Gait Synthesis and Sensory Control of Stair Climbing for a Humanoid Robot

Chenglong Fu and Ken Chen, *Member, IEEE*

***Abstract*—Stable and robust walking in various environments is one of the most important abilities for a humanoid robot. This paper addresses walking pattern synthesis and sensory feedback control for humanoid stair climbing. The proposed stair-climbing gait is formulated to satisfy the environmental constraint, the kinematic constraint, and the stability constraint; the selection of the gait parameters is formulated as a constrained nonlinear optimization problem. The sensory feedback controller is phase dependent and consists of the torso attitude controller, zero moment point compensator, and impact reducer. The online learning scheme of the proposed feedback controller is based on a policy gradient reinforcement learning method, and the learned controller is robust against external disturbance. The effectiveness of our proposed method was confirmed by walking experiments on a 32-degree-of-freedom humanoid robot.**

***Index Terms*—Cascade control, legged locomotion, motion control, motion planning, multisensor systems.**

I. Introduction

HUMANOID robots are expected to take an important role in assisting human activities in human daily environments because of their flexibility and friendly appearance. To realize this goal, stable and robust humanoid walking in various environments is one of the most fundamental requirements. The inherent instability of biped locomotion is the underlying problem that creates this challenge.

There are a number of humanoid robots that have been recently built throughout the world. After ten years of secret research, Honda Corporation developed the humanoid robots P2, P3, and Asimo, which can perform several complicated tasks, such as walking on flat ground, turning, climbing up/down stairs, balancing, and running [1]. The Technical University of Munich constructed the anthropomorphic autonomous biped robot JOHNNIE for the realization of dynamic 3-D walking and jogging motion [2]. JOHNNIE with 17 degrees of freedom (DOF) has a height of about 1.8 m and a weight of about 40 kg, whereas the operating power and a part of the computational power are supplied by external sources. The Ministry of Economy, Trade, and Industry of Japan had run the Humanoid Robotics Project (HRP) from 1998 to 2002. The final goal of HRP is to create "useful" humanoid robots. Toward this goal, HRP have developed a humanoid robot called HRP-2 that can walk, lie down, and get up [3]. Recently, humanoid robots cannot only walk but can also learn and avoid collision. Tan *et al.* studied a method for a humanoid robot to learn multiple tasks in an unknown environment [4]. Ohashi *et al.* [5] and Motoi *et al.* [6] investigated a collision avoidance method and a gait planning method for pushing motion for a biped robot with an upper body.

Although many papers have been published on the planning and executing walking gaits for humanoid robots, most of them were concerned with walking on level ground [3], [7]–[10]. Only a few researchers studied the walking control in the restricted situation of circumstance, such as stairs. For example, Shih [11] constructed a 7-DOF biped robot BR-1 with variable-length legs and a translatable balance weight in the body for stair climbing, and the walking stability was insured by large feet and carefully controlling the position of the center of gravity. Figliolini *et al.* [12] developed a biped robot EP-WAR3 composed of a pantograph and a double articulated parallelogram for descending stairs. Sugahara *et al.* [13] used Stewart Platforms to ascend and descend stairs by tuning up the waist yaw and preset zero moment point (ZMP) trajectories for motion pattern generation. However, these robots executed the planned gait without a sensory feedback controller; the robots may suddenly become unstable and tend to tip over with unexpected sudden events or real-world uncertainties.

Biological investigations suggest that humans regulate their muscles to adapt to the change of environments according to the sensory inputs [14]; therefore, for a humanoid robot to be able to stably and robustly walk in various environments, the walking controller must be able to adapt to the ground conditions based on the sensory information. Takenaka [8] proposed a stabilizing attitude control using an inverted pendulum model based on the inclinatory error of the robot's trunk, and its effectiveness has been confirmed by the Honda humanoid robots. Kun *et al.* [15] and Hu and Pratt [16] discussed adaptive control of biped robots using neural networks. Huang and Nakamura [17] proposed a reflex controller for walking on uneven ground; the reflex controller is simple but can rapidly respond to sensory input. Unfortunately, the parameters of the aforementioned controllers had to be tuned based on trial and error by a designer or on *a priori* knowledge of the robot's dynamics by simulation. The defects of these methods are that each aspect of the feedback design looks like a special case: If the environment is different, a designer should start the optimization from the very beginning. To act in various environments, it is necessary for humanoid robots that the

Manuscript received July 25, 2007; revised January 28, 2008. This work was supported in part by the National Natural Science Foundation of China under Grant 50575119 and in part by the National High-tech Research and Development Program (863 Program) under Grant 2006AA04Z253.

The authors are with Department of Precise Instruments and Mechanology in Tsinghua University, Beijing 100084, China (e-mail: fcl@mail.tsinghua.edu.cn; kenchen@mail.tsinghua.edu.cn).

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

Digital Object Identifier 10.1109/TIE.2008.921205

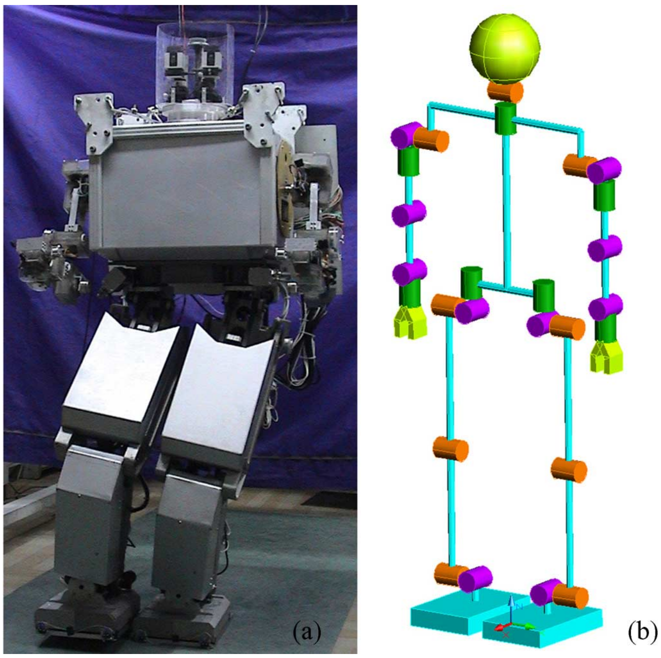


Fig. 1. THBIP-I humanoid robot and its DOF configuration.

feedback parameters can be automatically adjusted in each environment.

In this paper, we propose walking control for humanoid stair climbing, which consists of a stair-climbing gait and a sensory control strategy. The climbing gait can be considered as a feedforward control and is formulated to satisfy the environmental constraint, the kinematic constraint, and the ZMP restriction; the selection of the gait parameters is formulated as a constrained nonlinear optimization problem. The sensory control strategy can be considered as feedback control and consists of three phase-dependent sensory feedback controllers; the feedback control parameters are automatically adjusted in each step by a reinforcement learning (RL) method.

