is able to reduce the complexity of the vision feedback system and handle varying manipulation dynamics and uncertain perturbations. By the proposed two hierarchies plannings, the humanoidlike mobile robot has successfully fulfilled the given task of manipulating and grasping the objects with uncertain disturbance with good performance. The experimental results using the developed humanoidlike mobile robot have demonstrated the effectiveness of the proposed approach.

References

[1] 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.

[2] T. Endo, A. Kusakabe, Y. Kazama, and H. Kawasaki, "Haptic interface for displaying softness at multiple fingers: Combining a side-faced-type multifingered haptic interface robot and improved softness-display devices," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 5, pp. 2343–2351, Oct. 2016.

[3] S. Kim, H. Seo, S. Choi, and H. J. Kim, "Vision-guided aerial manipulation using a multirotor with a robotic arm," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 4, pp. 1912–1923, Aug. 2016.

[4] M. Stachowsky, T. Hummel, M. Moussa, and H. A. Abdullah, "A slip detection and correction strategy for precision robot grasping," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 5, pp. 2214–2226, Oct. 2016.

[5] G. Pajaka and I. Pajaka, "Sub-optimal trajectory planning for mobile manipulators," *Robotica*, vol. 33, no. 6, pp 1181–1200, Jul. 2015.

[6] P. Hebert *et al.*, "Mobile manipulation and mobility as manipulation-design and algorithms of RoboSimian," *J. Field Robot.*, vol. 32, no. 2, Mar. 2015, pp. 255–274.

[7] M. Phillips, V. Hwang, S. Chitta, and M. Likhachev, "Learning to plan for constrained manipulation from demonstrations," *Auton. Robots*, vol. 40, no. 1, pp 109–124, Jan. 2016.

[8] M. Beetz *et al.*, "Generality and legibility in mobile manipulation: Learning skills for routine tasks," *Auton. Robot.*, vol. 28, pp. 21–44, 2010.

[9] Y. Yang and O. Brock, "Elastic roadmaps—Motion generation for autonomous mobile manipulation," *Auton. Robot.*, vol. 28, pp. 113–130, 2010.

[10] C. L. Bottasso, D. Leonello, and B. Savini, "Path planning for autonomous vehicles by trajectory smoothing using motion primitives," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 6, pp. 1152–1168, Nov. 2008.

[11] A. Ude, C. G. Atkeson, and M. Riley, "Programming full-body movements for humanoid robots by observation," *Robot. Auton. Syst.*, vol. 47, no. 2–3, pp. 93–108, 2004.

[12] Y. Wada and M. Kawato, "A via-point time optimization algorithm for complex sequential trajectory formation," *Neural Netw.*, vol. 17, no. 3, pp. 353–364, 2004.

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

[14] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," *Int. J. Robot. Res.*, vol. 23, no. 4–5, pp. 363–377, 2004.

[15] Y. Zhang, S. S. Ge, and T. H. Lee, "A unified quadratic-programming-based dynamical system approach to joint torque optimization of physically constrained redundant manipulators," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 5, pp. 1083–4419, Oct. 2004.

[16] J. Fang, J. Zhao, and T. Mei, "Online optimization scheme with dual-mode controller for redundancy-resolution with torque constraints," *Robot. Comput.-Integr. Manuf.*, vol. 40, pp. 44–54, Aug 2016.

[17] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *J. Mach. Learn. Res.*, vol. 11, no. 11, pp. 3137–3181, 2010.

[18] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Phil. Trans. Royal Soc. B: Biol. Sci.*, vol. 358, pp. 537–547, 2003.

[19] V. Kamat and S. Ganesan, "A robust hough transform technique for description of multiple line segments in an image," *Proc. 1998 Int. Conf. Image Process.*, 1998, pp. 216–220.

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

[21] C. Yang, S. Amarjyoti, X. Wang, Z. Li, H. Ma, and C.-Y. Su, "Visual servoing control of baxter robot arms with obstacle avoidance using kinematic redundancy," *Intelligent Robotics and Applications* (Ser. Lecture Notes in Computer Science), vol. 9244, 2015, pp. 568–580.