The rest of this paper is organized as follows: Section II introduces the 32-DOF humanoid robot THBIP-I and the motion equations during the stair climbing. Section III presents a gait synthesis method for humanoid stair climbing. In Section IV, a sensory feedback strategy is proposed, and the control parameters are automatically learned by an online learning scheme. Section V introduces the walking experiments, and Section VI concludes this paper.

II. Modeling

A. 32-DOF Humanoid Robot THBIP-I

Fig. 1(a) shows the humanoid robot THBIP-I, which is supported by the Tsinghua University 985 Humanoid Biped Project [18], [19]. One goal of this project is to realize complete online motion control of the biped robot walking in various environments based on a sensory feedback control. Currently, this robot has the abilities of stable walking on the ground, turning in any direction, and continuously climbing up/down stairs. Fig. 1(b) shows the DOF configuration of the robot. This robot is self-contained, and its total weight with battery is about 130 kg, and its height is about 1.8 m.

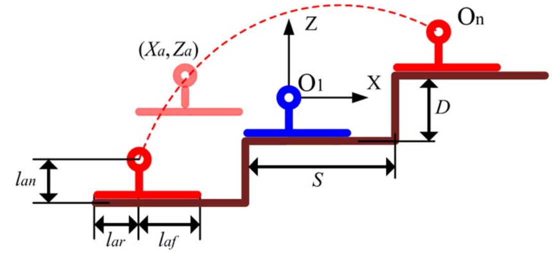


Fig. 2. Stair parameters and foot motion during stair climbing.

The humanoid robot has sensory devices including rotational encoders, six-axis force/torque sensors, gyrometers, and accelerometers. Rotational encoders are equipped in each joint and used to measure the real joint angles. The six-axis force/torque sensors are placed in the feet to measure the contact force between the feet and the ground. The gyrometers and accelerometers are mounted in the trunk of the robot to measure the body inclination relative to the gravitational direction. These sensors are used to realize feedback control in Section IV.

B. Motion Equations During the Stair Climbing

The stair-climbing gait can be considered as a sequence of steps. Each step is composed of two phases: 1) a single-support phase (SSP) and 2) a double-support phase (DSP). During the SSP, only one foot is stationary on the ground, and the other foot swings from the rear stair to the front stair. The locomotion system can be regarded as an open treelike kinematic chain. During the DSP, both feet are fixed on stairs, and the locomotion system can be regarded as an overactuated closed kinematic chain.

Let T_c denote the step time and T_d denote the period of the DSP. Considering human walking [20], the interval of the DSP is specified as 20% of the step time. Let n be the DOF number of the humanoid robot. It is assumed that the kinematic loop formed by the locomotion system during the DSP is cut at the contact point of the front foot. Therefore, in both phases, the same order coordinates, which are defined as $\boldsymbol{q} := (q_1, \ldots, q_n)^T$, will be used to formulate the motion equations of the humanoid.

In the DSP, vector $\boldsymbol{q}$ is subjected to holonomic constraints that can be written as (see Fig. 2 for the notations)

$$t \in [T_d, T_c], \qquad \begin{cases} C_1\left(\boldsymbol{q}(t)\right) \equiv O_1O_n \cdot X - S = 0 \\ C_2\left(\boldsymbol{q}(t)\right) \equiv O_1O_n \cdot Z - D = 0 \\ C_3\left(\boldsymbol{q}(t)\right) \equiv \varphi\left(\boldsymbol{q}(t)\right) = 0 \end{cases} \tag{1}$$

where C_1 and C_2 specify the Cartesian coordinates of the ankle joint of the front foot, C_3 specifies the inclination angle of the front foot, S is the stair size, and D is the stair height.

Combine C_1, C_2, and C_3 together in the vector-valued function $\boldsymbol{C}^h$ as

$$\boldsymbol{C}^h = \left(C_1(\boldsymbol{q}), C_2(\boldsymbol{q}), C_3(\boldsymbol{q})\right)^T. \tag{2}$$

Moreover, the Jacobian matrix

$$\Phi(\boldsymbol{q}) = \partial \boldsymbol{C}^h / \partial \boldsymbol{q} \tag{3}$$

will be used for deriving the Lagrange equation of motion.

As previously stated, in both phases, the humanoid is assumed to be an open kinematic chain since the front foot contact has been explicitly specified through closure constraints. This results in stating a Lagrangian dynamic model with Lagrange multipliers subjected to holonomic constraints. Using the method of Lagrange, the motion equations can be written in the form

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{G}(\boldsymbol{q},\dot{\boldsymbol{q}}) = \boldsymbol{B}\boldsymbol{\tau} + \Phi^T(\boldsymbol{q})\lambda \tag{4}$$

where $\boldsymbol{M}$ is the inertia matrix, matrix $\boldsymbol{C}$ contains Coriolis and centrifugal terms, vector $\boldsymbol{G}$ represents the gravity terms, $\boldsymbol{B}$ is the input matrix, τ is the vector of actuating joints torques, and λ is the Lagrange multiplier with respect to holonomic constraints $\boldsymbol{C}^h$.

III. Gait Synthesis for Humanoid Stair Climbing

If swing foot trajectory, hip joint trajectory, and torso angle trajectory are determined, all joint trajectories can be calculated by inverse kinematics [9]; therefore, the gait of stair climbing can be expressed as swing foot trajectory, hip joint trajectory, and torso angle trajectory.

A. Swing Foot Trajectory

To adapt to stair conditions, swing foot trajectory must be specified first. In this paper, we assume that the swing foot is always level with the ground during stair climbing; thus, the swing foot trajectory can be denoted only by the Cartesian coordinate of the swing ankle position $(x_a(t), z_a(t))$ in the SSP.

To describe the swing foot motion, it is necessary to specify one middle point (X_a, Z_a) to avoid collision with the stairs, as shown in Fig. 2. According to the kinematic constraints, the following position constraints must be satisfied:

$$(x_a(t), z_a(t)) = \begin{cases} (-S,-D), & t = T_d \\ (X_a, Z_a), & t = T_m \\ (S, D), & t = T_c \end{cases} \tag{5}$$

where T_m is the time of the middle point. T_m is one of the gait parameters to be optimized in Section III-D.

Since the swing foot is in contact with the ground at the beginning and the end of the SSP, the following derivative constraints must be satisfied:

$$(\dot{x}_a(t), \dot{z}_a(t)) = \begin{cases} 0, & t = T_d \\ 0, & t = T_c. \end{cases} \tag{6}$$

To satisfy constraints (5) and (6) and the continuity conditions of the first and second derivatives at all breakpoints, $(x_a(t), z_a(t))$ are characterized by two third-order polynomial expressions. Thereby, one can obtain the swing foot trajectory by third-order spline interpolation. By varying the values of S, D, X_a, and Z_a, the robot can produce different foot trajectories and easily negotiate different stairs.

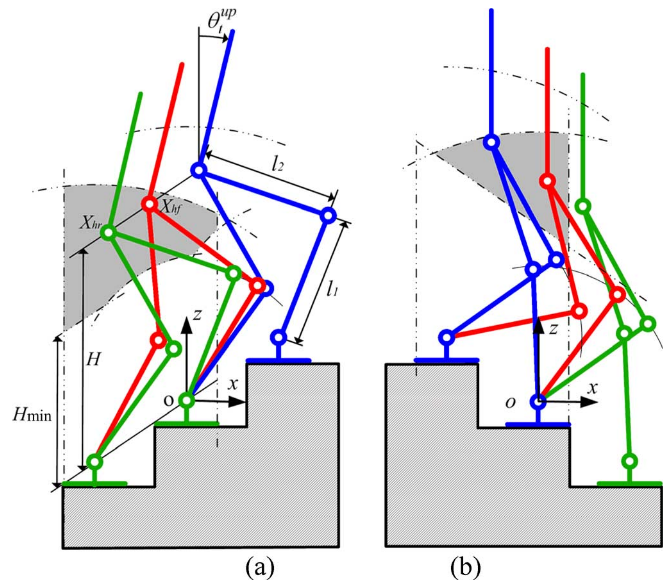


Fig. 3. Feasible region of the hip position in the DSP during stair climbing. (a) Climbing up stairs. (b) Climbing down stairs.

B. Hip Joint Trajectory

The torso trajectory of the humanoid robot can be denoted by a vector $[x_h(t), z_h(t), \theta_t(t)]^T$, where $(x_h(t), z_h(t))$ is the coordinate of the hip joint and $\theta_t(t)$ is the angle of the torso.

To satisfy kinematic constraints, the position of the hip joint in the DSP should be limited in a feasible region. Considering the maximum length of both legs in the DSP, the coordinate of the hip joint should satisfy

$$x_h^2 + z_h^2 \le (l_1 + l_2)^2 \tag{7}$$

$$(x_h + S)^2 + (z_h + D)^2 \le (l_1 + l_2)^2 \tag{8}$$

where l_1 and l_2 are the length of tibia and femur, respectively, as shown in Fig. 3.

To avoid interference between both the tibias and the stairs in the DSP, the coordinate of the hip joint should satisfy

$$(x_h - x_{\mathrm{kf}})^2 + (z_h - z_{\mathrm{kf}})^2 \ge l_2^2 \tag{9}$$

$$(x_h - x_{\mathrm{kr}})^2 + (z_h - z_{\mathrm{kr}})^2 \ge l_2^2 \tag{10}$$

where

$$x_{\mathrm{kf}} = \frac{l_1(S - l_{\mathrm{ar}})}{\sqrt{(S - l_{\mathrm{ar}})^2 + (D - l_{\mathrm{an}})^2}} \tag{11}$$

$$z_{\mathrm{kf}} = \frac{l_1(D - l_{\mathrm{an}})}{\sqrt{(S - l_{\mathrm{ar}})^2 + (D - l_{\mathrm{an}})^2}} \tag{12}$$

$$x_{\mathrm{kr}} = \frac{l_1(S - l_{\mathrm{ar}})}{\sqrt{(S - l_{\mathrm{ar}})^2 + (D - l_{\mathrm{an}})^2}} - S \tag{13}$$

$$z_{\mathrm{kr}} = \frac{l_1(D - l_{\mathrm{an}})}{\sqrt{(S - l_{\mathrm{ar}})^2 + (D - l_{\mathrm{an}})^2}} - D. \tag{14}$$

From the viewpoint of reducing the torque on the supporting knee joint, the hip joint should be avoided at a low position relative to the stair; therefore, the following constraint should be satisfied:

$$z_h - x_h \tan^{-1}\frac{D}{S} - H_{\min} \geq 0 \tag{15}$$

where $H_{\min}$ is the lowest possible position of the hip joint during stair climbing, as shown in Fig. 3.

During the DSP, we specify the abscissa of the hip joint to vary within a fixed range:

$$-S - l_{\mathrm{ar}} \leq x_h \leq l_{\mathrm{af}}. \tag{16}$$

According to (7)–(10), (15), and (16), the feasible region of the hip joint in the DSP can be determined, as shown in Fig. 3.

Let X_{hr} denote the abscissa of the hip joint at the initial start of the DSP and X_{hf} denote the abscissa of the hip joint at the terminal of the DSP. These two parameters have a great effect on the performance of stair climbing and must be selected in the aforementioned feasible region.

The abscissa of the hip joint during a step should satisfy the following position constraints:

$$x_h(t) = \begin{cases} X_{\mathrm{hr}}, & t = 0 \\ X_{\mathrm{hf}}, & t = T_d \\ X_{\mathrm{hr}} + S, & t = T_c. \end{cases} \tag{17}$$

To satisfy the continuity conditions of the first and second derivatives at the transition of each step, the following derivative constraints should be satisfied:

$$\begin{cases} \dot{x}_h(0) = \dot{x}_h(T_c) \\ \ddot{x}_h(0) = \ddot{x}_h(T_c). \end{cases} \tag{18}$$

By third-order polynomial interpolation, one can obtain the $x_h(t)$ during a one-step cycle.

We specify that the hip joint moves in a line parallel to the stair during the climbing; therefore, the hip joint motion $z_h(t)$ can be expressed as

$$z_h(t) = \tan^{-1}\left(\frac{D}{S}\right) \cdot x_h(t) + H \tag{19}$$

where H is the distance between the hip joint and the stair during stair climbing, as shown in Fig. 3. Since the hip joint moves in a line, all points of the hip position in both the DSP and SSP will satisfy the kinematic constrains (7)–(10), (15), and (16).

In a similar way, one can synthesize the hip joint trajectory of the climbing-down gait and transition gait between the flat ground and the stairs.

C. Torso Angle Trajectory

In previous studies, some researchers [11]–[13] have planned the torso angle trajectory of biped robots to be upright during stair climbing. However, for humanoid robots with a large-mass trunk, the torso angle has a major effect on the stability of walking; therefore, if the torso keeps upright during the stair climbing, the humanoid robot will need a large ankle torque to maintain the balance of the supporting foot. To solve this problem, we propose that the torso angle of the climbing-up gait be proportional to the grads of the stair, i.e.,

$$\theta_t^{\mathrm{up}} = k_t^{\mathrm{up}} \cdot \frac{D}{S} \tag{20}$$

where k_t^{up} is one of the gait parameters to be optimized in Section III-D.