[22] Y. Xia, J. Wang, and L.-M. Fok, "Grasping-force optimization for multifingered robotic hands using a recurrent neural network," *IEEE Trans. Robot. Autom.*, vol. 20, no. 3, pp. 549–554, Jun. 2004.

[23] H. Su and B. He, "A simple rectification method of stereo image pairs with calibrated cameras," in *Proc. 2010 2nd Int. Conf. Inf. Eng. Comput. Sci.*, 2010, pp. 1–4.

[24] A. M. C. Smith, C. Yang, H. Ma, P. Culverhouse, and A. Cangelosi, "Hybrid adaptive controller for manipulation in complex pertubation environments," *PLoS ONE*, vol. 10, no. 6, 2015, Art. no. e0129281.

[25] X. Wang, C. Yang, Z. Ju, H. Ma, and M. Fu, "Robot manipulator self-identification for surrounding obstacle detection," *Multimedia Tools Appl.*, vol. 76, no. 5, pp. 6495–6520, 2017, doi: 10.1007/s11042–016–3275–8.

[26] R. H. Castain and R. P. Paul, "An on-line dynamic trajectory generator," *Int. J. Robot. Res.*, vol. 3, no. 1, pp. 68–72, 1984.

[27] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*, New York, NY, USA: Springer-Verlag, 2009.

[28] B. Kim, J. Park, S. Park, and S. Kang, "Impedance learning for robotic contact tasks using natural actor-critic algorithm," *IEEE Trans. Systems, Man Cybern., B*, vol. 40, no. 2, pp. 433–443, Apr. 2010.

[29] M. Duguleana, F. G. Barbuceanu, A. Teirelbar, and G. Mogan, "Obstacle avoidance of redundant manipulators using neural networks based reinforcement learning," *Robot. Comput.-Integr. Manuf.*, vol. 28, pp. 132–146, 2012.

[30] S. Calinon, P. Kormushev, and D. G. Caldwell, "Compliant skills acquisition and multi-optima policy search with EM-based reinforcement learning," *Robot. Auton. Syst.*, vol. 61, no. 4, pp. 369–379, Apr. 2013.

[31] M. J. L. Boada, R. Barber, and M. A. Salichs, "Visual approach skill for a mobile robot using learning and fusion of simple skills," *Robot. Auton. Syst.*, vol. 38, pp. 157–170, 2002.

[32] G. Endo, J. Morimoto, T. Matsubara, J. Nakanishi, and G. Cheng, "Learning CPG-based biped locomotion with a policy gradient method: application to a humanoid robot," *Int. J. Robot. Res.*, vol. 27, no. 2, pp. 213–228, 2008.

[33] M. P. Kennedy and L. O. Chua, "Neural networks for nonlinear programming,"*IEEE Trans. Circuits Syst.*, vol. 35, no. 5, pp. 554–562, May 1988.

[34] F. T. Cheng, R. J. Sheu, and T. H. Chen, "The improved compact QP method for resolving manipulator redundancy,"*IEEE Trans. Syst. Man Cybern.*, vol. 25, no. 11, pp. 1521–1530, Nov. 1995.

[35] Z. Li, S. Xiao, S. S. Ge, and H. Su, "Constrained multi-legged robot system modeling and fuzzy control with uncertain kinematics and dynamics incorporating foot force optimization," *IEEE Trans. Syst. Man Cybern., Syst.*, vol. 46, no. 1, pp. 1–15, Jan. 2016.
[36] Y. Hu, Z. Li, G. Li, P. Yuan, R. Song, and C. Yang, "Development of sensory-motor fusion based manipulation and grasping control for a robotic hand-eye system," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 47, no. 7, pp. 1169–1180, Jul. 2017.
[37] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *J. Mach. Learn. Res.*, vol. 11, pp. 3137–3181, Oct. 2010.

Zhijun Li (M'07–SM'09) received the Ph.D. degree in mechatronics from Shanghai Jiao Tong University, Shanghai, China, in 2002.

From 2003 to 2005, he was a Postdoctoral Fellow in the Department of Mechanical Engineering and Intelligent Systems, University of Electro-Communications, Tokyo, Japan. From 2005 to 2006, he was a Research Fellow in the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, and Nanyang Technological University, Singapore. Since 2012, he has been a Professor with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China. His research interests include service robotics, teleoperation systems, nonlinear control, neural network optimization, etc.