During the transition between the flat ground and stairs, the torso angle is specified as

$$\theta_t^{\mathrm{tran}}(t) = 2\frac{\theta_t^{\mathrm{grnd}} - \theta_t^{\mathrm{up}}}{(T_c - T_d)^3}t^3 - 3\frac{\theta_t^{\mathrm{grnd}} - \theta_t^{\mathrm{up}}}{(T_c - T_d)^3}t^2 + \theta_t^{\mathrm{grnd}} \tag{21}$$

where θ_t^{grnd} is the torso angle during walking on the level ground.

D. Determining Gait Parameters Through Optimization

When stair size S, D, and step period T_c are determined, the climbing gait during a step is represented by the following five parameters: T_m, X_{hr}, X_{hf}, H, and k_t^{up}. Let $\boldsymbol{W} = [T_m \quad X_{\mathrm{hr}} \quad X_{\mathrm{hf}} \quad H \quad k_t^{\mathrm{up}}]$. To automatically obtain appropriate gait parameters $\boldsymbol{W}$, we formulate the selection problem as a constrained nonlinear optimization problem with available numerical optimization tools.

Posing the constrained nonlinear optimization problem requires two ingredients: set constraints and a cost function.

During humanoid stair climbing, the constrained conditions can be classified as the following three conditions: 1) unilateral contact condition; 2) ZMP condition; and 3) kinematic constraint condition.

The first constraint condition is to describe the unilateral contact between the supporting foot and the ground without slippage, which can be written as

$$F_z > 0 \tag{22}$$

$$\sqrt{F_x^2 + F_y^2} < \mu_s \cdot F_z \tag{23}$$

where F_x, F_y, and F_z are reactive forces on the support foot, and μ_s is the friction coefficient between the sole of the robot and the stairs.

The second one is the ZMP restriction. Having the ZMP inside the support polygon is the true dynamic condition that guarantees that the support foot will not rotate [21]. For stair climbing, define the support polygon as a region $S \subset \mathbb{R}^2$, which is the convex hull of the humanoid robot. We use ∂S to denote the set of all boundary points of S. Consequently, this constraint condition can be expressed as

$$\sqrt{(x' - x_{\mathrm{zmp}})^2 + (y' - x_{\mathrm{zmp}})^2} \geq d_{\mathrm{zmp}} \tag{24}$$

where $(x', y') \in \partial S$, and d_{zmp} denotes the stability margin in the sense of ZMP stability criterion.

The last one comes from the kinematic constraint of the hip joint position, which includes the restriction of leg length and the interference between the tibias and the stairs during stair climbing, as analyzed in Section III-D [see (7)–(16)].

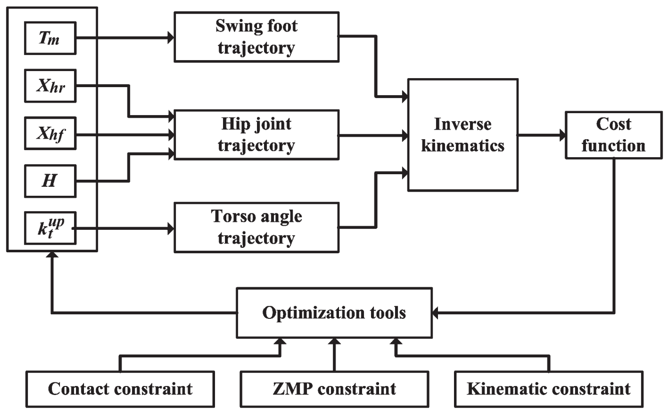


Fig. 4. Flow chart of the gait planning for stair climbing. The gait is represented by five parameters and optimized by available optimization tools.

We considered the minimum consumed energy (CE) as a cost function in this paper. There are two reasons for this choice. First, the autonomous humanoid robot with an energy-efficient gait will have a long walking distance. Second, according to [10], walking gaits with the minimum CE are similar to human gaits. For minimum CE cost function, it can be assumed that the energy to control the position of the robot is proportional to the integration of the square of the torque with respect to the time; therefore, the cost function can be defined as follows:

$$J = \frac{1}{2}\int_{0}^{T_d}\left[\|\tau(t)\|_2^2 + \|\lambda(t)\|_2^2\right]dt + \frac{1}{2}\int_{T_d}^{T_c}\|\tau(t)\|_2^2\,dt \qquad (25)$$

where the multipliers λ are implicitly considered as additional active forces that are needed to hold the front foot in its prescribed position [see (4)]. Minimizing λ is equal to minimizing antagonistic forces against the stair during the DSP, which could be the cause of break or sliding in contact.

In the preceding generic form, the parameter optimization problem may be solved with any available numerical optimization tool. In this paper, the optimization problem was solved with Matlab's constrained nonlinear optimization tool *fmincon*. Continuity of the cost function (25) with respect to the gait parameters, along with the use of a small optimization step size, makes the use of the gradient-based *fmincon* algorithm feasible. Fig. 4 shows the whole process that generates the stair-climbing gait.

IV. Sensory Feedback Control

Although the stair-climbing gait synthesized in Section III satisfies the stability constraint, the humanoid robot may also lose its balance and threaten to tip over because of the misalignment of the ideal condition and the real one. This misalignment comes from not only the humanoid robot but also the environment. Therefore, it is desirable to modify the original gait according to the information of sensors implemented in the humanoid robot.

The sensory feedback controller presented in this section is phase dependent and consists of a torso attitude controller, a ZMP compensator, and an impact reducer. Fig. 5 shows the overall control block diagram for stair climbing.

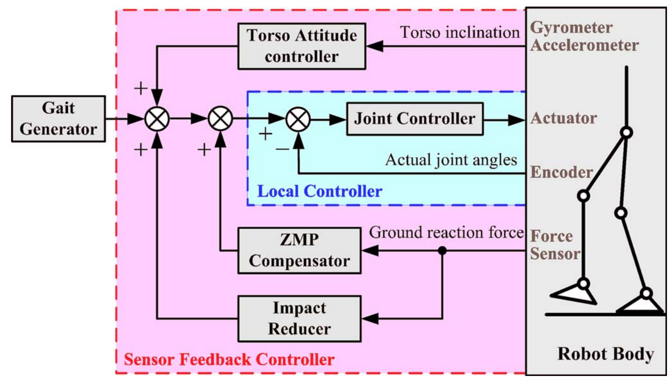


Fig. 5. Overall control strategy consisting of the sensory feedback controller and the local joint controller.

A. Torso Attitude Controller

One of the basic aspects of humanoid locomotion is to maintain a desired torso attitude. During stair climbing, the upper body of the humanoid robot tends to slant from the desired torso attitude. If the torso attitude cannot be recovered in time, the tipping moment will become large, and the humanoid may finally tip forward or backward. Since the hip joint is the nearest joint to the torso, the most effective way to recover the robot's tipping over posture is to modify the hip joint of the support leg [17].

The control process of the torso attitude controller can be summarized as follows: 1) Determine which leg is the support leg by the values of the force sensors. Since the hip joint of the swing leg does not affect the torso attitude, only the hip joint of the support leg needs to be modified. 2) Modify the desired value of the support-hip joint according to the information of inclination sensor, such as gyrometers and accelerometers. The modification of the support-hip joint, as shown in Fig. 6, is given as follows:

$$\Delta q_{h\_sp}(t) = K_T^P\left(\theta_t^d(t) - \theta_t^r(t)\right) + K_T^D\left(\dot{\theta}_t^d(t) - \dot{\theta}_t^r(t)\right) \qquad (26)$$

where K_T^P and K_T^D are the coefficient to be learned online in Section IV-D; $\theta_t^r(t)$ is the real torso inclination detected by the accelerometer and the gyrosensors, which are located on the torso of the humanoid; and $\theta_t^d(t)$ is the desired torso inclination determined from the expected gait.

It should be noted that the torso attitude controller is phase dependent and is effective only for the support-hip joint, as shown in Fig. 7.

B. ZMP Compensator

The ZMP compensator is used to make the actual ZMP trajectory coincide with the desired ZMP trajectory as near as possible. The actual ZMP is measured by the universal force-moment sensors implemented in the feet of the robot during walking. If the actual ZMP coincides with the desired ZMP

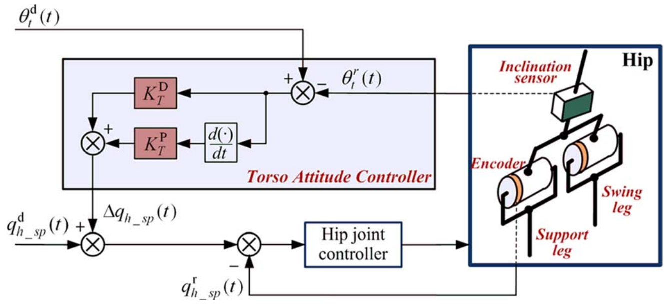


Fig. 6. Torso attitude controller. The input of this controller is the actual torso angle, and the output of the controller is modification of the support-hip joint.

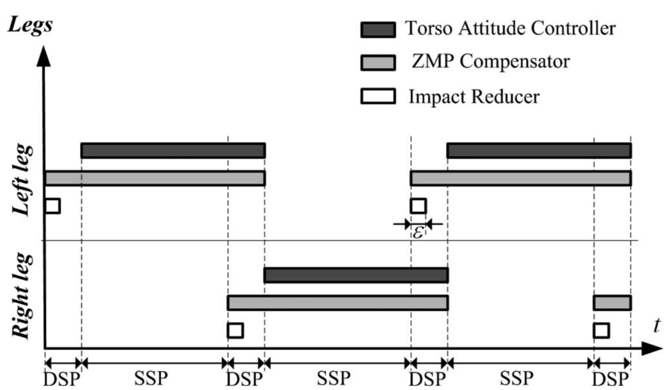


Fig. 7. Phases of the sensory feedback controller.

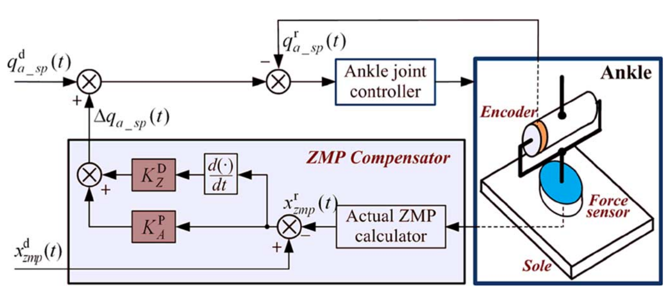


Fig. 8. ZMP compensator. The input of this controller is the ground forces, and the output of the controller is modification of the support-ankle joint.

determined from the expected gait, the robot will walk in the desired pattern; however, if a modeling error or a disturbance causes a large divergence between the actual ZMP and the desired ZMP, the robot may be in danger of tipping over.

Since the support-ankle joint has the largest effect on ZMP, the most effective way to compensate the actual ZMP is to modify the support-ankle joint. As shown in Fig. 8, the modification

$$\Delta q_{a\_sp}(t) = K_Z^P \left(x_{\mathrm{zmp}}^d(t) - x_{\mathrm{zmp}}^r(t)\right) + K_Z^D \left(\dot{x}_{\mathrm{zmp}}^d(t) - \dot{x}_{\mathrm{zmp}}^r(t)\right) \quad (27)$$

where K_Z^P and K_Z^D are the coefficient to be learned online in Section IV-D; $x_{\mathrm{zmp}}^r(t)$ is the real ZMP calculated based on the force sensors, which are located in the feet of the humanoid robot; and $x_{\mathrm{zmp}}^d(t)$ is the desired ZMP trajectory, which can be determined from the expected gait.

The control phase of the torso attitude controller is shown in Fig. 7, and it is effective only when the foot is in contact with the stair. When the foot is in the swing phase, the revised pattern of the foot is returned to the preset walking pattern gradually.

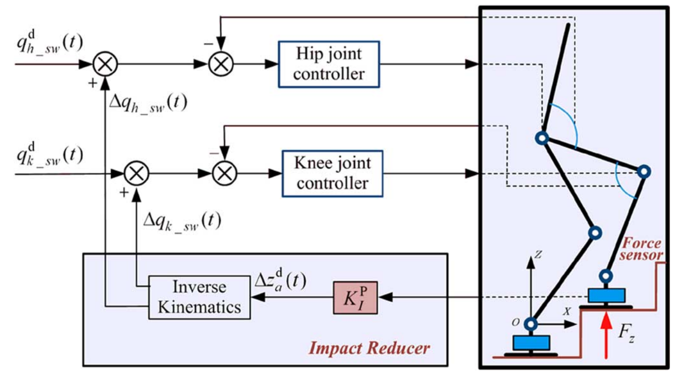


Fig. 9. Impact reducer. The input of this controller is the ground forces, and the output of the controller is modification of the swing-hip joint and the swing-knee joint.

C. Impact Reducer

The impact reducer is used to reduce the magnitude of impact and guarantee a stable foot at foot landings on the stair. During stair climbing, the swing foot may land on the stair faster or slower than the desired time; therefore, an impact force may occur between the swing foot and the stair. If this impact force cannot be reduced in time, a large tipping moment will be produced, and the robot may tip backward.