Dr. Li has been the Co-Chair both of the Technical Committee on Biomechatronics and Biorobotics Systems (B^2S), IEEE Systems, Man, and Cybernetics Society, and the Technical Committee on Neuro-Robotics Systems, IEEE Robotics and Automation Society since 2016. He is an Editor-at-Large of the *Journal of Intelligent and Robotic Systems*, and Associate Editor of several IEEE TRANSACTIONS. He has been the General Chair and Program Chair of 2016 and 2017 IEEE Conference on Advanced Robotics and Mechatronics, respectively.

Ting Zhao received the M.S. degree in control theory and control engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2009. He is currently working toward the Ph.D. degree in control theory and control engineering with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China.

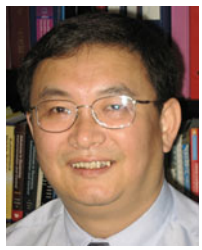
His current research interests include bionic arms and prosthesis hands, neural network control, and optimization.

Fei Chen (M'12) received the B.S. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2006; the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2008; and the Dr. Eng. degree from the Department of Micro-Nano Systems, Fukuda Laboratory, Nagoya University, Nagoya, Japan, in 2012.

In June 2013, he joined the Department of Advanced Robotics, Italian Institute of Technology, Genoa, Italy. Since then, he has been leading the Mobile Manipulation Group, focusing on robot learning and teleoperation for robotic mobile manipulators. He is the Principal Investigator of the AutoMAP Project, which is funded by the EU FP7 EUROC Project. His research interest include robotic mobile manipulation, robot three-dimensional vision and imitation learning, and human–robot collaboration.

Yingbai Hu received the B.S. degree in automation from the Hubei University of Technology, Hubei, China, in 2014. He is currently working toward the master's degree in control engineering with the College of Automation Science and Engineering, South China University of Technology, Guangzhou, China.

He has authored one paper in international conferences. His research interests include bionic arms and prosthesis hands, and electromyography signals processing.



Chun-Yi Su (SM'99) received the Ph.D. degree in control engineering from the South China University of Technology, Guangzhou, China, in 1990.

He is currently with the Department of Mechanical, Industrial, and Aerospace Engineering, Concordia University, Montreal, QC, Canada. After a seven-year stint at the University of Victoria, Victoria, BC, Canada, he joined Concordia University, in 1998. He is the author or co-author of more than 400 publications, which have appeared in journals, as book chapters, and in conference proceedings. His current research focuses on the application of automatic control theory to mechanical systems. He is particularly interested in the control of systems involving hysteresis nonlinearities.

Dr. Su was the Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, and the *Journal of Control Theory and Applications*. He has been on the editorial board of 18 journals, including the International Federation of Automatic Control journals *Control Engineering Practice* and *Mechatronics*. He was an Organizing Committee Member for many conferences, including the General Co-Chair of the 2012 IEEE International Conference on Mechatronics and Automation, and the Program Chair of the 2007 IEEE Conference on Control Applications.

Toshio Fukuda (M'82–SM'94–F'95) received the graduate degree from Waseda University, Tokyo, Japan, in 1971, and the Master of Engineering and Doctor of Engineering degrees from the University of Tokyo, Tokyo, Japan, in 1973 and 1977, respectively.

In 1977, he joined the National Mechanical Engineering Laboratory in Japan, the Science University of Tokyo, Tokyo, in 1981, and then joined the Department of Mechanical Engineering, Nagoya University, Nagoya, Japan, in 1989. He is currently one of the Thousand Talented Foreign Professors at Beijing Institute of Technology, Beijing, China. He is the Professor Emeritus with Nagoya University, having worked as Professor in the Department of Micro and Nano System Engineering and the Department of Mechanoinformatics and Systems, and as the Director of the Center for Micro and Nano Mechatronics. He was a Professor at the Shenyang University of Technology, Shenyang, China, Suzhou University, Suzhou, China, Institute of Automation, Chinese Academy of Science, Beijing; the Russell Springer Chaired Professor at the University of California, Berkeley, Berkeley, CA, USA, Seoul National University, Seoul, South Korea; and the Advisory Professor of the Industrial Technological Research Institute, Taiwan, etc.