Observing human walking, we find that a human being controls his/her leg muscles for shock absorption. The muscles are relaxed to absorb the impact force just before landing, whereas the muscles are hardened to maintain the posture after landing. To imitate the elasticity of human muscles, the impact reducer is designed. When the force sensors detect the impact, the robot retracts its swing foot to reduce the magnitude of the impact force. The value of the foot retraction is given as follows:

$$\Delta z_a^d(t) = K_I^P \cdot F_z \quad (28)$$

where K_I^P is the coefficient to be learned online, and F_z is the foot contact force with the stair, which is measured by a foot-force sensor. Fig. 9 shows the diagram of the impact reducer, where $q_{h\_sw}^d$ and $q_{k\_sw}^d$ denote the desired values of the swing-hip joint and the swing-knee joint, respectively.

The control phase of the impact reducer is shown in Fig. 7, and it is effective only in the initial of the DSP. In particular, this period is set as $\varepsilon = 0.2$ s in this paper. In the last DSP, the revised pattern of the landing leg is gradually returned to the desired walking pattern.

D. Learning Feedback Parameters

As analyzed in Section I, in previous studies, the parameters of sensory feedback controllers of the humanoid robots had to be tuned based on trial and error by a designer [17] or on *a priori* knowledge of the robot's dynamics by simulation [8].

To act in various environments, it is necessary for humanoid robots that the feedback parameters can be automatically adjusted in each environment. To solve this problem, we propose

TABLE I
LEARNING PROCEDURE

```
//i : Step number
SET the feedback parameters K to an initial value
SET t0=1, i=1
REPEAT // Learning episodes
    FOR t = ti to ti + tmax
        Execute the feedback parameters k with probability PK(k)
        Observe the step reward rs
        Calculate ej(t) and Aj(t) as
            ej(t) = ∂/∂kj ln(PK(k))
            Aj(t) = ej(t) + γAj(t−1)
        Calculate Δkj(t) as
            Δkj(t) = (rs − b)Aj(t)
        IF policy_update_condition THEN
            ΔK(t) = (Δk1(t),···,Δkj(t),···,Δk5(t))
            K = K + α(1−γ)ΔK(t)
        END IF
        IF terminal_condition THEN
            SET ti= t + 1, i =i +1
            break
        END IF
    END FOR
UNTIL
```

an RL method, which learns from the interaction with the environment to learn the feedback parameters. In particular, we employ a stochastic policy gradient method [22] in RL to obtain the feedback parameters $\boldsymbol{K} = \left[K_T^P \;\; K_T^D \;\; K_Z^P \;\; K_Z^D \;\; K_I^P\right]$.

The aim of RL is to obtain the optimal policy, which maximizes the reward accumulation toward the future. The RL algorithm is applied on a step-by-step basis, which means that the learning system observed a reward and updated the feedback parameters in every step.

Table I shows the pseudocode of the learning algorithm, where γ and $\alpha \in (0, 1]$ are a discount factor and a learning rate factor, respectively [23], $\boldsymbol{e}$ is the eligibility, A is the eligibility trace, r_s is the step reward, and b is a constant offset.

The stochastic policy was defined as a normal distribution and given as follows:

$$P_K(\boldsymbol{k}) = N(\boldsymbol{k}; \boldsymbol{K}, \Sigma) = \frac{1}{\sqrt{2\pi}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{k}-\boldsymbol{K})^T\Sigma^{-1}(\boldsymbol{k}-\boldsymbol{K})\right\} \quad (29)$$

and covariance Σ is given by

$$\Sigma = \begin{bmatrix} \left(\sigma_T^P\right)^2 & 0 & 0 & 0 & 0 \\ 0 & \left(\sigma_T^D\right)^2 & 0 & 0 & 0 \\ 0 & 0 & \left(\sigma_Z^P\right)^2 & 0 & 0 \\ 0 & 0 & 0 & \left(\sigma_Z^D\right)^2 & 0 \\ 0 & 0 & 0 & 0 & \left(\sigma_I^P\right)^2 \end{bmatrix} \quad (30)$$

where σ_T^P, σ_T^D, σ_Z^P, σ_Z^D, and σ_I^P are a constant standard deviation of the policy. For a standard deviation set to be too large, the locomotion becomes unstable due to the noisy control signal, whereas, at a too small standard deviation, the system's ability to explore a better policy decreases.

The step reward is calculated when a step is achieved, and it can be defined as

$$r_s = r_{\text{torso}} + r_{\text{zmp}} + r_{\text{impact}} + r_{\text{penalty}} \quad (31)$$

where r_{torso} is calculated as

$$r_{\text{torso}} = -\frac{1000}{T_c} \times \int_0^{T_c} \left(\theta_t^d(t) - \theta_t^r(t)\right) dt, \quad (32)$$

r_{zmp} is calculated as

$$r_{\text{zmp}} = -\frac{1000}{T_c} \times \int_0^{T_c} \left(x_{\text{zmp}}^d(t) - x_{\text{zmp}}^r(t)\right) dt, \quad (33)$$

r_{impact} is calculated as

$$r_{\text{impact}} = -\int_0^{\varepsilon} F_z(t)dt, \quad (34)$$

and r_{penalty} is calculated as

$$r_{\text{penalty}} = \begin{cases} 0, & \text{succesful step forward} \\ -100, & \text{otherwise.} \end{cases} \quad (35)$$

V. EXPERIMENTAL RESULT

In this section, we provide several experimental results of THBIP-I toward checking the proposed method.

A. *Learning a Sensory Feedback Controller*

The learning scheme of stair climbing consists of two stages. In the first stage, the robot learns the control parameters from walking on level ground. Since no prior knowledge can be utilized, the initial values of the control parameters are all set at zero at this stage. The accumulated reward received by the learning algorithm is shown in Fig. 10. The robot learns a walking parameter within 50 trials. These values are used for the initial setting of the robot in the next stage. In the second stage, the robot is controlled to climb up/down stairs with the initial parameters learned from the level ground walking. After 20 episodes, the accumulated reward reached a relatively high value. Fig. 11 shows the ZMP trajectories with/without the sensory feedback controller; from Fig. 11, one can see that the ZMP trajectories with the aforementioned controller are much smoother than those without the proposed controller. With the learned controller, the robot can stably and smoothly walk under such a gait with a step length of 300 mm and a step duration of 6 s. Fig. 12 shows the snapshot of climbing up and down stairs.

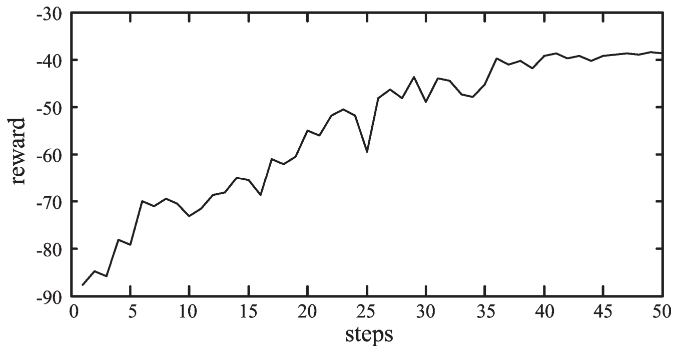
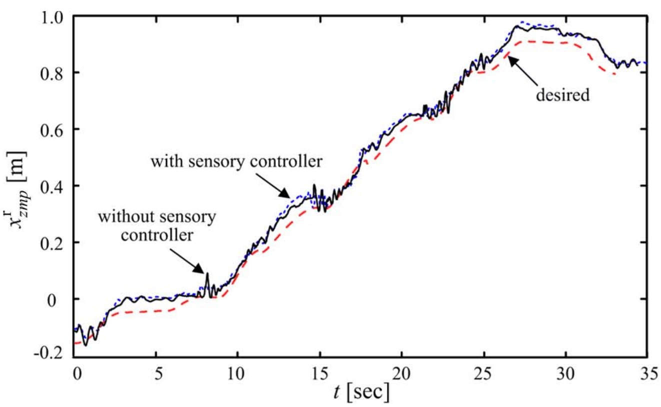
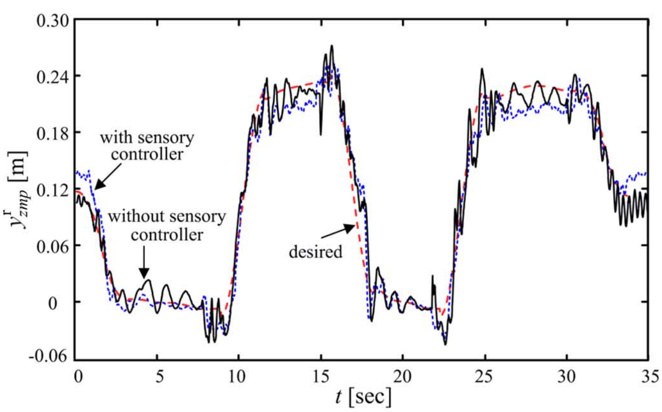


Fig. 10. Cumulative reward during level-ground walking for 50 steps.



(a)



(b)

Fig. 11. Comparison of the ZMP trajectories with and without the sensory feedback controller. (a) ZMP trajectory along the x-axis. (b) ZMP trajectory along the y-axis.

B. Robustness Against External Disturbance

To evaluate the robustness of the proposed sensory feedback controller against external disturbance, a sudden push was applied to the torso of the robot by a person during stair climbing. The push occurs in the SSP, because the robot has a relatively small supporting region in this period. Fig. 13(a) shows that the torso attitude radically changes at the collision, which means that the torso of the robot slants forward due to the unexpected push. From Fig. 13(b), it is known that the support-hip joint radically changes, and this is the result of the sensory feedback controller. With the torso attitude controller,

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 12. Snapshot of climbing up and down stairs.

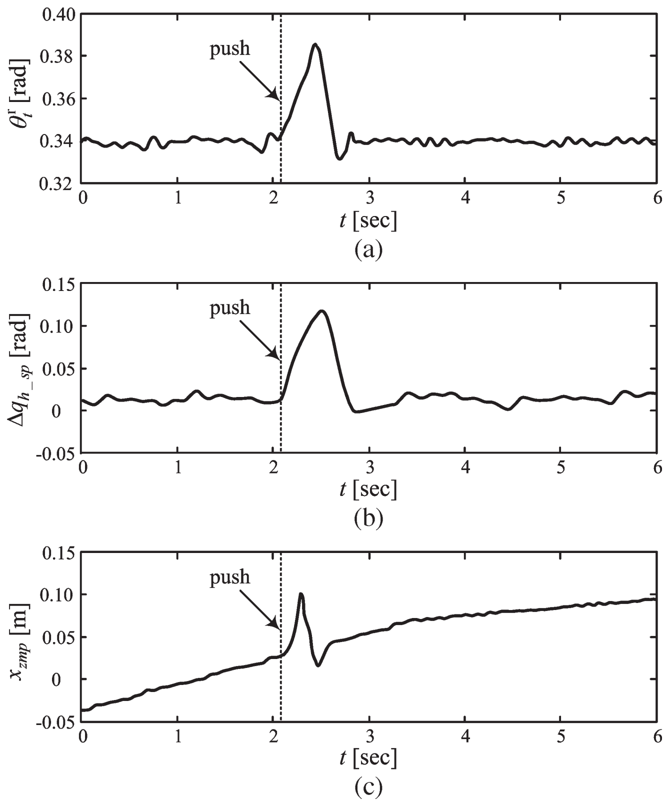


Fig. 13. Results of robustness against external disturbance. (a) Torso attitude. (b) Modification of the support-hip joint. (c) ZMP trajectory.

the desired value of the support-hip joint should be radically recovered. From Fig. 13(c), one can see that the ZMP trajectory radically changes at the moment of the push, but it quickly returns to its desired trajectory, due to the ZMP compensator.

VI. Conclusion

This paper describes walking control for a humanoid robot to realize stable and robust stair climbing.

The proposed method consists of a feedforward stair-climbing gait and a feedback sensory controller. The stair-climbing gait is parameterized by the swing foot trajectory, hip joint trajectory, and torso angle trajectory; the selection of the gait parameters is formulated as a constrained nonlinear optimization problem with available optimization tools. The sensory controller consists of the torso attitude controller, ZMP compensator, and impact reducer; the parameters of these controllers are automatically regulated in each step by a two-stage policy gradient RL method. The validity of the proposed method was confirmed by stair-climbing experiments of an actual 32-DOF humanoid robot.

We believe that our two-stage learning approach is plausible in the perspective of walking in various environments for humanoid robots, because such an approach is, at least, found in humans [24], [25]: After performing a primitive task in a low-dimensional and simple environment, human beings tend to utilize learned experiences to realize more complicated movements in high-dimensional and complex environments.

Acknowledgment

The authors would like to thank all the members of the 985 Anthropopathic Biped Project at Tsinghua University for their cooperative construction of the robot and the reviewers for their constructive comments and suggestions.

References

[1] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, "The development of Honda humanoid robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1998, pp. 1321–1326.

[2] K. Löffler, M. Gienger, F. Pfeiffer, and H. Ulbrich, "Sensors and control concept of a biped robot," *IEEE Trans. Ind. Electron.*, vol. 51, no. 5, pp. 972–980, Oct. 2004.

[3] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, and T. Isozumi, "Humanoid robot HRP-2," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2004, pp. 1083–1090.

[4] K. C. Tan, Y. J. Chen, K. K. Tan, and T. H. Lee, "Task-oriented developmental learning for humanoid robots," *IEEE Trans. Ind. Electron.*, vol. 52, no. 3, pp. 906–914, Jun. 2005.

[5] E. Ohashi, T. Aiko, T. Tsuji, H. Nishi, and K. Ohnishi, "Collision avoidance method of humanoid robot with arm force," *IEEE Trans. Ind. Electron.*, vol. 54, no. 3, pp. 1632–1641, Jun. 2007.

[6] N. Motoi, M. Ikebe, and K. Ohnishi, "Real-time gait planning for pushing motion of humanoid robot," *IEEE Trans. Ind. Informat.*, vol. 3, no. 2, pp. 154–163, May 2007.

[7] S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, and H. Hirukawa, "The 3D linear inverted pendulum mode: A simple modeling for a biped walking pattern generation," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, 2001, pp. 239–246.

[8] T. Takenaka, "Attitude stabilization control system for a legged mobile robot," U.S. Patent 5 459 659, Oct. 17, 1995.

[9] Q. Huang, K. Yokoi, S. Kajita, K. Kaneko, H. Arai, N. Koyachi, and K. Tanie, "Planning walking patterns for a biped robot," *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 280–289, Jun. 2001.

[10] G. Capi, Y. Nasu, L. Barolli, and K. Mitobe, "Real time gait generation for autonomous humanoid robots: A case study for walking," *Robot. Auton. Syst.*, vol. 42, no. 2, pp. 107–116, Feb. 2003.

[11] C. Shih, "Ascending and descending stairs for a biped robot," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 3, pp. 255–268, May 1999.

[12] G. Figliolini, M. Ceccarelli, and M. Gioia, "Descending stairs with EP-WAR3 biped robot," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, 2003, pp. 747–752.

[13] Y. Sugahara, A. Ohta, H. Lim, and A. Takanishi, "Walking up and down stairs carrying a human by a biped locomotor with parallel mechanism," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 1489–1494.

[14] G. Melvill and D. Watt, "Observations on the control of stepping and hopping movements in man," *J. Physiol.*, vol. 219, no. 3, pp. 729–737, Dec. 1971. (London).

[15] A. Kun and W. Miller, "Adaptive dynamic balance of a biped robot using neural networks," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1996, pp. 240–245.

[16] J. Hu and J. Pratt, "Adaptive dynamic control of a biped walking robot with radial basis function neural networks," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1998, pp. 400–405.

[17] Q. Huang and Y. Nakamura, "Sensory reflex control for humanoid walking," *IEEE Trans. Robot.*, vol. 21, no. 5, pp. 977–984, Oct. 2005.

[18] M. Zhao, L. Liu, J. Wang, K. Chen, J. Zhao, and K. Xu, "Control system design of THBIP-I humanoid robot," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2002, pp. 2253–2258.

[19] C. Fu, M. Shuai, K. Xu, J. Zhao, J. Wang, Y. Huang, and K. Chen, "Planning and control for THBIP-I humanoid robot," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, 2006, pp. 1066–1071.

[20] V. T. Inman, H. J. Ralston, and F. Todd, *Human Walking*. Baltimore, MD: Williams & Wilkins, 1981.

[21] M. Vukobratovic and D. Juricic, "Contribution to the synthesis of biped gait," *IEEE Trans. Biomed. Eng.*, vol. BME-16, no. 1, pp. 1–6, Jan. 1969.

[22] H. Kimura and S. Kobayashi, "Reinforcement learning for continuous action using stochastic gradient ascent," in *Proc. Intell. Auton. Syst.*, 1998, pp. 288–295.

[23] M. Sato and S. Ishii, "Reinforcement learning based on on-line EM algorithm," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 1998, pp. 1052–1058.

[24] K. M. Newell and D. E. Vaillancourt, "Dimensional change in motor learning," *Hum. Mov. Sci.*, vol. 20, no. 4, pp. 695–715, Nov. 2001.

[25] K. Hitomi, T. Shibata, Y. Nakamura, and S. Ishii, "Reinforcement learning for quasi-passive dynamic walking of an unstable biped robot," *Robot. Auton. Syst.*, vol. 54, no. 12, pp. 982–988, Dec. 2006.

Chenglong Fu was born in Heilongjiang, China, in 1980. He received the B.S. degree in mechatronics engineering from Tongji University, Shanghai, China, and the Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2002 and 2007, respectively.

He is currently an Assistant Professor with the Department of Precision Instruments and Mechanology, Tsinghua University. He has published about 20 refereed papers in several domestic and international academic journals, and international conference proceedings. His research interests include biped robots, humanoid robots, and bio-inspired control.

Dr. Fu was the recipient of the Excellent Paper Award from the Chinese Mechanical Engineering Society in 2007 and the Youth Excellent Paper Award from the Annual Conference of the Chinese Mechanical Engineering Society in 2007.

Ken Chen (M'93) was born in Sichuan, China, in 1954. He received the B.S. degree in mechanical engineering from Sichuan University, Chengdu, China, and the M.S. and Ph.D. degrees in mechanical engineering from Zhejiang University, Hangzhou, China, in 1982, 1984, and 1987, respectively.

From 1991 to 1992, he was a Visiting Professor with the University of Illinois, Chicago. From 1992 to 1995, he was a Postdoctoral Researcher with Purdue University, West Lafayette, IN. He is currently a Professor and the Director of the Institute of Manufacturing Engineering, Tsinghua University, Beijing, China. He was the author of *Robot Technology and Application* (Beijing, China: Tsinghua University Press, 2006, in Chinese) and a coauthor of *Manufacturing Systems* (Beijing, China: Tsinghua University Press, 2001, in Chinese), with L. Zheng and B. Zhang, and *Modern Manufacturing Technology* (Beijing, China: Tsinghua University Press, 2006, in Chinese), with D. Yu. He has published more than 100 refereed papers in several domestic and international academic journals and international conference proceedings. His research interests include robotics and intelligent control, humanoid robots, microrobots and small robots, medical and space robots, manufacturing automation systems, and hydraulic servo systems.

Dr. Chen is a member of the American Society of Mechanical Engineers and a Senior Member of the Chinese Mechanical Engineering Society. He was the recipient of the First Award of Chinese Youth Technology in 1988 and the Second Award of Technology Advancement from People's Liberation Army in 2004.